Results and Problems in Cell Differentiation 57

Ellen Hsu
Louis Du Pasquier *Editors*

Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations

Springer

Results and Problems in Cell Differentiation

Volume 57

More information about this series at http://www.springer.com/series/400

Ellen Hsu • Louis Du Pasquier
Editors

Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations

Springer

Editors
Ellen Hsu
Department of Physiology and Pharmacology
SUNY Downstate Medical Center
Brooklyn, New York
USA

Louis Du Pasquier
Institute of Zoology and Evolutionary Biology
University of Basel
Basel, Switzerland

Series Editors
Dietmar Richter
Center for Molecular Neurobiology
University Medical Center Hamburg-Eppendorf (UKE)
University of Hamburg
Hamburg
Germany

Henri Tiedge
The Robert F. Furchgott Center for Neural and Behavioral Science
Department of Physiology and Pharmacology
Department of Neurology
SUNY Health Science Center at Brooklyn
Brooklyn, New York
USA

ISSN 0080-1844 ISSN 1861-0412 (electronic)
Results and Problems in Cell Differentiation
ISBN 978-3-319-20818-3 ISBN 978-3-319-20819-0 (eBook)
DOI 10.1007/978-3-319-20819-0

Library of Congress Control Number: 2015955263

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Créer, c'est recombiner.—François Jacob

As changes over time—evolution—may be called adaptation at the species level, we refer to somatic adaptations as those genetic changes occurring in an animal's cells during its lifetime. The latter phenomenon, with which this volume is concerned, deals with the genetic diversity generated in vegetative cells of individuals, whether parasite or host, enabling individualized responses for survival. This strategy allows members of the population to survive, procreate, or infect. And because the survival phenotypic constellation was somatically based, subsequent generations continue to take individual risks. Thus, it is not so much the resultant genetic diversity in itself but rather the mechanism generating this diversity that has been evolutionarily selected for.

The model systems here exemplify diversification pathways involving DNA, RNA, or posttranslational modifications and range from protozoan to mollusk to fishes to humans. In most cases, the diversifying mechanism, be it in host or in parasite, enables competing responses that escalate in scissors-rock-cloth fashion. The advantage of genetic diversification is rapid adjustment to novel (read: potentially adverse) circumstances, as in the switching of the variant surface glycoprotein coating a trypanosome during the host's mounting immune response, or as in the proliferation and expansion of vertebrate lymphocytes bearing immunoglobulin, T-cell receptor, or variable lymphocyte receptor (VLR) when activated in the course of an infection.

As respiration and nourishment have to be acquired from the external environment, complex interactions have originated at these interfaces; the epithelium of respiratory or filtering-feeding organs and the gut must be also be immunocompetent. The immune aspect of the vertebrate gut, with its associated lymphoid tissues and molecules, is represented here in bony fishes and in mammals, which have independently evolved specialized secretory immunoglobulins. Whereas V(D)J recombination provides fish lymphocytes with a variable region that directly splices to the IgT C region exons, in mammals it is additional somatic signaling that

induces class switch recombination to IgA, diversifying the immunoglobulin effector function. Junctions of contact with the environment may have been the primary sites where decisive selective advantages were exerted.

Just as the host systems evolved to recognize invaders, the invaders evolved to become evaders. However, there is yet another important facet: the host needs to distinguish beneficent microorganisms. Negotiation with gut microbiota is thus not unilaterally defensive but complex in ways we are only beginning to explore. The protochordate VCBP molecules constitute the only germline-based system presented in this volume. These highly diverse molecules have undergone selection through evolution, and their role in the gut involves not only immune function but also regulating microbiota homeostasis. And, intriguingly, stimulation of lymphocytes by the gut flora is not always a host protective response. As demonstrated in rabbit appendix, emerging B lymphocytes interact with gut commensals, and antibody repertoire expansion through gene conversion and somatic hypermutation is driven by the local superantigen-like molecules. The developing immune system within an individual is subjected to selection for tolerance versus immunity, self versus nonself, but the interaction between its immune system and microbiome is emerging as multilayered and, to a great but unknown extent, symbiotic.

If the CRISPR system in bacteria and archaea is included, then in most branches of the tree of life organisms have evolved somatic mechanisms in this "arms race" between host and parasite. As shown in models reviewed in this volume, when the diversification pathway involves combinatorial associations of elements, the gene organization tends toward minimalism. That is, in stark contrast to their diversified output, there are far fewer germline elements. Fewer components facilitate regulation of the system. Fewer building blocks also permit more stringent selection for efficacy; mutations in a germline component will have amplified effects. What we discover in gene systems to be simplicity and perceive as elegance, it is in Nature mere functionality.

Basel, Switzerland, 2015

Ellen Hsu
Louis Du Pasquier

Contents

Evolution of Immunity and Pathogens

Robert S. Jack

Abstract Host and pathogen engage in a constant evolutionary struggle known as a "Red Queen Paradigm". In this struggle, natural selection favours the pathogen which evolves effective virulence mechanisms and the host which is able to field adequate resistance strategies. A number of factors limit what each side can do. These include the fact that the elaboration of virulence or resistance mechanisms results in costs in genetic fitness and requires the use of ever more of the limited number of genes available in the genome. In addition, since the pathogen usually has a very much shorter generation time than the host, it can fix new virulence mutations much more quickly than the host can evolve matching resistance mechanisms. Finally, the host must ensure that its defence system does not result in unacceptable levels of collateral damage to its own tissues. This chapter briefly outlines how these considerations shape host–pathogen interactions.

1 The Red Queen

Alice stepped through the looking glass and emerged on the other side in a strange country. No sooner was she there than an energetic lady in a long red dress grabbed her by the hand and started to run as fast as she could. After a while Alice was out of breath and asked why they had to run so fast—especially as they didn't seem to be getting anywhere. The Red Queen looked at her in surprise and told her that it would take all the running she could do, to keep in the same place. If she wanted to get somewhere else, she'd have to run at least twice as fast (Carroll 1871). The idea that resources must be invested, not to get ahead but simply to maintain the status quo, was taken up by those interested in evolution to describe a situation very common in biology—The Red Queen Paradigm (RQP). Host–pathogen interactions are classical RQPs—arms races in which the pathogens are forever evolving new virulence strategies while the host responds with ever more effective resistance mechanisms.

R.S. Jack (✉)
University of Greifswald, Department of Immunology, Sauerbruchstrasse DZ 7, Greifswald D-17487, Germany
e-mail: bobjack@hotmail.de

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_1

The host's problem is to quickly recognise a potential pathogen, for only after the intruder has been detected can it be attacked with pore-forming proteins, hydrolytic enzymes, free radicals and phagocytosing cells. The pathogen, for its part, is under equally strong selective pressure to find a means of avoiding the host's defence system. In this struggle, fundamental constraints limit the range of possible responses of host and pathogen. First, the investment of resources in virulence or resistance mechanisms means that these resources can no longer be assigned to the production of progeny and this inevitably leads to a reduction in genetic fitness. Second, genomes do not contain unlimited numbers of genes, and therefore both virulence and defence must be organised with the minimum input of genetic information. Third, the severe asymmetry commonly found between the generation time of the host and that of the pathogen means that the pathogen can fix new resistance mechanisms very much faster than the host can fix complementary resistance strategies. This "generation gap" gives the pathogen a large head start. Finally, the host faces the problem of collateral damage: a pathogen must be rapidly detected and then attacked with all the means at the host's disposal, but, in the course of this battle, collateral damage to host tissue must be kept to the minimum.

1.1 Genetic Fitness

Fitness is a measure of an organism's ability to project its genes into subsequent generations. Resources invested in activities which do not support this central goal inevitably lead to a reduction of the individual's contribution to the genetic future. Because of this, both host and pathogen must be cautious in the way they organise their sides of the RQP—resources must not be squandered. For an obligate pathogen, the equation has an added level of complexity because it must ensure that its progeny can move to a new host on the death of the present one. If the host has a low population density, then a highly virulent pathogen may rapidly wipe out all locally available hosts and thus assure its own destruction. Each pathogen must therefore work out an optimal modus vivendi with its host. In many cases, the pathogen will find it advantageous to reduce its virulence, extend the life of the host, and thus increase the chance of further transmission (Boots and Mealor 2007). If pushed to the limit, this sort of development may lead to mutual advantage for both partners and the pathogen turns into a commensal. In other cases, the relationship may develop into a state of armed neutrality as is the case with *Staphylococcus aureus* which lives as a harmless symbiont in the human nose until a temporary reduction in the host's defences allows it to cross the epithelial barrier whereupon it may turn into a truly fearsome pathogen (Kluytmans et al. 1997).

The advantages to the pathogen of investment in a virulence strategy will be weighed by natural selection against the fitness costs which this diversion of resources entails. By the same token, successful hosts will have opted for a trade-off that maximises the fitness benefit achieved for the lowest investment of resources.

1.2 Gene Number

A second constraint which limits the genetic options open to both pathogen and host concerns the number of genes which are available to construct a virulence mechanism or an immune defence system. Humans, mice and indeed most other multicellular animals only have around 20,000 genes. This is not a lot of genetic information with which to programme the development and functioning of a complex multicellular organism and consequently few genes can be allocated solely to building an immune defence system. The eukaryotic parasites discussed in this volume all have substantially fewer than 20,000 genes (Berriman et al. 2005, 2009; Gardner et al. 2002) and thus have a problem in genetically funding the virulence mechanisms they require. On both sides of the battlefront, ingenious solutions to the gene number problem have emerged. Three groups of solutions are commonly found.

1.2.1 Making the Best Use of a Few Genes

Since the number of potential pathogens is limitless and the number of genes in the host genome is restricted, it follows that immune defence cannot work by assigning a unique germ line encoded gene to counter each different pathogen. Instead, the immune systems of the vast majority of animals function by detecting pathogens with the help of a small number of hardwired receptors which have been selected to be directed against essential components of a wide range of microorganisms. In this way, the innate immune system is able to detect many different pathogens with just a small number of sensor molecules. But even this is not parsimonious enough and many of these receptors justify their existence by being involved in more than just pathogen detection.

When single-cell eukaryotes learned to come together and benefit from the advantages of life as a multicellular organism, they had to move from being individuals at the single-cell level to becoming individuals at a higher, multicellular level (Michod 2007). To do this, they had to develop the capacity to recognise "self". At the simplest level, cell adhesion molecules provide the component cells with the ability to interact with other "self" cells and these molecules help define the place of the various cell lineages in the multicellular structure. This sort of thing may be sufficient in a simple animal like *Hydra*, but as the complexity of the organism increases there arises a requirement for constant quality control of the tissues and for a means to ensure tissue homeostasis. Supernumerary cells must be quietly removed by apoptosis; cell debris resulting from sterile trauma must be cleared away; transformed cells should be identified and destroyed, and infections must be quickly detected and repulsed. All of this is the job of the innate immune system. Armed with just a handful of dedicated genes, the sentinel cells of innate immunity—granulocytes, macrophages, dendritic cells and innate lymphoid cells—must quickly detect "apoptotic self", "damaged self" or "altered self", and

be able to determine whether the problem was caused by sterile trauma, by transformation or by infection so that appropriate action can be taken.

1.2.2 Germ line Encoded Genes of the Innate Immune System

The innate immune system is often described as being "non-specific", but this is a serious misnomer for the system's receptors are typically of very high affinity and specificity. Some of these receptors, like the Toll-like Receptor 5 (TLR-5), which is directed against bacterial flagellin, bind (as far as we know) only to a microbial structure (Yoon et al. 2012). Others have been selected to bind both to microbe associated ligands and to ligands which signal "damaged self". Examples of these are TLR4 which, in association with MD2, binds bacterial lipopolysaccharide (Park et al. 2009), but in association with CD36 and TLR6, it binds the oxidised form of LDL which is an "altered self" danger signal (Stewart et al. 2010). Likewise, the C-type lectin Mincle has two independent binding sites one of which binds carbohydrate ligands associated with mycobacteria and fungi (Wells et al. 2008), while the other binds the SAP-130 ribonucleoprotein that is released from necrotic cells (Yamasaki et al. 2008). In a similar vein, TLR2 may be triggered not only by microbial lipoproteins but also by the pyrroles formed as breakdown products of the membranes of necrotic cells (Jin et al. 2007; West et al. 2010).

There are a number of families of innate receptors including Toll-like Receptors, C-Type Lectin Receptors, NOD-like Receptors, Natural Killer Cell Receptors and scavenger receptors. Each sentinel cell of the innate immune system expresses a number of different innate receptors so that a disturbance of homeostasis may lead to the engagement of several of them. Each of these receptors will then initiate its respective signal transduction pathway and, since these pathways engage in a considerable amount of crosstalk (Sancho and Reis e Sousa 2013), the sentinel cell's response will not be stereotyped but rather form a balanced retort to the array of inputs emanating from the zone of disturbed homeostasis. Apoptotic cells, for example, have the capacity to "silence" both macrophages (Henson and Bratton 2013) and dendritic cells (Steinman et al. 2000) so that, in general, the immune response to them is restricted to phagocytosis. The importance of this can be seen in the fact that in the bone marrow of an adult human, some 10^{11} apoptotic neutrophils are silently removed in this way every day. In contrast, microorganisms invading a sterile tissue provoke a pro-inflammatory activation of the sentinel cells which then release inflammatory mediators that diffuse out to the neighbouring capillaries and in turn activate the endothelial cells lining the luminal surface of the vessels. These activated endothelial cells promote the recruitment of neutrophils which enter the infected tissue and destroy invading bacteria. Activated neutrophils are enormously violent cells and if left to themselves can destroy not only the invading pathogen but host tissue as well. Nature has quaintly chosen to solve this problem by giving activated neutrophils a very short half-life (Amulic et al. 2012). Because of this, an inflammatory site quickly fills up with neutrophil corpses which must be removed by inflammatory macrophages (Nguyen et al. 2012).

Neither the silently phagocytic response to apoptotic cells nor the inflammatory responses to microorganisms are simple automatic reflexes, but rather each can be modulated to suit the precise nature of a problem. One example of this involves the C-type lectin DNGR-1 which binds F-actin released from dead cells. This interaction leads neither to inflammatory activation of sentinel cells nor is it necessary for phagocytosis. Instead, it leads to a change in the endosome maturation pathway so that a non-degradative fate of the phagocytosed material is favoured. This is particularly important in the clearance of virus infected or transformed cells whose contents may thus be cross-presented to MHC class 1 molecules and hence lead to the recruitment of $CD8^+$ T-cells (Sancho and Reis e Sousa 2013). Thus, the pattern of innate receptors engaged will have far-reaching consequences on how the ensuing immune response is shaped.

1.2.3 Germ line Encoded Virulence Genes of Pathogens

Obligate pathogens too face restrictions in the number of genes which can be devoted to elaborating virulence strategies and yet the effectiveness and sophistication of the immune system may force them into making expensive compromises. One example of this is seen in *Plasmodium falciparum*, the causative agent of malaria. This parasite infects erythrocytes and then arranges for the expression on the erythrocyte surface of a molecule known as *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) which mediates attachment to the vascular endothelium and by doing so helps the infected red blood cell avoid destruction by the erythrocyte quality control system in the spleen. One PfEMP1 gene is, however, not sufficient because the protein expressed on the erythrocyte surface is readily recognised by the adaptive immune system and antibodies to it are synthesised. For this reason, the parasite has a family of around 60 pseudo alleles coding for PfEMP1 variants. The choice of which gene is to be expressed is made at the transcriptional level and at any one time only one PfEMP1 gene is active. However, the pathogen can switch the transcriptional apparatus to the expression of a pseudo allele encoding a different and immunologically non-cross-reacting PfEMP1 protein, and can do so at the astonishingly high rate of 1–2 % per cell generation even in the absence of immune selection (Roberts et al. 1992). In this way, although the host does mount an effective immune response, the pathogen uses its pool of PfEMP1 genes to constantly flee into the immunological future (see chapter "Antigenic Variation in *Plasmodium falciparum*").

1.2.4 Polymorphism: Distributing Variants Throughout the Population

Both for host and pathogen, a minimalistic genetic strategy soon reaches its numerical limits and other means of extracting the maximum benefit from the minimal number of genes must be found. One widely used solution is to make use of the fact that natural selection operates on populations. A small number of

genes in an individual genome can be effectively expanded by having a large number of polymorphic alleles spread through the population. Thus, an individual *P. falciparum* cell has around 60 PfEMP1 pseudo alleles, but there are numerous polymorphic variants of these genes. Because of this, even extensive host immunity to any given parasite clone will not necessarily provide protection against subsequent infection with a new clone which expresses a different set of PfEMP1 alleles.

The same reliance on polymorphisms can be seen on the host side of the RQP in the MHC class I and Class II genes whose protein products are required to present antigens to T-cells. Each individual human will express three to six of these molecules and each class I or class II molecule will bind a limited range of antigen peptides. Because of this, it is readily conceivable that a pathogen might arise whose proteins do not yield peptides which can be recognised by any of the MHC molecules in a given individual. However, since there are hundreds of alleles of the class I and of class II genes in the population, it is most unlikely that a pathogen will be able to evolve a proteome which is invisible to all of the polymorphic MHC molecules. The chance that a pathogen might manage to wipe out the whole of humanity with this sort of molecular "lucky punch" is therefore not very high.

1.2.5 Discrete Cassette Systems to Build Variant Gene Products

Systems in which the information for the crucial functional parts of a molecule is recombined as cassettes into the region coding for the molecule's structural backbone provide a further means of obtaining a large number of different products using a small amount of genetic space. One well-investigated example of this sort of thing is given by the DSCAM gene in arthropods which is described in detail in the chapter "Somatic and germline diversification of a putative immunoreceptor within one phylum: Dscam in arthropods". In *Drosophila melanogaster*, each DSCAM molecule is made of 24 exons, but the gene is transcribed as a messenger RNA in which there are 12 alternatives forms of exon 4, 48 alternative forms of exon 6, 33 alternative copies of exon 9 and two alternative copies of the trans-membrane exon 17. Exons 4, 6 and 9 contribute to the formation of that part of the extracellular domain which determines the molecule's binding specificity and there are therefore $12 \times 48 \times 33 = 19{,}008$ different DSCAM binding isoforms. While these DSCAM isoforms are involved in wiring the Drosophila nervous system (Hattori et al. 2009), they are also expressed in the principal immune tissues of the fly—the fat body and hemocytes. Certain DSCAM isoforms can bind to bacteria and by doing so stimulate their phagocytosis by hemocytes (Watson et al. 2005). Interestingly, this sort of DSCAM variability due to alternative splicing is not seen in mammals.

Generating molecular diversity by recombination of discrete coding units need not be restricted to RNA splicing but can also be carried out at the DNA level. For example, the functional properties of immunoglobulins may be modulated by an Activation Induced Deaminase (AID) driven DNA recombination event which results in joining the region of the gene coding for the N-terminal antigen binding part of an immunoglobulin to one of several different C-terminal gene segments. In

a similar way, *Trypanosoma brucei* uses gene conversion to move a silent copy of the Variant Surface Glycoprotein (VSG) into an expression site during the early phases of an infection (Hall et al. 2013) (see chapter "A Host–Pathogen Interaction Reduced to First Principles: Antigenic Variation in *T. brucei*").

It is important to note that whether this sort of recombination event takes place at the RNA or DNA level, one discrete piece of information is being substituted by another, and both of these pieces of information are present in the germ line and have therefore been scrutinised by natural selection. A very different situation arises with the Red Queen's current trump—"indiscrete" recombination in which genetic information is somatically mixed and matched to provide a vast repertoire of products. These systems will be considered separately later.

1.3 The "Generation Gap"

A successful virulence or resistance strategy must, in general, be passed on to the next generation through the germ line. This makes the rate of fixation of advantageous changes dependent on the generation time of the organism. Humans have a generation time which is currently somewhat greater than 20 years. Viruses and bacteria may have generation times of 20 min. The difference between 20 min and 20 years is a factor of somewhat more than 500,000 which means that by this criterion prokaryotic pathogens may be able to fix virulence strategies around a half a million times faster than we can respond with hardwired resistance mechanisms. Eukaryotic pathogens have considerably longer doubling times than do many viruses or bacteria, but even so *T. brucei*—with an in vivo doubling time of 5.3 h—has a considerable advantage over its human host (Haanstra et al. 2012).

Given this substantial evolutionary life cycle advantage for the pathogens, one might reasonably ask how metazoan eukaryotes manage to survive at all. The answer lies—at least in part—in a curious property of Natural Selection. Natural Selection is a race. The chance that an individual wins a race is partly dependent on that person's sporting capabilities and partly on the nature of the competition. An average person would have a reasonable chance of coming in first, second or third in a race against ten randomly chosen members of humanity. That average person would stand no chance whatever in a race against the entire human population. This illustrates the simple fact that the sorting power of Natural Selection is directly proportional to the effective population size—that is to say to the number of individuals in the population producing progeny. The effective population size of *Escherichia coli* is equivalent to the total population, and since, at the end of the day, all *E. coli* on the planet are competing with each other, this population is astronomical in size. Given this huge population, competition is fierce and natural selection will ensure that only the fittest *E. coli* survive for long. For a multicellular eukaryote however the situation is different since the total population is, by comparison to that of *E. coli*, miniscule, and, since only a tiny fraction of them are involved in producing progeny, the effective population is orders of magnitude

smaller than the total population (Lynch 2007). Selection pressure is therefore relaxed and as a result multicellular eukaryotes accumulate mildly disadvantageous mutations by genetic drift to an extent that prokaryotes or single-celled eukaryotes do not. This can be dramatically illustrated in two quite different ways. First, in a series of elegant experiments Lindquist and her collaborators have demonstrated that a large number of known mutant phenotypes in *D. melanogaster* can be phenocopied by interfering with the function of the chaperone Hsp 90. This suggests that many gene products in higher eukaryotes are literally falling apart and require the chaperone to hold them together. Many of these variants may be far from perfect at doing their assigned job, but, by the same token, they may contribute to evolvability by providing the start of an answer to new evolutionary challenges (Jarosz and Lindquist 2010; Rutherford and Lindquist 1998). This accumulation of disadvantageous mutations has recently been directly demonstrated from genomic data. Using highly restrictive criteria to search for Loss-of-Function (LoF) alleles, a total of 1285 such LoF alleles were identified in the first 185 genomes of healthy humans from the 1000 human genome project. On average, each genome carried 253 LoF alleles in heterozygous form plus an astonishing 20 LoF alleles in homozygous form (MacArthur et al. 2012).

1.4 Collateral Damage

One final genetic constraint affects the elaboration of an immune defence system and that is the requirement that the receptors be able to distinguish accurately between "self" and "non-self". For many years, it was thought that the innate immune system solves this problem by the selection of receptors directed against structures like LPS or flagellin which are present exclusively on certain prokaryotes. Recently, however, it has become clear that innate immunity plays a broader role in detecting departures from normal homeostasis and the fundamental ability to recognise self structures that this brings with it certainly raises the possibility that innate immune receptors might contribute to the induction of autoimmune disease. However, the number of receptors involved is limited and they have been subjected to a long period of natural selection during which potentially dangerous ones would have been eliminated from the population. Nevertheless, as outlined above, these receptors will in many cases not act singly but rather in combinations in response to changes in homeostasis and this brings with it the danger that combinations of rare alleles at these loci may persist and then, under certain circumstances, provoke pathological responses. However, the bottom line is that autoimmunity is clearly not a major penalty for the possession of an innate immune system. The situation is dramatically different for the adaptive immune systems outlined below.

2 The Red Queen's Current Trump: Mosaic and "Sloppy" Recombination Systems

The systems we have looked at so far provide both pathogen and host a wide range of options in their respective sides of an RQP. By using carefully screened genes, by expanding the sets of polymorphic alleles in the population or by utilising discrete recombination systems, both host and pathogen can extract the maximum benefit from a restricted amount of genetic space. However, a large increase in the number of virulence or resistance options available to pathogen or host can be achieved by using recombination systems which are "indiscrete". Here a large repertoire of products can be built by somatically generating hybrid coding sequences which are not present in the organism's germ line.

2.1 Indiscrete Recombination

One example of an indiscrete recombination system is to be seen in the *T. brucei* Variable Surface Glycoprotein (VSG) system (see chapter "A Host–Pathogen Interaction Reduced to First Principles: Antigenic Variation in *T. brucei*"). The trypanosome expresses a huge number of copies of a VSG on the surface and these immunogenic molecules effectively shield the other surface components of the parasite from the immune system. The *T. brucei* genome contains thousands of copies of VSG genes both on its conventional chromosomes and on the roughly 100 mini chromosomes. The functional members of this collection can be moved intact into an expression site, but only about 5 % of the VSG gene copies are intact—the rest are pseudogenes. Late in infection the VSG copy in the expression site is frequently found to be a mosaic formed by a recombination process that stitches together DNA fragments derived from several of these pseudogenes. In this way, the pathogen has, in principle, access to a massive—almost limitless—number of possible VSG structures. In striking contrast to the discrete cassette systems, the mosaic VSG genes so formed are not encoded in the parasite's germ line and have therefore never been screened by natural selection (Hall et al. 2013). Some of this vast array of potential VSG structures will be nonfunctional for one reason or another, but this is not a major problem to the parasite: individual cells expressing mosaic genes which give rise to inappropriate VSG structures will simply be eliminated in the host while the remaining cells have access to so many VSGs that the host immune system will be unable to cope.

A rather similar situation—but with very different consequences—is seen on the other side of the RQP in the adaptive immune systems based on indiscrete recombination which so far have only been rigorously established in vertebrates. In cyclostomes—the jawless lamprey and hagfish—a potential repertoire of at least 10^{14} different Variable Lymphocyte Receptors (VLRs) is formed by using somatic gene conversion to insert fragments derived from a limited number of germ line

encoded Leucine-Rich Repeat (LRR) sequence elements into an expression site (Nagawa et al. 2007; Pancer et al. 2004). The resulting receptors are therefore not encoded in the germ line and have never been examined by natural selection (see chapter "Variable lymphocyte Receptors: A Current View"). In jawed vertebrates, equivalently large receptor repertoires are produced by somatic recombination of the germ line V, D and J segments which code for the variable regions of the B-cell receptor and of the T-cell receptor. Here the initial recombination event in V gene rearrangement is carried out by the RAG recombinase. RAG recombinase is derived from a site-specific transposase and would therefore be expected to be best suited to a cassette-type form of discrete recombination which would yield only a modest repertoire of receptors (Kapitonov and Jurka 2005). However, the site-specific recombination process has been rendered sloppy by introducing a hairpin formation step and by the deliberate addition of un-templated random nucleotides by Terminal-deoxynucleotidyl-Transferase (TdT). Since the sequences at the RAG recombination termini code for the antigen recognition part of the receptors, a vast repertoire of different antigen specificities is formed.

The majority of the products of this sloppy recombination process are nonfunctional because of frameshifts induced during hairpin opening or by the action of TdT or because the recombination site gives rise to a structurally inappropriate protein product. Lymphocytes expressing such rearrangements are detected by a quality control system and discarded by apoptosis. This makes the generation of the lymphocyte population a very costly and wasteful procedure which can only be of selective value to an animal able to exploit the repertoire to the full. Since a naïve repertoire formed in this way is vastly greater than the total number of lymphocytes in the body, investing in such a system would be of little advantage to a small, short-lived animal with few immune cells, for it would never be able to sample more than a tiny fraction of the repertoire in its own lifetime. Even if the population of such an organism was large, selection of those members expressing the advantageous receptor at the appropriate time would not constitute an evolutionary advantage because the useful receptor is not germ line encoded. The consequence is that the vast repertoire formed by indiscrete recombination will only be of significant survival value to a long-lived species with a large number of short-lived lymphocytes. For such an animal, there are substantial advantages—but also drawbacks—to the generation of the huge repertoire.

The first advantage of indiscrete recombination is obvious: an almost limitless V-gene repertoire can be built from just a handful of V, D and J gene elements which occupy only a tiny amount of genetic space. In this way, the gene number problem is solved.

The second advantage—and it is huge—is that because of the nature of the rearrangement process, every lymphocyte formed is, by definition, a mutant. The sequences coding for the all-important antigen binding site are not present in the germ line and hence classical natural selection is not involved in sorting out the best products. Instead, currently useful mutant receptors are selected from the pool of lymphocytes by clonal selection. The result is that mutant selection and fixation has been moved from the level of the organism's germ line to the level of the individual

lymphocyte. Because of this, the rate of mutational fixation is now determined by the rate of lymphocyte production—and millions of lymphocytes are produced each day. By using this elegant trick, the rate of evolution of immune receptors is now matched to the rate of evolution of new pathogen virulence strategies (Berek and Milstein 1988), and the generation gap problem is effectively solved. The corollary of this is, however, that since the advantageous receptors are generated as somatic mutants, they cannot be passed on to the next generation through the germ line. What the next generation gets is the germ line collection of V, D and J elements plus the sloppy RAG recombination system—beyond that it must then look out for itself.

A third consequence of indiscrete recombination is that a rare somatic mutation leading to the formation of a receptor able to provide protection against a pathogen will be lost when the clone of lymphocytes producing it ages and dies, leaving the host subject to re-infection with the same pathogen. Since each lymphocyte expresses only one receptor specificity, and since the number of specificities in the naïve repertoire is substantially greater than the number of lymphocytes in the body, the danger is that on re-infection no suitable receptor will be available in the pool of circulating lymphocytes. This gap in adaptive defence has been covered by the setting aside of some of the clonally selected cells as immunological memory.

Finally, there is the downside to indiscrete recombination in immune defence. The vast repertoire which is generated can "see" in principle every conceivable antigen and a receptor repertoire which "sees" everything "sees" far too much. Many of the receptors formed will interact with "self" antigens. Indeed, it has been demonstrated that the majority of well-formed antibody structures in the mouse interact with self molecules and the cells expressing them must be driven into apoptosis to avoid autoimmunity (Wardemann et al. 2003). This problem of autoimmunity is something which will inevitably result from any indiscrete recombination system. A fascinating central question to which we have at the moment no satisfactory answer thus centres on how such indiscrete systems could arise without leading to lethal autoimmunity.

2.2 The Emergence of Indiscrete Recombination Systems

Ten years ago immunologists believed that the initial emergence of indiscrete recombination in immunological defence was based uniquely on the chance appearance of RAG mediated recombination and that this was such an unlikely event that it could only have happened once. The discovery of a second apparently completely independent system of adaptive immunity based on indiscrete DNA-level recombination in cyclostomes (lamprey and hagfish) therefore came as a large surprise (Pancer et al. 2004). However, with the benefit of hindsight, two points are worth noting. The first is that DNA repair and recombination systems are present in every nucleated cell so that this machinery is available, and could in principle be easily hijacked, to contribute to immune defence. That this is indeed possible is shown by the fact that even bacteria have found ways to use DNA recombination for immune

defence in the form of the CRISPR system with which they protect themselves against phage attack. The second important point is that there is nothing special about the Ig protein scaffold, which is used as the RAG recombination target in jawed vertebrates (gnathostomes), that makes it predestined for building a vast antigen-specific repertoire. Other protein domains including the LRR domain and the ankyrin domain can also serve the same purpose (Kummer et al. 2012).

The crucial feature permitting the positive selection of indiscrete adaptive immunity may have had less to do with the nuts and bolts of the recombination system and more with the emergence of long-lived animals possessing large numbers of circulating immunocytes. Only such creatures can exploit a large repertoire and thus provide the evolutionary niche in which adaptive immunity can contribute to genetic fitness and hence becomes selectable. The emergence of adaptive immunity in such animals becomes—if not inevitable—then at least unsurprising.

In this context, the fundamental importance to immunity of circulating cells with the basic properties of lymphocytes can be seen in the astonishing results which emerged from the comparison of the indiscrete recombination systems of mammals and of lampreys. These two systems are unquestionably separate "inventions" because they use different protein scaffolds to build the repertoires—Ig domains in gnathostomes and LRR domains in cyclostomes—and different primary recombination enzymes—RAG in gnathostomes and cytidine deaminase in cyclostomes. Nevertheless, there are astonishing similarities in the pattern of genes expressed in gnathostome and cyclostome circulating lymphocytes. Before turning to the results, it is worth emphasising that the expression in lamprey lymphocytes of a gene also expressed in mouse lymphocytes begs us to draw the conclusion that the gene has the same function in the two organisms. That may be so, but since genetic manipulation of the lamprey or hagfish is not currently possible, this line of argument, though certainly intriguing, is not compelling. However, the persuasiveness of the argument is increased by every additional gnathostome lymphocyte feature which is mirrored in cyclostome—and there are many. The first is that in gnathostome lymphocytes come in three major flavours—B-cells, αβT-cells and γδT-cells. The B-cells synthesise receptors which recognise intact antigen and these receptors can be released in the form of soluble antibody. The T-cell lineages also express antigen-specific receptors, but these do not interact with intact antigen and they are not released in soluble form. Astonishingly the lamprey VLR-A and VLR-C lymphocytes resemble mammalian T-cells in these features and they are preferentially activated by a classical T-cell mitogen. The lamprey VLR-B lymphocytes, in contrast, mirror mouse B-cells in that they synthesise receptors with specificity for intact antigen and these receptors can be released in soluble form after activation (Han et al. 2008). But the similarities do not end there: mammalian lymphocyte populations express cytokines and cytokine receptors through which they can communicate with each other and the lamprey lymphocytes express orthologs of some of these molecules. This suggests that all vertebrate lymphocytes share to a remarkable extent this inter-lymphocyte communication system and hence that this is a property which was present in the last common ancestor of

cyclostomes and gnathostomes (Guo et al. 2009). Furthermore, the lamprey lymphocyte populations express transcription factors which are reminiscent of the expression pattern in mouse lymphocyte populations. For example, Pax 5 is the master regulator transcription factor for B-cell development in the mouse and a Pax 5 ortholog is expressed in the lamprey VLR-B lineage but not in the T-cell-like VLR-A or VLR-C lineages (Hirano et al. 2013).

All of this suggests that these three basic lymphocyte populations were established before the divergence of cyclostomes and gnathostomes. What selective advantage such presumptive pre-lymphocytes might have had and what antigen specific receptors, if any, they expressed remain matters of speculation, but the identification of lymphocyte-like cells in amphioxus gives hope that these questions may not be forever intractable (Huang et al. 2007).

2.3 The Requirement for Repertoire Purging

It is reasonable to suppose that the indiscrete recombination systems evolved from simpler discrete systems. Indeed in the case of the Gnathostomes, there is indirect evidence that this is the case because the rearrangement of certain germ line V genes has been shown to provide protection against common bacterial infections. In particular, rearrangement of the mouse $V_{H\ S107.1}$—$D_{FL16.1}$—J_{H1} heavy chain together with a Vk_{22}—Jk_5 light chain yields a "natural antibody with specificity for phosphorylcholine. This antibody provides protection from *Streptococcus pneumoniae* (Briles et al. 1982) and at the same time recognises the oxidised phospholipids expressed on apoptotic cells and in artherosclerotic plaques (Shaw et al. 2000). This suggests that these sequences, like many of the receptors of the innate immune system, may have been naturally selected to recognise both a common bacterial antigen and an endogenous danger signal.

However, the transformation of a discrete recombination system to an indiscrete one is anything but trivial because the indiscrete systems bring with them enormous problems in terms of autoimmunity. How is the autoimmune potential of an indiscrete recombination system to be suppressed? It is clear that there is no simple straightforward solution to this problem, but in mice and humans central T-cell tolerance and peripheral mechanisms like anergy, ignorance and T-regulator cells are all essential elements. The great evolutionary mystery is: how did they arise—and when? An indiscrete recombination system automatically produces an enormous naïve repertoire and this repertoire will be lethal in the absence of a system to rigorously purge it of autoimmune receptors. The naive repertoire therefore cannot evolve in the absence of the purging machine. Yet, on the other hand, a repertoire purging system would seem to have no selective value prior to the appearance of the "indiscretely" formed repertoire and since evolution cannot store up currently useless things for long, the essential tolerance mechanisms must be "invented" at the same time as the indiscrete recombination system. How this was achieved is for

the moment a great mystery. Unravelling that mystery will certainly lead to a large advance in our understanding of the evolution of immunity.

3 The Evolutionary Trajectory of Immunity

Immune defence is a fundamental attribute of living systems, yet, though it extends back to the dawn of life, our knowledge of the evolution of immunity is patchy at best. The reason for this is very simple: the immune system has been studied in detail in a small number of vertebrates, and in flies while, on the other hand, little is known about immune defence in most other lower animals.

3.1 Genomic Analysis and Lower Eukaryote Immune Systems

The flood of genomic sequence data which has become available in recent years has transformed efforts to trace the evolution of immunity across phylogeny. The genomic approach has been an enormous stimulus to the field, though it is important to understand that it has its limitations. The principal drawback is that the presumptive immune function of a gene in a lower eukaryote is often ascribed merely on the basis of its sequence homology to a bone fide immune defence gene in vertebrates. This is of course a valid first step, and yet it has its pitfalls. For example, the Toll-like Receptors (TLRs) in mammals act as so-called "Pattern Recognition Receptors". They bind to their microbial ligands and by doing so initiate an immune response. However, of the nine TLRs in Drosophila, none binds a microbial ligand and only one—Toll 1—is involved in immune defence (Narbonne-Reveau et al. 2011). In the light of this, the immunological significance of the presence of more than 200 genes coding for TLR like molecules in the sea urchin must remain a matter of speculation.

Sequence homology to immunologically important molecules in vertebrates, ligand binding in in vitro assays, expression in immunocytes and upregulation on contact with pathogens are all valuable hints of an involvement in immune defence—but they do not constitute final proof. This makes the assessment of immune functions difficult, particularly in basal chordates like the urochordate Ciona or the cephalochordate Amphioxus for which "knock out" and "knock down" technologies do not currently exist (Sasaki et al. 2009). In short, the definitive demonstration of the involvement of a molecule in immune responses in lower eukaryotes is very often a technically hard nut to crack. It is, however, not impossible as has been demonstrated for DSCAM in Drosophila (Watson et al. 2005) and for Fibrinogen Domain Immunolectins in Anopheles mosquitoes (Dong and Dimopoulos 2009; Hanington and Zhang 2011). Thus, despite the

genome sequencing revolution, our knowledge of immunity in most invertebrates is sketchy at best and, because of this lack of detailed functional data, any discussion of immune system evolution at the present moment remains somewhat speculative. It requires that one seeks a reasonable middle way between Baconian fundamentalism on the one hand and an "anything goes" approach on the other.

3.2 Inheritance by Descent Versus Intercalary Evolution

Few restrictions are placed on the evolutionary trajectories open to pathogens, and since different hosts have faced very different challenges, it is perhaps not surprising that immune defence viewed across phylogeny is a curious mix of highly conserved elements, such as phagocytosis or the use of agglutinins, together with wildly different group-specific responses to pathogen attack. Examples of the latter might be the DSCAM experiment in Arthropods or the LRR domain-based indiscrete repertoire of immune receptors in cyclostomes. How are we to explain the evolution of such apparently different immune system architectures across phylogeny? To approach this question, it is perhaps helpful to look at a rather similar situation for which more data are available—the evolution of vision. A visual system allowing an animal to orient itself with respect to a light source would bring a large fitness benefit with it. However, the probability that such a primitive visual system would be brought about by random mutation is very small and hence vision would not be expected to have been "invented" many times. Nevertheless, once this initial unlikely event had taken place, inheritance by descent, coupled with selection for ever more acute visual systems, would be expected to lead step by step to a gradual increase in complexity as one followed the development of eyes through phylogeny. This expectation, however, is not met.

In triploblastic animals—those with three germ layers—eyes come in all sorts of shapes and sizes and they pop up and disappear with astonishing rapidity across phylogeny. True, down near the bottom of triploblastic life, the lowly flat worm *Polycelis auricularia* does have simple eye spots consisting of just one photoreceptor cell and a pigment cell while mice and men boast complex camera-type eyes. But there is no gradual step-by-step accumulation of complexity as one moves from flat worms to mammals—no slow improvement of a basic morphological design by evolutionary "tinkering" (Jacob 1977). The insect compound eye has a completely different morphological structure from the camera-type eye of a mouse or the mirror eye of a scallop; the cephalopod camera eye has a different developmental origin from the eye of man. Worse still is the case of the adult jelly fish *Tripedelia cystophora* which as a diploblast lies phylogenetically way below the flat worms and yet nevertheless sports a complex camera-type eye (Kozmik et al. 2003). An attempt to make sense of all this led to the conclusion that visual systems had been "invented" 40–60 times over the course of phylogeny (Salvini-Plawen and Mayr 1977). That conclusion accurately reflects what can be seen at the morphological level and yet the idea that a vanishingly unlikely event could be generated 40–60

times left this as one of the great apparent paradoxes in biology. A second equally astonishing evolutionary paradox of the same sort emerged much more recently with the discovery that the inherently unlikely generation of a vast repertoire of immune receptors by an indiscrete recombination system had been "invented" twice within a short period (Pancer et al. 2004).

The resolution of the visual paradox has come largely from the work of Gehring and his associates at the Biozentrum in Basel who noted that loss-of-function alleles of the transcription factor Pax 6 resulted in disruption of eye development in man, mouse and Drosophila. Astonishingly, transfer of the mouse Pax 6 gene into the developing leg anlagen of Drosophila led to the production of an ectopic eye on the fly's leg. The ability of the mouse gene to direct eye development in the fruit fly was remarkable in itself—the last common ancestor of mouse and fly existed some 500 million years ago—but what was even more astonishing was that the mouse Pax 6 gene specifies a camera-type eye in the mouse and an insect's compound eye in the fly. The explanation for this extraordinary state of affairs, and the resolution of the paradox of multiple random "inventions" of the eye, was that the unlikely initial event in eye development was the specification of a photoreceptor cell under the control of the transcription factor Pax 6. This was the crucial initial step in establishing visual systems. All else flows from this monophyletic "big bang". The master regulator Pax 6 controls the expression of downstream transcription factors some of which are also highly conserved in eye development across phylogeny. In such a transcription factor network, downstream effector genes can be recruited by mutational insertion of the appropriate transcription factor binding sites into their enhancer elements. Since mutation is random, different organisms could recruit different downstream genes and hence end up with morphologically very distinct eye structures. Pax 6 and the photoreceptor cell are thus the two poles of a visual system and between them lies the assortment of effector genes which the transcription factor cascade has recruited. In the case of the Drosophila larva, some 370 - eye-specific genes have been recruited into the network while if pupal and adult expression is included the number extends to around 1000 (Gehring 2012; Michaut et al. 2003). This form of evolutionary trajectory involving the recruitment of effector genes to a transcription factor cascade has been termed "intercalary evolution" (Gehring and Ikeo 1999).

If this example is to help us visualise the evolution of immune defence, then we must first define what the irreducible minimum properties of an immune defence system might be. To do this, it may be helpful to look at immunity in the cellular slime mould *Dictyostelium discoideum*, which is a primitive "conditional" metazoan. *D. discoideum* lives as a free-living single-cell phagocyte until its bacterial prey becomes scarce, at which point thousands of scattered individual cells coalesce to form a so-called "slug" that moves around searching for food and, if it doesn't find any, then it adheres to the substrate and forms a stem on top of which a meiotic fruiting body produces spores. A specialised set of cells—around 1 % of the total—have the job of patrolling through the slug and ingesting any non-self particles such as microorganisms or debris which they find. Their ability to do this is dependent on the presence of a Toll/Interleukin-1 Receptor Domain (TIR-A) containing gene

which is also required at the single-cell stage for successful foraging. The remaining 99 % of the cells in the slug have a much reduced capacity to carry out phagocytosis (Chen et al. 2007). The immune system of the slime mould thus consists of mobile cells equipped with the capacity to distinguish self from non-self.

From this simple starting point, intercalary evolution provides a model for the evolution of immune defence in the sense that it explains the curious mix of retained components together with the sudden appearance of group-specific novelties. On the one hand, complex multicellular eukaryotes all use mobile cells able to distinguish self from non-self as the heart of their immune system. Nevertheless, one repeatedly sees that what is sauce for the immunological goose is by no means sauce for the immunological gander: Toll receptors in Drosophila are a very different kettle of fish from Toll receptors in mice or man; arthropods exploit DSCAM in ways completely different from what happens in mammals. Basic features of immunity are retained throughout phylogeny, but, just as in the evolution of visual systems, many of the details will vary from species to species depending on their history of pathogen interactions and because there is usually more than one way to solve a problem.

4 Conclusions

Each host–pathogen pair faces a unique set of challenges and evolves a unique set of virulence and resistance mechanisms. Our current understanding of the Red Queen paradigm of infection biology comes largely from studies of the immune systems in humans, mice and a few other economically important species. In contrast, immune defence in lower eukaryotes has received much less attention. However, as the age of antibiotics now gradually comes to an end, it will be increasingly important to understand the dynamics of infection biology in a much broader way. Since many aspects of immune defence are retained throughout phylogeny, studying the evolution of successful pathogen virulence mechanisms highlights the potential weak spots in the architecture of the host's immune system. On the other hand, studying the resistance mechanisms which other species have evolved to counter their pathogens may identify novel pathways and molecules and by doing so suggest intervention strategies with which to tip the balance in an infection in the host's favour.

References

Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A (2012) Neutrophil function: from mechanisms to disease. Annu Rev Immunol 30:459–489

Berek C, Milstein C (1988) The dynamic nature of the antibody repertoire. Immunol Rev 105:5–26

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B et al (2005) The genome of the African trypanosome Trypanosoma brucei. Science 309:416–422

Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama ST, Al-Lazikani B, Andrade LF, Ashton PD et al (2009) The genome of the blood fluke Schistosoma mansoni. Nature 460:352–358

Boots M, Mealor M (2007) Local Interactions Select for Lower Pathogen Infectivity. Science 315:1284–1286

Briles DE, Forman C, Hudak S, Claflin JL (1982) Anti-phosphorylcholine antibodies of the T15 idiotype are optimally protective against Streptococcus pneumoniae. J Exp Med 156:1177–1185

The Penguin Complete Lewis Carroll. Penguin Books, ISBN 014 00.9004 5, p 152

Chen G, Zhuchenko O, Kuspa A (2007) Immune-like phagocyte activity in the social amoeba. Science 317:678–681

Dong Y, Dimopoulos G (2009) *Anopheles* fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. J Biol Chem 284:9835–9844

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S et al (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419:498–511

Gehring WJ (2012) The evolution of vision. WIREs Dev Biol 3:1–40

Gehring WJ, Ikeo K (1999) Pax 6: mastering eye morphogenesis and eye evolution. Trends Genet 15:371–377

Guo P, Hirano M, Herrin BR, Li J, Yu C, Sadlonova A, Cooper MD (2009) Dual nature of the adaptive immune system in lampreys. Nature 459:796–801

Haanstra JR, van Tuijl A, van Dam J, van Winden W, Tielens AG, van Hellemond JJ, Bakker BM (2012) Proliferating bloodstream-form *Trypanosoma brucei* use a negligible part of consumed glucose for anabolic processes. Int J Parasitol 42:667–673

Hall JP, Wang H, Barry JD (2013) Mosaic VSGs and the scale of Trypanosoma brucei antigenic variation. PLoS Pathog 9, e1003502

Han BW, Herrin BR, Cooper MD, Wilson IA (2008) Antigen recognition by variable lymphocyte receptors. Science 321:1834–1837

Hanington PC, Zhang SM (2011) The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. J Innate Immun 3:17–27

Hattori D, Chen Y, Matthews BJ, Salwinski L, Sabatti C, Grueber WB, Zipursky SL (2009) Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. Nature 461:644–648

Henson PM, Bratton DL (2013) Antiinflammatory effects of apoptotic cells. J Clin Invest 123:2773–2774

Hirano M, Guo P, McCurley N, Schorpp M, Das S, Boehm T, Cooper MD (2013) Evolutionary implications of a third lymphocyte lineage in lampreys. Nature 501:435–438

Huang G, Xie X, Han Y, Fan L, Chen J, Mou C, Guo L, Liu H, Zhang Q, Chen S et al (2007) The identification of lymphocyte-like cells and lymphoid-related genes in amphioxus indicates the twilight for the emergency of adaptive immune system. PLoS One 2, e206

Jacob F (1977) Evolution and tinkering. Science 196:1161–1166

Jarosz DF, Lindquist S (2010) Hsp90 and environmental stress transform the adaptive value of natural genetic variation. Science 330:1820–1824

Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik S-G, Lee H, Lee J-O (2007) Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell 130:1071–1082

Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 3, e181

Kluytmans J, van Belkum A, Verbrugh H (1997) Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. Clin Microbiol Rev 10:505–520

Kozmik Z, Daube M, Frei E, Norman B, Kos L, Dishaw LJ, Noll M, Piatigorsky J (2003) Role of Pax genes in eye evolution: a cnidarian PaxB gene uniting Pax2 and Pax6 functions. Dev Cell 5:773

Kummer L, Parizek P, Rube P, Millgramm B, Prinz A, Mittl PRE, Kaufholz M, Zimmermann B, Herberg FW, Plückthun A (2012) Structural and functional analysis of phosphorylation-specific binders of the kinase ERK from designed ankyrin repeat protein libraries. Proc Natl Acad Sci 109:E2248–E2257

Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA 104:8597–8604

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB et al (2012) A systematic survey of loss-of-function variants in human protein-coding genes. Science 335:823–828

Michaut L, Flister S, Neeb M, White KP, Certa U, Gehring WJ (2003) Analysis of the eye developmental pathway in *Drosophila* using DNA microarrays. Proc Natl Acad Sci USA 100:4024–4029

Michod RE (2007) Evolution of individuality during the transition from unicellular to multicellular life. Proc Natl Acad Sci USA 104:8613–8618

Nagawa F, Kishishita N, Shimizu K, Hirose S, Miyoshi M, Nezu J, Nishimura T, Nishizumi H, Takahashi Y, Hashimoto S-I et al (2007) Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. Nat Immunol 8:206

Narbonne-Reveau K, Charroux B, Royet J (2011) Lack of an antibacterial response defect in *Drosophila* Toll-9 mutant. PLoS One 6, e17470

Nguyen HH, Tran BT, Muller W, Jack RS (2012) IL-10 acts as a developmental switch guiding monocyte differentiation to macrophages during a murine peritoneal infection. J Immunol 189:3112–3120

Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD (2004) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 430:174–180

Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O (2009) The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458:1191

Roberts DJ, Craig AG, Berendt AR, Pinches R, Nash G, Marsh K, Newbold CI (1992) Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature 357:689–692

Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. Nature 396:336

Salvini-Plawen L, Mayr E (1977) On the evolution of photoreceptors and eyes. Plenum, New York

Sancho D, Reis e Sousa C (2013) Sensing of cell death by myeloid C-type lectin receptors. Curr Opin Immunol 25:46–52

Sasaki N, Ogasawara M, Sekiguchi T, Kusumoto S, Satake H (2009) Toll-like receptors of the ascidian *Ciona intestinalis*. J Biol Chem 284:27336–27343

Shaw PX, Hörkkö S, Chang M-K, Curtiss LK, Palinski W, Silverman GJ, Witztum JL (2000) Natural antibodies with the T15 idiotype may act in atherosclerosis, apoptotic clearance, and protective immunity. J Clin Invest 105:1731–1740

Steinman RM, Turley S, Mellman I, Inaba K (2000) The induction of tolerance by dendritic cells that have captured apoptotic cells. J Exp Med 191:411–416

Stewart CR, Stuart LM, Wilkinson K, van Gils JM, Deng J, Halle A, Rayner KJ, Boyer L, Zhong R, Frazier WA et al (2010) CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. Nat Immunol 11:155–161

Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC (2003) Predominant autoantibody production by early human B cell precursors. Science 301:1374–1377

Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI, Schmucker D (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 309:1874–1878

Wells CA, Salvage-Jones JA, Li X, Hitchens K, Butcher S, Murray RZ, Beckhouse AG, Lo Y-L-S, Manzanero S, Cobbold C et al (2008) The macrophage-inducible C-type lectin, mincle, is an essential component of the innate immune response to *Candida albicans*. J Immunol 180:7404–7413

West XZ, Malinin NL, Merkulova AA, Tischenko M, Kerr BA, Borden EC, Podrez EA, Salomon RG, Byzova TV (2010) Oxidative stress induces angiogenesis by activating TLR2 with novel endogenous ligands. Nature 467:972–976

Yamasaki S, Ishikawa E, Sakuma M, Hara H, Ogata K, Saito T (2008) Mincle is an ITAM-coupled activating receptor that senses damaged cells. Nat Immunol 9:1179–1188

Yoon SI, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, Wilson IA (2012) Structural basis of TLR5-flagellin recognition and signaling. Science 335:859–864

Part I
Pathogens: Antigen Repertoires and Variations

A Host–Pathogen Interaction Reduced to First Principles: Antigenic Variation in *T. brucei*

Galadriel Hovel-Miner, Monica Mugnier, F. Nina Papavasiliou, Jason Pinger, and Danae Schulz

Abstract Antigenic variation is a common microbial survival strategy, powered by diversity in expressed surface antigens across the pathogen population over the course of infection. Even so, among pathogens, African trypanosomes have the most comprehensive system of antigenic variation described. African trypanosomes (*Trypanosoma brucei* spp.) are unicellular parasites native to sub-Saharan Africa, and the causative agents of sleeping sickness in humans and of n'agana in livestock. They cycle between two habitats: a specific species of fly (*Glossina* spp. or, colloquially, the *tsetse*) and the bloodstream of their mammalian hosts, by assuming a succession of proliferative and quiescent developmental forms, which vary widely in cell architecture and function. Key to each of the developmental forms that arise during these transitions is the composition of the surface coat that covers the plasma membrane.

The trypanosome surface coat is extremely dense, covered by millions of repeats of developmentally specified proteins: procyclin gene products cover the organism while it resides in the *tsetse* and metacyclic gene products cover it while in the fly salivary glands, ready to make the transition to the mammalian bloodstream. But by far the most interesting coat is the Variant Surface Glycoprotein (VSG) coat that covers the organism in its infectious form (during which it must survive free living in the mammalian bloodstream). This coat is highly antigenic and elicits robust VSG-specific antibodies that mediate efficient opsonization and complement mediated lysis of the parasites carrying the coat against which the response was made. Meanwhile, a small proportion of the parasite population switches coats, which stimulates a new antibody response to the prevalent (new) VSG species and this process repeats until immune system failure. The disease is fatal unless treated, and treatment at the later stages is extremely toxic.

Because the organism is free living in the blood, the VSG:antibody surface represents the interface between pathogen and host, and defines the interaction of

G. Hovel-Miner • M. Mugnier • F.N. Papavasiliou (✉) • J. Pinger • D. Schulz
Laboratory of Lymphocyte Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
e-mail: papavasiliou@rockefeller.edu

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_2

the parasite with the immune response. This interaction (cycles of VSG switching, antibody generation, and parasite deletion) results in stereotypical peaks and troughs of parasitemia that were first recognized more than 100 years ago. Essentially, the mechanism of antigenic variation in *T. brucei* results from a need, at the population level, to maintain an extensive repertoire, to evade the antibody response. In this chapter, we will examine what is currently known about the *VSG* repertoire, its depth, and the mechanisms that diversify it both at the molecular (DNA) and at the phenotypic (surface displayed) level, as well as how it could interact with antibodies raised specifically against it in the host.

List of Abbreviations

VSG	Variant surface glycoprotein
BF	Bloodstream form
PF	Procyclic form
BES	Bloodstream expression site
GC	Gene conversion
TE	Telomeric exchange
ESB	Expression site body
PolI	RNA polymerase I
PolII	RNA polymerase II
DSB	Double-stranded break

1 The VSG Coat and Its Roles in Immune Evasion

In its bloodstream form (BF), *Trypanosoma brucei* lives freely in the blood, lymph, and interstitial fluids of its host, fully exposed to both innate and adaptive immune responses. In that milieu, the VSG coat shields invariant components of the cell surface from the immune system, and elicits an antibody response against only the components of the coat that are altered by antigenic variation.

VSG molecules, which comprise the most abundant protein in BF *T. brucei,* accounting for about 10 % of the total protein content of the cell (Overath & Engstler 2004), are bound to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor and displayed on the cell surface as an array of 5×10^6 identical homodimers. This dense (12–15 nm thick) surface layer does not elicit production of opsonins or allow assembly of the membrane attack complex unless mediated by bound antibodies, thus evading the action of many innate immune components in the absence of an adaptive humoral immune response (Ferrante & Allison 1983; Black et al. 2010) and effectively functioning as a physical shield.

VSGs are highly immunogenic and elicit robust antibody responses, but their structure supports tight packing on the cell surface and shielding of all but a few epitopes. All VSGs consist of two domains: a long, highly variable N-terminal

domain, and a smaller, more conserved C-terminal domain to which the GPI anchor is attached (Blum et al. 1993). Early studies characterizing the VSG coat indicated that only a small surface region of the large VSG protein is exposed to antibodies and that only a small portion of antibodies generated against soluble VSGs are able to bind VSGs expressed on living trypanosomes (Barry 1979; Masterson et al. 1988). These observations, along with measurements of coat density and thickness, and later determinations of VSG crystal structures (Blum et al. 1993; Freymann et al. 1990; Chattopadhyay et al. 2005), have informed our current model of the larger structure of the VSG coat. In this model, individual VSG homodimers extend vertically from the cell surface, and close packing between homodimers prohibits antibody access to all surfaces except for the highly variable "top" of the molecule. Thus, the more conserved C-terminal domain and many epitopes of the N-terminal domain are hidden by steric hindrance. It is generally accepted that the density of VSGs on the cell surface is also sufficient to hide other invariant surface antigens from B-cell recognition. This theory has not been rigorously tested, and it is possible that other determinants are recognized to some degree, but the bulk of the humoral response is directed against the VSG (Field et al. 2009).

Individual VSGs do not remain externally exposed for extended periods at a time; they are constantly shuttled on and off the cell surface. All endo- and exocytosis in *T. brucei* occurs at the flagellar pocket, a specialized structure that makes up 2 % of the total surface membrane, and is situated at the base of the flagellum at the posterior end of the cell. VSGs at the flagellar pocket are internalized by endocytosis in clathrin-coated vesicles (Allen et al. 2003) and delivered to RAB5-positive early endosomes. The vast majority are then sorted to RAB11-positive recycling endosomes (directly or via RAB7-positive late endosomes) where they are delivered back to the flagellar pocket and reemerge onto the cell surface (Manna et al. 2014; Pal et al. 2003; Grünfelder et al. 2003; Engstler et al. 2004). Roughly 9 % of the total cellular VSG content is in this internal recycling pathway at steady state (Engstler et al. 2004). Despite the high abundance of surface VSGs, this process is rapid and efficient. The entire surface coat goes through one round of internalization and redistribution to the cell surface in approximately 12.5 min (Engstler et al. 2004).

It has long been observed that live trypanosomes rapidly redistribute surface-bound antibody to the flagellar pocket (Barry 1979). VSG and bound antibody are then internalized together, but upon internalization, the antibody is separated from the VSG, transported to the lysosome, and degraded (Pal et al. 2003; O'Beirne et al. 1998). It is not entirely clear how VSGs are specifically sorted from other proteins in the endosome, but the current model suggests that this selection is based on default sorting of GPI-anchored proteins to the recycling pathway (Manna et al. 2014; Grünfelder et al. 2003). VSG recycling thus provides a mechanism for rapid clearance of host antibodies from the trypanosome surface. Trypanosomes have been demonstrated to fully remove a single layer of surface-bound IgG in 120 s at 37 °C (Engstler et al. 2007), far faster than the rate of internalization of the VSG coat. This increased antibody clearance rate has been accounted for by the influence of hydrodynamic forces created by parasite motility. Beating of the

trypanosome flagellum produces forward motion, causing bound antibodies to act as "molecular sails," which preferentially drives VSG–antibody complexes to the posteriorly located flagellar pocket (Engstler et al. 2007). Overall, the high rate of endocytosis and surface coat recycling appears to be another adaptation for immune evasion in the mammalian host.

2 Antigenic Variation of VSG Coats

2.1 The VSG Genomic Repertoire

In 1909, Ross and Thomson applied a new method for counting parasites to the blood samples of a patient infected with *T. brucei gambiense*. Their methodical counting revealed, for the first time, the periodic peaks and valleys in parasitemia characteristic of *T. brucei* infections (Ross and Thomson 1910). These waves, we now understand, are the result of periodic "switching" of the parasite's variant surface glycoprotein (VSG) coat.

Recent work sequencing the "VSGnome" of *T. brucei* (Lister427 strain) has provided great insight into the genomic *VSG* repertoire (Cross et al. 2014). In contrast with previous predictions (Van der Ploeg et al. 1982), this analysis revealed more than 2000 different *VSG*-encoding genes. Interestingly, the great majority of these are incomplete *VSGs* or *VSG* pseudogenes. This might at first suggest a very limited repertoire of *VSG* genes from which the parasite can draw during an infection. What purpose could a repertoire of partial *VSG* genes serve? It appears that this repertoire of partial *VSGs* and *VSG* pseudogenes may be extremely important to *T. brucei* infection. When expressed *VSGs* were cloned from late stages of *T. equiperdum* infections in rats, it was discovered that of three clones, isolated based on recognition by a single polyclonal antibody, all were distinct and derived from three or four *VSG* pseudogenes (Roth et al. 1989). This type of chimeric *VSG* sequence is referred to as a "mosaic *VSG*," and the appearance of such mosaic *VSGs* late in infection suggests a role for the large pool of *VSG* pseudogenes in the *T. brucei* genome.

Mosaic *VSGs* have also been described in *T. brucei* infections (Kamper & Barbet 1992). A recent study looked in depth at the formation of mosaic *VSGs* in vivo by cloning and sequencing hundreds of expressed *VSGs* over the course of mouse infections (Hall et al. 2013). This work again confirmed the late emergence of mosaic *VSGs* during infection. In 4–5 week-long infections, mosaic *VSGs* only began to emerge at week 3 of infection. Moreover, these experiments demonstrated that mosaics with similar sequence were nonetheless antigenically distinct, as antibodies raised against one variant would not cross-react with a related variant from the same infection. Modeling of the protein structure of these VSGs showed that many of the distinct residues on related mosaics occurred in regions likely to be exposed on the trypanosome surface and thus most immunogenic. Altogether,

reports of mosaic *VSGs* suggest that the repertoire of complete *VSG* genes at a given time is perhaps not sufficient for escape from the host immune system in a chronic infection. Switching may occur at a rate that exhausts the intact *VSG* repertoire relatively early during infection. As a result, the much larger repertoire of *VSG* pseudogenes and partial *VSGs* must be accessed to produce mosaic *VSGs*, in which segments of partial *VSGs* are combined by gene conversion to build a new, intact, and (ideally) antigenically distinct VSG coat. Some studies have suggested an even greater extension of the *VSG* repertoire through the introduction of point mutations during gene conversion events (Lu et al. 1993, 1994), although others have suggested this is not a significant source of VSG diversity (Graham & Barry 1996).

It is important to note that while mosaic *VSGs* are expressed later in *T. brucei* infection, it is not known when or how these variants are actually formed within the genome. Hall et al. described mosaic *VSGs* whose boundaries of segmental gene conversion between donors showed absolutely no homology (Hall et al. 2013). Donors typically showed homology with one another, but it did not appear that homology at the site of gene conversion was required for the formation of mosaic *VSG* genes. Moreover, it is not clear where in the genome mosaic *VSGs* form or how frequently they arise.

2.2 Immune Response to VSG Coats

Given the immense repertoire of *VSGs T. brucei* may use during infection, and the parasite's extracellular lifestyle, it is no surprise that the immune response to *T. brucei* is primarily B-cell mediated. The immune response to the parasite has been thought to be mediated by IgM B cells. This is based on a study tracking VSG-specific antibody titers during *T. brucei* infections (Dempsey & Mansfield 1983), which reported that IgM appeared earliest in infection and persisted longest. VSG-specific IgG did not appear until the first peak of parasitemia was cleared, leading the authors to conclude that the IgM response was responsible for clearance of each variant. Other studies, however, paint a more complex picture of the arms race between host and parasite. One study reported polyclonal B-cell activation (nonspecific) in response to *T. brucei* infection (Diffley 1983), while VSG-specific antibodies could also be detected during infection (Dempsey & Mansfield 1983; Musoke et al. 1981). In addition, certain B-cell populations have been reported to be lost during *T. brucei* infection, specifically IgM marginal zone B cells in the spleen. Similarly, coinfection with *T. brucei* resulted in a demonstrated loss in vaccine-induced protection against *B. pertussis* infection in mice (Radwanska et al. 2008). Based on these observations, it is interesting to speculate why a parasite that undergoes antigenic variation might simultaneously destroy immune memory.

The T-cell response to *T. brucei* has also been characterized (Schleifer et al. 1993), and it appears to be very important in eliciting an IFN-γ response during infection, which is a major determinant in resistance and susceptibility to infection (Hertz et al. 1998). Interestingly, although T cells could theoretically recognize peptide fragments from the entire VSG protein, including the more

conserved C-terminal domain, an analysis of T cells activated during *T. brucei* infection indicated that these cells only recognized epitopes originating from the more variable N-terminal domain (Dagenais et al. 2009). Somehow, the parasite manages to avoid immune recognition at invariant regions of the VSG. Thus, both arms of the adaptive immune system primarily recognize the variable, and exposed, N-terminal domain of the VSG.

2.3 Characteristics of Antigenic Variation (VSG Coat Switching)

It is obvious that the parasite interferes with the host immune response in a number of ways. It is unclear, however, what the mechanism of this interference is. There are many factors that would contribute to *VSG* expression dynamics in vivo. These can be grouped into two general categories: (1) the rate of switching (is the parasite exhausting humoral immunity simply by virtue of very frequent switching?) and (2) the characteristics of each new variant (including growth rate differences and immune cross-reactivity between VSGs).

Despite *T. brucei's* large repertoire of *VSGs*, it is thought that the rate of switching must be tightly controlled. If switching is too slow, the immune system will recognize *T. brucei* and clear the infection, preventing transmission to the tsetse fly. If switching is too fast, it is possible that the *VSG* repertoire will be exhausted prematurely. Many measurements of *VSG* switching rates have been made, both in vivo and in vitro. In laboratory-adapted strains, the rate of switching ranges between 10^{-5} and 10^{-7} switch events per parasite per population doubling both in vivo (Lamont et al. 1986; Doyle et al. 1980; Turner 1997) and in vitro (Lamont et al. 1986; Doyle et al. 1980; Aitcheson et al. 2005). Many interpret the in vitro data to suggest that the process of switching is purely stochastic. However, some studies have shown the rate of switching in naturally occurring (fly-transmitted) infections to be much higher, perhaps as high as 1 in 100 (Turner 1997; Turner & Barry 1989). These findings suggest that there may be some environmental cue initiating a *VSG* switch, or at least an environmental influence on the frequency of *VSG* switching.

Determining the true rate of *VSG* switching presents experimental challenges. Trypanosomes expressing different *VSGs* may grow at different rates, or trypanosomes may switch to a *VSG* for which the immune system has already mounted a full or partial response. Moreover, trypanosomes can sequester in extravascular tissues (Seed et al. 1984), and most work examines only populations in the bloodstream. As a result, even if a switch occurs, that switched variant may never establish in the measured population. It is also difficult, if not impossible, to examine the response to a changing antigen without an understanding of how that antigen changes. Unfortunately, while our understanding of the *VSG* repertoire has improved greatly in recent years due to next-generation sequencing approaches, our understanding of how that repertoire is accessed in vivo has not yet benefitted from more modern techniques (Fig. 1).

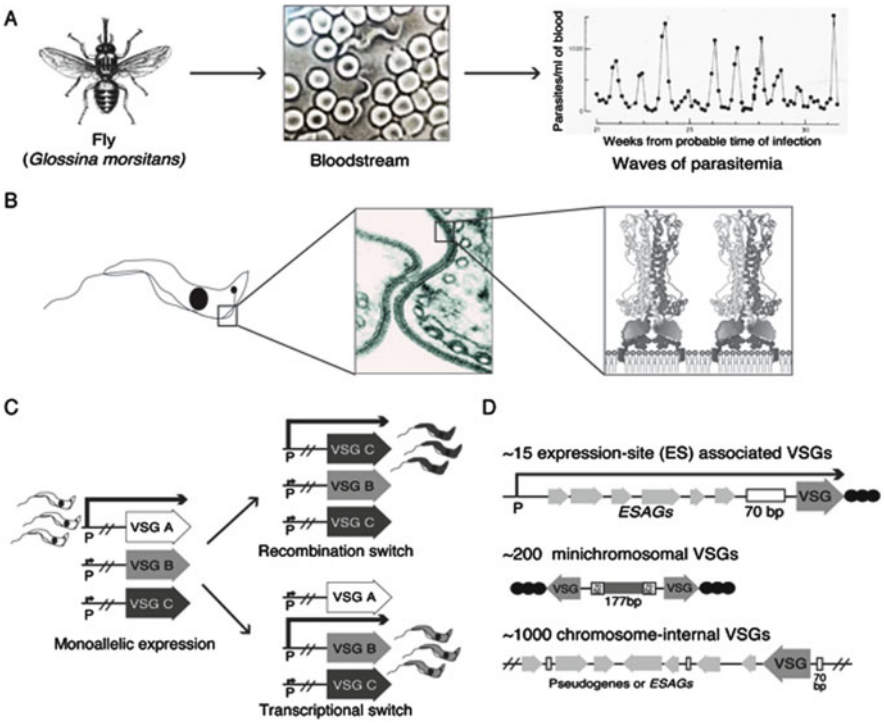


Fig. 1 (**a**) *Trypanosoma brucei* is transmitted from the salivary gland of the fly, *Glossina morsitans*, to its mammalian host. There it lives extracellularly in the bloodstream. Patients infected with *T. brucei* display waves of parasitemia that reflect the interaction between the parasite and the host immune system. Host antibodies are able to mostly clear parasites from the bloodstream, but a small number of parasites are able to escape the response, leading to a resurgence in parasite number (image adapted from Ross and Thomson 1910). Escape from the immune system is largely mediated by the Variant Surface Glycoprotein (VSG) that densely coats the parasite surface, as shown by the electron micrograph (http://tryps.rockefeller.edu/trypsru2_avariation_intro.html) displayed in (**b**). Ten million VSG molecules homodimerize on the surface of the parasite cell, shielding all remaining surface epitopes from exposure to the host immune system (image courtesy of Markus Engstler). (**c**) VSG genes are expressed from ~15 telomeric expression sites, only one of which is transcriptionally active at any given time. The parasite can switch expression to an antigenically distinct VSG by (1) silencing the initial expression site and transcriptionally activating a new one, thus activating a new VSG (*bottom*, transcription switch) or (2) copying a VSG from a separate expression site or from elsewhere in the genome into the active transcription site (*top*, Recombination switch). Using a repertoire of ~2000 distinct VSG genes, the parasite is able to effectively disguise itself from existing host antibodies by remodeling its surface coat. (**d**) The extensive repertoire of VSG genes is located in multiple places throughout the genome. Around 15 VSGs are associated with expression sites, where they are transcribed with a cohort of Expression Site Associated Genes (ESAGs). Expression sites also contain repetitive 70 bp repeat elements. Around 200 VSGs are located on minichromosomes, which contain 177 bp repetitive elements. Finally, a large number of VSGs are located in chromosome-internal regions, where they are often found in combination with ESAGs and repetitive elements

To date, the resolution of most studies measuring switching rate in *T. brucei* has been quite low. In fact, nearly all of the studies on *VSG* switching rate and *VSG* switching hierarchy have measured less than ten *VSGs*, for which either the sequence of the *VSG* gene was known or antibodies against the *VSG* existed (Turner 1997; Turner & Barry 1989; Miller & Turner 1981; Liu et al. 1985; Morrison et al. 2005). Even more recent work cloning hundreds of *VSGs* over the course of infection (Hall et al. 2013) could not detect rare variants at any given point, due to limited sampling. Thus, if some variants arise but never establish within the population (due to cross-reactivity, physical location, or intrinsic fitness), they cannot be measured using existing methods. This makes it difficult to estimate the effects of competition between variants or immune cross-reactivity.

Although limited in resolution, studies looking at a handful of *VSGs* have shed some light on the characteristics of new variants that arise during infection. It has been observed in many in vivo studies that certain VSGs are more likely to arise early in infection than others (Miller & Turner 1981; Liu et al. 1985; Morrison et al. 2005; Myler et al. 1985). While some work suggests that this can be explained in part by differences in growth rate (i.e., a variant that grows the fastest will appear first) (Myler et al. 1985; Lythgoe et al. 2007), this view has been challenged (Kosinski 1980), and it is clear that the genomic position of a *VSG* has an influence on when that *VSG* appears during infection. Telomeric *VSGs* are much more likely to arise early in infection (Liu et al. 1985), while more complicated forms of recombinant *VSGs* tend to appear later (Hall et al. 2013). The mechanism for this preference is not well understood, although it makes logical sense that *VSGs* that comprise segmental gene conversion events may be more likely to arise later in infection, when the repertoire of complete *VSGs* has been exhausted. Finally, other studies report that the immune system is the major player in determining *VSG* order, primarily through cross-reactive epitopes (Antia et al. 1996; Recker et al. 2004), thus highlighting the role of the immune repertoire in this host:pathogen arms race.

Recently, we developed a high-throughput method for identifying and measuring expressed *VSGs* during infection (Mugnier et al. 2015). Our study of expressed *VSGs* over the course of chronic mouse infections confirmed many of the findings from previous low-resolution studies. Indeed, *VSG* populations appear to be biased toward certain *VSGs* at different times during infection, mosaic *VSGs* tend to appear later infection, and immune cross-reactivity appears to play a key role in determining variant success. On the other hand, this work brought to light aspects of infection that could not be accessed using other approaches. During infection, many *VSGs* are expressed at a single time (28 variants, on average), and most of these variants never establish to levels at which they could be detected using low-resolution techniques. Thus, the complete functional *VSG* repertoire (~400 *VSGs* in Lister 427) is almost certainly insufficient to support the long *T. brucei* infections frequently observed in the field. *VSG* diversification through the formation of mosaic *VSGs*, or perhaps other yet to be identified mechanisms, may play a more significant role in sustaining infection than previously thought.

A better understanding of how the genomic *VSG* repertoire is accessed during infection would certainly shed light on the mechanisms of switching in *T. brucei*.

The recently catalogued VSGnome reference (Cross et al. 2014) can now be used to study how the *VSG* repertoire is used, in high resolution. Simultaneously, the immune response of the parasite can be measured. Now that the system in vivo can be described accurately, it may be possible to disentangle some of these many variables and reveal the fundamental mechanistic parameters that define this extremely dynamic host–pathogen interface.

3 Mechanisms of *VSG* Switching

Antigenic variation is required for *T. brucei* to effectively evade the host immune response during the bloodstream stage of the life cycle. Switching from the expression of one *VSG*-encoding gene to the next occurs through specific genetic processes that allow access to an extensive genetic repertoire (discussed in Sect. 2.1). At the same time, it is essential that only one VSG be displayed on the cell surface at a given time to ensure that the parasite has time to replicate before the immune system marshals its forces. This dual requirement for switching *VSG* expression and for monoallelic expression of each coat (discussed in detail in Sect. 4) controls antigenic variation at the molecular level.

Both requirements are facilitated by the special structure of the locus that harbors the expressed *VSG*. Transcription of the active *VSG* gene occurs from one of approximately 15 specialized subtelomeric Bloodstream Expression Sites (BES) that have a similar organization. Each BES can be likened to an operon with several unique features: (1) transcription initiates from an RNA Polymerase I promoter (PolI), which in all other organisms studied is exclusively used for ribosomal RNA production; (2) the promoter is followed by a series of expression site associated genes (ESAGs); (3) then a long region of repetitive DNA; and (4) a single *VSG* gene in close proximity to the telomeric repeats of the chromosome (Shea et al. 1986; Hertz-Fowler et al. 2008). The active BES produces a single polycistronic transcript (30–60 kb in length depending on the specific site) that is subsequently spliced and targeted for nuclear export (Alarcon et al. 1994). BESs are highly similar at the nucleotide sequence level with minor variations in the number and order of ESAG genes (Hertz-Fowler et al. 2008). *VSG*-encoding genes are very dissimilar at the level of DNA sequence, bearing only a short conserved tract at the 3′-end. Thus, the most unique BES region is the *VSG* gene.

This location also facilitates the two major classes of *VSG* switching, namely the transcriptional switch (or "in situ") between BESs (i.e., inactivation of one and activation of another) or recombination-based switching, which results in a new *VSG* gene in the actively transcribed BES [reviewed extensively Barry & McCulloch (2001), Borst & Ulbert (2001), Pays et al. (1994), Borst et al. (1998)]. The 15–20 *VSGs* that reside in BESs are a small portion of the total *VSG* archive (~2000) encoded in the *T. brucei* genome (Cross et al. 2014). The majority of *VSG* genes are encoded in gene arrays on megabase chromosomes and the remainder is subtelomerically located on intermediate chromosomes (~10) or minichromosomes

(~100), which have the potential for bearing a *VSG* at each telomere (Cross et al. 2014; Berriman et al. 2005; Wickstead et al. 2004). Thus, 1–5 % of the *VSG* archive is encoded in sites that can be expressed (BESs), while the remainder resides in physically distinct, transcriptionally inactive (comparatively), and actively silenced (see Sect. 4) genomic sites. The following sections will describe the discoveries and remaining questions regarding BES enforced monoallelic expression as well as review current knowledge regarding the two broad classes of *VSG* switching and their contribution to the repertoire.

3.1 Transcriptional Switching

VSG expression is not only required for infection, it is essential for life in African trypanosome species. Safeguards must be in place within the cell to ensure that expression of the active *VSG* is not compromised during switching from one *VSG* gene to the next. Transcription of the active BES by PolI occurs rapidly and processively through many tens of kilobases and, ultimately, the *VSG*-encoding gene (Zomerdijk et al. 1990, 1991). Modest amounts of transcription occur from silent PolI promoters, but only one VSG coat is expressed on the surface of the cell. Transcriptional *VSG* switching occurs by the simultaneous activation of a new BES and inactivation of the previously expressed site (Johnson & Borst 1986; Horn & Cross 1997). This type of switch can only promote access to the small fraction of the *VSG* repertoire that is encoded on BESs. It is nonetheless worth considering the dynamics of transcriptional switching, why it occurs, and whether or not transcription is a layer of switching regulation.

The actively expressed BES resides in a subnuclear compartment termed the expression site body (ESB), from which the inactive BESs are believed to be excluded [reviewed in Navarro et al. (2007)]. Similarly, the active BES has a much lower level of nucleosome occupancy than its silenced counterparts (both of these phenomena are covered in detail in the monoallelic exclusion section of this chapter) (Figueiredo & Cross 2010; Stanne & Rudenko 2010). Thus, a transcriptional switch of *VSG* expression would be accompanied by both a considerable change in the chromatin state of the newly active site and the active repositioning of this site into the ESB coupled with the predicted ESB exclusion of the newly silenced site. The natural conditions under which transcriptional switching occurs are unknown; however, transcriptional switching has been monitored in a number of potentially informative laboratory experiments.

Genetic analysis of trypanosomes that have switched during in vivo infections under laboratory conditions suggests that transcriptional switching accounts for approximately 20–60 % of the events observed (Myler et al. 1984a, b). As with most early in vivo analysis, these measurements are based on very few switched isolates. More recent in vitro approaches have enabled enrichment of switchers from a population for more quantitative genetic determination. These in vitro studies suggest that transcriptional switching in wild-type strains occurs in as few

as 2 % and as much as 30 % of the total switchers, depending on the isolation method (Kim & Cross 2010; Hovel-Miner et al. 2012). Thus, while the exact amount that transcriptional switching contributes to switching in vivo, in vitro, or during natural infections is unclear, it is a mode of switching that must be considered.

Due to its low diversity contribution (<10 % of the total repertoire) and potentially infrequent occurrence, some have wondered: why have more than one BES and why retain a system of transcriptional switching? About 20 years ago, a theory emerged that the system provided a safeguard against irreparable BES damage. Keeping in mind that *VSG* expression is essential, it was proposed that transcriptional switching only occurs when the active BES can no longer produce sufficient *VSG* transcript. Subsequent experiments have shown that when the active BES promoter is artificially shut off, or if the end of the active chromosome is sufficiently damaged, transcriptional switching is favored (Glover et al. 2007). These data support the hypothesis that either a damaged active site or loss of *VSG* expression can trigger the transcription switch. However, exactly how BES damage is sensed and whether this is the only source of transcriptional switching are unknown.

One of the most unusual and fascinating aspects of *T. brucei* genetics is that transcription is largely unregulated. Transcription of most of the genome occurs in a largely constitutive manner by RNA Polymerase II (PolII) without the benefit of transcription factors and expression is primarily regulated at the level of translation (De Gaudenzi et al. 2011; Vasquez et al. 2014). In contrast, BESs have PolI promoters that are not only identical to one another at the DNA sequence level but also indistinguishable from most rRNA promoters, which can be swapped with BES promoters to retain active BES function (Rudenko et al. 1995). Until recently these observations seemed to preclude any traditional model of transcriptional regulation of switching, in which signaling may occur through transcription factors to control genetic outcomes. Recent studies have identified and characterized a multicomponent novel Class I Transcription Factor A (CITFA) that binds to PolI promoters and is required for BES transcription (Brandenburg et al. 2007; Nguyen et al. 2012). In addition, CITFA subunits predominantly occupy the active BES promoter and are localized with the ESB (Nguyen et al. 2014). Whether CITFA directs activation and ESB localization of the active BES or reinforces the existing state (active or silent) of the BES is unclear. It is intriguing to speculate, and worth investigating, if this newly identified level of transcriptional regulation could be a component in a pathway that results in transcriptional switching.

3.2 *Recombinatorial Switching*

Chronic infection with *T. brucei* is dependent upon its ability to activate the expression of *VSGs* encoded throughout the genome. Since most *VSG* genes are located in transcriptionally silent sites, recombination-based mechanisms exist to

copy a silent *VSG* gene into an actively expressed BES. The two major types of recombinatorial switching are reciprocal telomeric exchange and duplicative gene conversion. Telomeric exchange (TE) is the result of a homologous recombination event between telomeric ends that results in the reciprocal translocation of a *VSG* from a silent site into the active BES and the previously active *VSG* into the corresponding silent site. By definition, TE facilitates selection of a limited repertoire because it can only occur between telomeric ends that contain functional bloodstream *VSGs* (~200). TE is also not considered to be a major contributor to the overall rate of switching. In contrast, duplicative gene conversion (GC) is an unbalanced chromosomal translocation in which a donor *VSG* encoded in a silent site is duplicated into the active BES resulting in deletion of the previously active *VSG*. Thus, TE and GC can be distinguished experimentally by their genetic outcomes. There are multiple subtypes of GC associated with their specific genetic outcomes; for instance, expression site gene conversion (ESGC) is a GC event in which the entire BES sequence is replaced with another BES. Unlike transcriptional switching and TE, GC can (theoretically) access the entire functional *VSG* repertoire regardless of genomic site (BES, internal array, or MC) as well as being the putative source of mosaic *VSG* formation (discussed in Sect. 2). GC is observed to be a major contributor to switching in vitro and considered by many to be the predominant mechanism in natural infections (Robinson et al. 1999). While transcriptional switching and TE are naturally restricted to telomeric *VSGs*, subtelomeres are also preferred sites of *VSG* donor selection during GC. In fact, there is a natural hierarchy in *VSG* donor selection that first favors BES-encoded *VSGs*, followed by non-BES subtelomeres (MC and others), and finally chromosome internal arrays (Morrison et al. 2005; Lythgoe et al. 2007). The observed semi-predictable order and inherent considerations of switching frequency and timing during natural infections have resulted in many speculations regarding how recombinatorial switching is monitored and controlled. The following sections will provide an overview of what is known and what remains to be discovered about recombinatorial switching with an emphasis on gene conversion.

Gene conversion events have been observed in most chromosomally organized genomes. In mammalian genomes, they are usually spontaneous, often detrimental, and can result in genetic disorders and specific cancers (Chen et al. 2007; Kobayashi 1992). The essential steps of GC, which include DNA break formation, ssDNA strand migration, and DNA duplication, are likely conserved among eukaryotic genomes (Kobayashi 1992), but how they are accomplished in *T. brucei* antigenic variation is an active research area. BES-encoded *VSGs* are located near the end of the chromosome downstream from a long region (4–16 kb) of "70-bp" DNA repeats and upstream from the *VSG* and telomeric repeats (Hertz-Fowler et al. 2008). Naturally occurring DNA lesions that may precipitate GC have been observed throughout BESs (Glover et al. 2013a) with possible preferable formation within the 70-bp repeats of the active BES (Boothroyd et al. 2009). Induction of site-specific DNA double-stranded break (DSB) in the active BES results almost exclusively in switching by GC, but the exact position of the DNA break formed can affect the amount of switching and the specific genetic outcomes

(Glover et al. 2008, 2013a; Boothroyd et al. 2009). A proposed source of naturally occurring DNA breaks is subtelomeric fragility of the actively expressed site, which may acquire DNA lesions from collisions between transcription and DNA replication machinery [reviewed in Dreesen et al. (2007)]. In support of this hypothesis, it was shown that artificial shortening of the active site telomere results in an increase in switching and GC (Hovel-Miner et al. 2012). It is also intriguing to note that natural *T. brucei* isolates have shorter telomeres than laboratory-adapted strains (Dreesen & Cross 2008), which may, in part, account for the fact that laboratory-adapted strains have lower rates of switching.

VSG switching by GC occurs by homologous recombination events in which homologs of specific recombination proteins have been implicated. A deficiency in *T. brucei* RAD51, which forms nucleoprotein filaments on ssDNA at the site of damage and catalyzes the transfer to a homologous sequence, results in reduced switching (McCulloch & Barry 1999). Mutations of *T. brucei* BRCA2 homolog, which is a key mediator of RAD51 in other systems, impair *VSG* switching and display genome instability (Hartley & McCulloch 2008). In contrast, mutation of MRE11, a sensor of DSBs in other organisms, resulted in DNA damage and chromosomal rearrangements but did not affect *VSG* switching (Robinson et al. 2002). Another interesting question is why GC occurs in response to DNA break formation rather than the crossover events that result in TE. Mutation of *T. brucei* homologs of either TOPO3a or RIM1, which function in the same genome maintenance pathway in other eukaryotes, tips the balance between these events toward TE (Kim & Cross 2010, 2011). While many of the events and factors associated with recombinatorial switching can be determined based on conservation with other eukaryotes, *T. brucei's* genomic divergence results in unexpected differences and the potential for novel protein functions.

The fact that conserved recombination proteins are required for GC and that *VSG* expression occurs from unstable genomic sites supports a stochastic model for *VSG* switching and its genetic outcomes. In this model, DNA breaks form randomly as a result of genomic positioning, the resulting ssDNA uses 5′ 70-bp homology and 3' *VSG* homology to pair with the new *VSG* donor gene, and conserved DNA duplication and recombination proteins complete translocation into the active site [reviewed recently Horn (2014)]. The implication of this model is that no switching specific factors exist that either control the amount of switching or direct donor selection toward specific *VSG*-encoding sites. Yet, the frequency and source of naturally occurring DNA breaks in BESs is still speculative and it is not known if the breakage observed in vitro is sufficient to support antigenic variation in a natural infection (Glover et al. 2013a; Boothroyd et al. 2009). Selection of the *VSG* donor during gene conversion is semi-ordered (as described above) and predicted to be driven by homology of 70-bp upstream and *VSG* 3′ conserved sequences in the donor sites. However, conservation of the 3′ end is consistent among all *VSG* genes (Cross et al. 2014) and recently data suggest that the repetitive element that composes the 70-bp repeat has a highly conserved sequence throughout the genome (BES, internal array, and MC—Hovel-Miner et al. unpublished data). Thus, variations in the conserved DNA elements flanking *VSG* genes may not be a sufficient

predictor of *VSG* donor selection. Similarly, BESs have both long regions of 70-bp repeats and extensive sequence identity in common with one another (Hertz-Fowler et al. 2008) and yet sites other than BESs are selected frequently even early in switching (Boothroyd et al. 2009). Alternative explanations to homology-driven donor selection include physical proximity of *VSG* containing sites in the nucleus or hypothetical factors that guide ssDNA in homologous pairing with donor sites. Much has been, and will continue to be, learned using functional homologs from other eukaryotic genomes. It is important to keep in mind that the findings from those studies do not preclude the involvement of functionally novel or highly divergent factors in *T. brucei* antigenic variation that have not yet been identified.

4 *VSG* Monoallelic Expression

As already mentioned, trypanosomes actively transcribe a single *VSG* gene whose product is displayed on the surface of the cell while transcriptionally repressing the rest. The parasite is thus faced with the problem of expressing one of the variants at a high enough level to completely coat the surface of the cell, while keeping nearly a quarter of the genes in its genome transcriptionally silenced. This is referred to as monoallelic expression and is essential for a number of developmental processes in mammalian cells. These include olfactory gene receptor usage (Rodriguez 2013), differential rhodopsin gene usage in the retina (Rister et al. 2013), as well as immunoglobulin allelic exclusion in the development of B and T cells of the immune system (Vettermann & Schlissel 2010). In these diverse systems, regulation of chromatin structure has emerged as a common mechanism by which monoallelic expression is maintained (Peng & Chen 2007; Magklara et al. 2011; Stanhope-Baker et al. 1996). Recent evidence from the malaria field has also implicated chromatin structure in the maintenance of monoallelic expression of the *var* genes (Volz et al. 2012). It is indeed striking that evolutionarily distinct organisms use such similar mechanisms to maintain monoallelic expression, and trypanosomes prove to be no exception to this rule. Elucidating the mechanisms by which chromatin structure is used to maintain monoallelic expression in *T. brucei* is sure to provide insight into how these systems evolved across evolutionary time.

4.1 Chromatin State Facilitates Expression

Monoallelic expression is regulated, at least in part, through chromatin state. Early evidence for chromatin regulation of monoallelic *VSG* expression came from the finding that a transcriptionally active T7 promoter placed at the BES in bloodstream forms became inactivated upon differentiation to procyclic forms, implicating a chromatin remodeling mechanism that rendered the locus “inaccessible” in procyclic forms (Navarro et al. 1999). While the histones in *T. brucei* are extremely

diverged from those found in yeast and metazoans, there is ample evidence for the existence of both acetylation and methylation marks on the highly diverged N-terminal tails [reviewed in Figueiredo et al. (2009)]. The existence of histone variants has also been verified in *T. brucei*, and their location at sites of transcription initiation and termination indicates that they may play a role in the regulation of PolII transcription (Siegel et al. 2009). The fact that the active BES is depleted of nucleosomes when compared to its silent counterparts indicates that chromatin structure may also play a role in PolI transcriptional regulation and the maintenance of monoallelic expression (Figueiredo & Cross 2010). More definitive evidence comes from the fact that depletion of both H3 and the linker histone H1 results in higher transcription at inactive BES promoters (Povelones et al. 2012; Alsford & Horn 2012). The factors responsible for histone deposition have also been shown to be important for maintaining transcriptional repression at inactive BESs, as depletion of FACT, NLP (Narayanan et al. 2011), Asf1A, or Caf-1b (Alsford & Horn 2012) results in derepression at these sites. Finally, nucleosome remodelers have been implicated in this regulation, as inhibition of ISWI causes depression at the BES promoters as well as the *VSGs*. Thus, the presence or absence of the histones themselves appears to be important in maintaining monoallelic expression at BESs. With respect to maintaining high levels of transcription at the active BES, the high mobility group protein TDP1 is enriched at PolI transcription sites and its depletion results in lower levels of transcription at the active BES as well as the rDNA loci, which are also transcribed by PolI (Narayanan & Rudenko 2013). The transcription factor CITFA binds to PolI promoters and is also essential for both PolI-driven rDNA transcription and BES transcription (Nguyen et al. 2012).

In addition to their physical location on the DNA, the "code" provided by the covalent modifications on the histone tails provides an additional layer of regulation for transcriptional activity. It is thus not surprising that perturbing proteins that "write" or "erase' these covalent modifications in *T. brucei* result in disruption of monoallelic *VSG* expression. This appears to be the case for the histone deacetylase DAC3, as inhibition results in derepression of BES promoters (Wang et al. 2010). Additionally, inhibiting the methyltransferase Dot1B, which is responsible for modifying H3K79, results in a striking derepression phenotype, whereby two distinct VSG proteins can be detected on a single cell (Figueiredo et al. 2008). While histone acetyltransferase HAT1 and DAC1 maintain silencing of a telomeric reporter, disruption of these factors does not appear to lead to derepression of the *VSGs* themselves (Wang et al. 2010; Kawahara et al. 2008). The role for histone "writers" and "erasers" in regulating monoallelic expression in *T. brucei* has been well established, but not as much work has been done to investigate how histone "readers," proteins that recognize specific modifications on the histone tails and coordinate downstream transcriptional effects, affect monoallelic expression. However, an intriguing early report for a role for Bdf5 indicates that these proteins may be worthy of future study (Alsford & Horn 2012).

4.2 A Role for Telomeric Silencing

In yeast, telomeric silencing is controlled by the sirtuin family of proteins (Pillus & Rine 1989). As *VSGs* are located near telomeres, a similar silencing mechanism could be in play in *T. brucei*. Surprisingly, disrupting the sirtuin SIR2rp1 leads to derepression of a telomeric reporter while not affecting transcription of *VSGs* in inactive BESs, a phenotype similar to the one seen upon HAT1 and DAC1 inhibition (Alsford et al. 2007). However, the factor RAP1, which is recruited to telomeres through its interaction with telomere-bound TRF2, has been shown to be essential for silencing telomeric *VSGs*, but not *VSGs* located elsewhere in the genome (Yang et al. 2009). Surprisingly, when a telomere is removed through a programmed DNA break, telomere-mediated silencing is disrupted, but the *VSG* proximal to the removed telomere remains transcriptionally repressed, suggesting that there may be both telomere-dependent and independent modes of *VSG* silencing (Glover et al. 2007).

With respect to the broader nuclear architecture, the actively PolI transcribed BES is located in the ESB (Navarro & Gull 2001), a region distinct from the nucleolus where rDNA transcription takes place. Interestingly, upon differentiation to the procyclic form, the active BES promoter is repositioned to the nuclear envelope, an event that is followed by chromatin condensation (Landeira & Navarro 2007). While direct evidence is lacking for the nuclear location of the silent BESs, heterochromatic localization is implied by the fact that depletion of NUP-1, a large repetitive protein that in *T. brucei* appears to functionally substitute for mammalian lamins, results in derepression of inactive BESs (DuBois et al. 2012). Little is known about how the inheritance of the epigenetic state for the active and silent BESs is maintained across generations, but DNA replication processes may play a role. It appears that the separation of the sister chromatids belonging to the active BES is delayed relative to other regions of the genome and that this is at least partially mediated by cohesin. Depletion of cohesin results in a rapid transcriptional switch to a new BES, indicating that it may have a role in epigenetic VSG inheritance (Landeira et al. 2009). Studies have also implicated ORC1/CDC6 as being important to maintain the repressive state of the inactive BESs. ORC1/CD6 is functionally related to both Orc1 of the mammalian Origin Recognition Complex and CDC6, a factor that mediates interaction with the minichromosome maintenance (MCM) helicase. Knocking down ORC1/CDC6 derepresses BESs in both procyclic and bloodstream form cells, as well as metacyclic *VSGs* in procyclic form cells (Tiengwe et al. 2012). Attention was also drawn to the MCM helicase when it was found that an MCM binding protein (MCM-BP), a member of a complex including MCM4-8, is essential for maintaining repression of both BESs and the procyclin surface protein genes that cover the parasite during the procyclic life cycle stage (Kim et al. 2013). Since ORC1 has been shown in other systems to interact with cohesin, this might be the mechanism by which epigenetic inheritance is perturbed (Guillou et al. 2010).

In conclusion, *VSG* monoallelic expression is maintained by both the nuclear architecture and chromatin state of the DNA in which the *VSGs* reside, while DNA replication processes are likely involved in the placement of epigenetic marks and their inheritance. A recent comprehensive review on monoallelic *VSG* expression and other aspects of antigenic variation is available from Glover et al (2013b).

5 Concluding Remarks: Open Questions

Unlike many other pathogens, *T. brucei* exists in the bloodstream of the mammalian host, completely exposed to the host's immune system. In vivo, the parasite must switch its VSG coat in order to evade recognition by the immune system. By packing extremely densely on the parasite's surface, VSG also obscures invariant antigens that might be recognized by the host immune system. This packing poses an interesting requirement, however, on the expressed VSG repertoire: VSGs must be structurally conserved, in order to pack densely on the parasite surface, but also antigenically distinct, to evade detection by host antibody.

It turns out that the parasite balances these requirements quite well. Only two (partial) VSG crystal structures exist (Blum et al. 1993; Freymann et al. 1990), but these structures are superimposable despite having only 16 % sequence identity (Blum et al. 1993). Analysis of the sequence of *VSG* genes has identified conserved cysteine residues that likely play a role in maintaining VSG structure while allowing for significant sequence divergence (Carrington et al. 1991). Global analysis of the genomic repertoire of *VSG* genes, both full length and partial, has provided useful information regarding the depth of the genomic repertoire. However, how this repertoire is used during an infection and how that correlates with utilization of the antibody repertoire are key questions that remain open. Recently developed high-throughput approaches for examining expressed *VSGs* during infection show great promise for answering these questions (Mugnier et al. 2015).

The relative contribution of surface antibody clearance to immune evasion has not been determined. The notion that this process is, in fact, an immune evasion strategy is supported by observed differences in endocytosis rates between BF trypanosomes and those in other life cycle stages, which are not subject to similar immune pressure. BF trypanosomes induced to differentiate to the tsetse fly midgut-inhabiting procyclic form show downregulation of components of the endocytic pathway and a concurrent tenfold reduction in endocytosis rate (Natesan et al. 2007). A recent report indicated that mice infected with trypanosomes with a motility defect did not have a survival advantage over mice infected with motile trypanosomes (Kisalu et al. 2014). However, the trypanosomes used for these experiments were derived from the highly virulent Lister 427 strain, and all mice succumbed to infection within 2 weeks. Thus, the capability of trypanosomes defective in internalization of surface antibody to persist within the host has not been properly addressed in an in vivo model of chronic infection.

Furthermore, the molecular mechanisms of monoallelic expression as well as of switching remain to be fully elucidated. Switching is a semi-ordered process, and there is a loose relationship between the expressed *VSG* and the identities of the new *VSGs* that might arise as a result of switching. This relationship might be partially based on sequence homology or on nuclear proximity, but must have additional layers of dynamic complexity. For example, if *VSG-X* possesses a preference to pair with *VSG-Y*, this results not only in replacement of *VSG-X* but also in duplication of *VSG-Y*. The subsequent recombination event must now avoid the locus with highest homology, namely, the original donor. Though genetic experiments continue to be informative, we have very little information on the *trans*-acting factors that might mediate these events in the cell.

Finally, once a new *VSG* is recombined into, and expressed from the active BES, it must fully replace the old VSG coat. Because trypanosomes cannot survive without their coat, this requires the rapid turnover and replacement of 10 % of cellular protein. Yet, the half-life of VSG proteins is quite long [determined to be approximately 32.6 h in two independent studies Bülow et al. (1989) and Seyfang et al. (1990)], and this is likely a consequence of the trypanosome's need to consistently maintain a sufficient surface coat density. Unless the process of coat replacement is specifically accelerated as a consequence of switching, a switched cell could well express the original VSG on their surface for an extended period after expression of the new VSG begins, and would be at significant risk of antibody clearance. Because antigenic variation occurs in a prohibitively small fraction of trypanosome populations in vitro, this coat replacement process has not been directly observed. Thus, questions regarding the rate of VSG coat replacement in the context of switching remain, for the moment, unanswered; however, new tools can and will be applied to this and all the other problems we note, and we are looking forward to significant new knowledge in the near future. This new knowledge will directly impact treatment modalities for this neglected tropical disease, but beyond that, it is our hope that it will also (re)introduce *T. brucei* as a model system for studies well beyond VSG switching, such as telomere biology, DNA repair, and even cancer biology through the modeling of chromosomal translocations.

Acknowledgments The authors would like to thank the NIH and NIAID for their generous support from NIAID #R01AI097127 and NIAID #R01AI085973.

References

Aitcheson N, Talbot S, Shapiro J et al (2005) VSG switching in *Trypanosoma brucei*: antigenic variation analysed using RNAi in the absence of immune selection. Mol Microbiol 57: 1608–1622. doi:10.1111/j.1365-2958.2005.04795.x

Alarcon CM, Son HJ, Hall T, Donelson JE (1994) A monocistronic transcript for a trypanosome variant surface glycoprotein. Mol Cell Biol 14:5579–5591

Allen CL, Goulding D, Field MC (2003) Clathrin-mediated endocytosis is essential in *Trypanosoma brucei*. EMBO J 22:4991–5002. doi:10.1093/emboj/cdg481

Alsford S, Horn D (2012) Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in *Trypanosoma brucei*. Nucleic Acids Res 40:10150–10160. doi:10.1093/nar/gks813

Alsford S, Kawahara T, Isamah C, Horn D (2007) A sirtuin in the African trypanosome is involved in both DNA repair and telomeric gene silencing but is not required for antigenic variation. Mol Microbiol 63:724–736. doi:10.1111/j.1365-2958.2006.05553.x

Antia R, Nowak MA, Anderson RM (1996) Antigenic variation and the within-host dynamics of parasites. Proc Natl Acad Sci USA 93:985–989

Barry JD (1979) Capping of variable antigen on *Trypanosoma brucei*, and its immunological and biological significance. J Cell Sci 37:287–302

Barry JD, McCulloch R (2001) Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. Adv Parasitol 49:1–70

Berriman M, Ghedin E, Hertz-Fowler C et al (2005) The genome of the African trypanosome *Trypanosoma brucei*. Science 309:416–422. doi:10.1126/science.1112642

Black SJ, Guirnalda P, Frenkel D et al (2010) Induction and regulation of *Trypanosoma brucei* VSG-specific antibody responses. Parasitology 137:2041–2049. doi:10.1017/S003118200999165X

Blum ML, Down JA, Gurnett AM et al (1993) A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*. Nature 362:603–609. doi:10.1038/362603a0

Boothroyd CE, Dreesen O, Leonova T et al (2009) A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*. Nature 459:278–281. doi:10.1038/nature07982

Borst P, Ulbert S (2001) Control of VSG gene expression sites. Mol Biochem Parasitol 114:17–27

Borst P, Bitter W, Blundell PA et al (1998) Control of VSG gene expression sites in *Trypanosoma brucei*. Mol Biochem Parasitol 91:67–76

Brandenburg J, Schimanski B, Nogoceke E et al (2007) Multifunctional class I transcription in *Trypanosoma brucei* depends on a novel protein complex. EMBO J 26:4856–4866. doi:10.1038/sj.emboj.7601905

Bülow R, Nonnengässer C, Overath P (1989) Release of the variant surface glycoprotein during differentiation of bloodstream to procyclic forms of *Trypanosoma brucei*. Mol Biochem Parasitol 32:85–92

Carrington M, Miller N, Blum M et al (1991) Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. J Mol Biol 221:823–835

Chattopadhyay A, Jones NG, Nietlispach D et al (2005) Structure of the C-terminal domain from *Trypanosoma brucei* variant surface glycoprotein MITat1.2. J Biol Chem 280:7228–7235. doi:10.1074/jbc.M410787200

Chen J-M, Cooper DN, Chuzhanova N et al (2007) Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8:762–775. doi:10.1038/nrg2193

Cross GAM, Kim H-S, Wickstead B (2014) Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. Mol Biochem Parasitol 195:59–73. doi:10.1016/j.molbiopara.2014.06.004

Dagenais TR, Demick KP, Bangs JD et al (2009) T-cell responses to the trypanosome variant surface glycoprotein are not limited to hypervariable subregions. Infect Immun 77:141–151. doi:10.1128/IAI.00729-08

De Gaudenzi JG, Noé G, Campo VA et al (2011) Gene expression regulation in trypanosomatids. Essays Biochem 51:31–46. doi:10.1042/bse0510031

Dempsey WL, Mansfield JM (1983) Lymphocyte function in experimental African trypanosomiasis. V. Role of antibody and the mononuclear phagocyte system in variant-specific immunity. J Immunol 130:405–411

Diffley P (1983) Trypanosomal surface coat variant antigen causes polyclonal lymphocyte activation. J Immunol 131:1983–1986

Doyle JJ, Hirumi H, Hirumi K et al (1980) Antigenic variation in clones of animal-infective *Trypanosoma brucei* derived and maintained in vitro. Parasitology 80:359–369

Dreesen O, Cross GAM (2008) Telomere length in *Trypanosoma brucei*. Exp Parasitol 118:103–110. doi:10.1016/j.exppara.2007.07.016

Dreesen O, Li B, Cross GAM (2007) Telomere structure and function in trypanosomes: a proposal. Nat Rev Microbiol 5:70–75. doi:10.1038/nrmicro1577

DuBois KN, Alsford S, Holden JM et al (2012) NUP-1 Is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions. PLoS Biol 10, e1001287. doi:10.1371/journal.pbio.1001287

Engstler M, Thilo L, Weise F et al (2004) Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in *Trypanosoma brucei*. J Cell Sci 117:1105–1115. doi:10.1242/jcs.00938

Engstler M, Pfohl T, Herminghaus S et al (2007) Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. Cell 131:505–515. doi:10.1016/j.cell.2007.08.046

Ferrante A, Allison AC (1983) Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat. Parasite Immunol 5:491–498

Field MC, Lumb JH, Adung'a VO et al (2009) Macromolecular trafficking and immune evasion in African trypanosomes. Int Rev Cell Mol Biol 278:1–67. doi:10.1016/S1937-6448(09)78001-3

Figueiredo LM, Cross GAM (2010) Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. Eukaryot Cell 9:148–154. doi:10.1128/EC.00282-09

Figueiredo LM, Janzen CJ, Cross GAM (2008) A histone methyltransferase modulates antigenic variation in African trypanosomes. PLoS Biol 6, e161. doi:10.1371/journal.pbio.0060161

Figueiredo LM, Cross GAM, Janzen CJ (2009) Epigenetic regulation in African trypanosomes: a new kid on the block. Nat Rev Microbiol 7:504–513. doi:10.1038/nrmicro2149

Freymann D, Down J, Carrington M et al (1990) 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei*. J Mol Biol 216:141–160. doi:10.1016/S0022-2836(05)80066-X

Glover L, Alsford S, Beattie C, Horn D (2007) Deletion of a trypanosome telomere leads to loss of silencing and progressive loss of terminal DNA in the absence of cell cycle arrest. Nucleic Acids Res 35:872–880. doi:10.1093/nar/gkl1100

Glover L, McCulloch R, Horn D (2008) Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. Nucleic Acids Res 36:2608–2618. doi:10.1093/nar/gkn104

Glover L, Alsford S, Horn D (2013a) DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. PLoS Pathog 9, e1003260. doi:10.1371/journal.ppat.1003260

Glover L, Hutchinson S, Alsford S et al (2013b) Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in VSG expression control. Cell Microbiol 15:1984–1993. doi:10.1111/cmi.12215

Graham VS, Barry JD (1996) Is point mutagenesis a mechanism for antigenic variation in *Trypanosoma brucei*? Mol Biochem Parasitol 79:35–45

Grünfelder CG, Engstler M, Weise F et al (2003) Endocytosis of a glycosylphosphatidylinositol-anchored protein via clathrin-coated vesicles, sorting by default in endosomes, and exocytosis via RAB11-positive carriers. Mol Biol Cell 14:2029–2040. doi:10.1091/mbc.E02-10-0640

Guillou E, Ibarra A, Coulon V et al (2010) Cohesin organizes chromatin loops at DNA replication factories. Genes Dev 24:2812–2822. doi:10.1101/gad.608210

Hall JPJ, Wang H, Barry JD (2013) Mosaic VSGs and the scale of *Trypanosoma brucei* antigenic variation. PLoS Pathog 9, e1003502. doi:10.1371/journal.ppat.1003502

Hartley CL, McCulloch R (2008) *Trypanosoma brucei* BRCA2 acts in antigenic variation and has undergone a recent expansion in BRC repeat number that is important during homologous recombination. Mol Microbiol 68:1237–1251. doi:10.1111/j.1365-2958.2008.06230.x

Hertz CJ, Filutowicz H, Mansfield JM (1998) Resistance to the African trypanosomes is IFN-gamma dependent. J Immunol 161:6775–6783

Hertz-Fowler C, Figueiredo LM, Quail MA et al (2008) Telomeric expression sites are highly conserved in *Trypanosoma brucei*. PLoS One 3, e3527. doi:10.1371/journal.pone.0003527

Horn D (2014) Antigenic variation in African trypanosomes. Mol Biochem Parasitol 195: 123–129. doi:10.1016/j.molbiopara.2014.05.001

Horn D, Cross GA (1997) Analysis of *Trypanosoma brucei* vsg expression site switching in vitro. Mol Biochem Parasitol 84:189–201

Hovel-Miner GA, Boothroyd CE, Mugnier M et al (2012) Telomere length affects the frequency and mechanism of antigenic variation in *Trypanosoma brucei*. PLoS Pathog 8, e1002900. doi:10.1371/journal.ppat.1002900

Johnson PJ, Borst P (1986) Mapping of VSG genes on large expression-site chromosomes of *Trypanosoma brucei* separated by pulsed-field gradient electrophoresis. Gene 43:213–220

Kamper SM, Barbet AF (1992) Surface epitope variation via mosaic gene formation is potential key to long-term survival of *Trypanosoma brucei*. Mol Biochem Parasitol 53:33–44

Kawahara T, Siegel TN, Ingram AK et al (2008) Two essential MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric silencing in trypanosomes. Mol Microbiol 69:1054–1068. doi:10.1111/j.1365-2958.2008.06346.x

Kim H-S, Cross GAM (2010) TOPO3alpha influences antigenic variation by monitoring expression-site-associated VSG switching in *Trypanosoma brucei*. PLoS Pathog 6, e1000992. doi:10.1371/journal.ppat.1000992

Kim H-S, Cross GAM (2011) Identification of *Trypanosoma brucei* RMI1/BLAP75 homologue and its roles in antigenic variation. PLoS One 6, e25313. doi:10.1371/journal.pone.0025313

Kim H-S, Park SH, Günzl A, Cross GAM (2013) MCM-BP is required for repression of life-cycle specific genes transcribed by RNA polymerase I in the mammalian infectious form of *Trypanosoma brucei*. PLoS One 8, e57001. doi:10.1371/journal.pone.0057001

Kisalu NK, Langousis G, Bentolila LA et al (2014) Mouse infection and pathogenesis by *Trypanosoma brucei* motility mutants. Cell Microbiol 16:912–924. doi:10.1111/cmi.12244

Kobayashi I (1992) Mechanisms for gene conversion and homologous recombination: the double-strand break repair model and the successive half crossing-over model. Adv Biophys 28: 81–133

Kosinski RJ (1980) Antigenic variation in trypanosomes: a computer analysis of variant order. Parasitology 80:343–357

Lamont GS, Tucker RS, Cross GA (1986) Analysis of antigen switching rates in *Trypanosoma brucei*. Parasitology 92(Pt 2):355–367

Landeira D, Navarro M (2007) Nuclear repositioning of the VSG promoter during developmental silencing in *Trypanosoma brucei*. J Cell Biol 176:133–139. doi:10.1083/jcb.200607174

Landeira D, Bart J-M, Van Tyne D, Navarro M (2009) Cohesin regulates VSG monoallelic expression in trypanosomes. J Cell Biol 186:243–254. doi:10.1083/jcb.200902119

Liu AY, Michels PA, Bernards A, Borst P (1985) Trypanosome variant surface glycoprotein genes expressed early in infection. J Mol Biol 182:383–396

Lu Y, Hall T, Gay LS, Donelson JE (1993) Point mutations are associated with a gene duplication leading to the bloodstream reexpression of a trypanosome metacyclic VSG. Cell 72:397–406

Lu Y, Alarcon CM, Hall T et al (1994) A strand bias occurs in point mutations associated with variant surface glycoprotein gene conversion in *Trypanosoma rhodesiense*. Mol Cell Biol 14: 3971–3980

Lythgoe KA, Morrison LJ, Read AF, Barry JD (2007) Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. Proc Natl Acad Sci USA 104: 8095–8100. doi:10.1073/pnas.0606206104

Magklara A, Yen A, Colquitt BM et al (2011) An epigenetic signature for monoallelic olfactory receptor expression. Cell 145:555–570. doi:10.1016/j.cell.2011.03.040
Manna PT, Boehm C, Leung KF et al (2014) Life and times: synthesis, trafficking, and evolution of VSG. Trends Parasitol 30:251–258. doi:10.1016/j.pt.2014.03.004
Masterson WJ, Taylor D, Turner MJ (1988) Topologic analysis of the epitopes of a variant surface glycoprotein of *Trypanosoma brucei*. J Immunol 140:3194–3199
McCulloch R, Barry JD (1999) A role for RAD51 and homologous recombination in *Trypanosoma brucei* antigenic variation. Genes Dev 13:2875–2888
Miller EN, Turner MJ (1981) Analysis of antigenic types appearing in first relapse populations of clones of *Trypanosoma brucei*. Parasitology 82:63–80
Morrison LJ, Majiwa P, Read AF, Barry JD (2005) Probabilistic order in antigenic variation of *Trypanosoma brucei*. Int J Parasitol 35:961–972. doi:10.1016/j.ijpara.2005.05.004
Mugnier MR, Cross GAM, Papavasiliou FN (2015) The *in vivo* dynamics of antigenic variation in *Trypanosoma brucei*. Science 347(6229):1470–1473
Musoke AJ, Nantulya VM, Barbet AF et al (1981) Bovine immune response to African; trypanosomes: specific antibodies to variable surface glycoproteins of *Trypanosoma brucei*. Parasite Immunol 3:97–106
Myler PJ, Allison J, Agabian N, Stuart K (1984a) Antigenic variation in African trypanosomes by gene replacement or activation of alternate telomeres. Cell 39:203–211
Myler P, Nelson RG, Agabian N, Stuart K (1984b) Two mechanisms of expression of a predominant variant antigen gene of *Trypanosoma brucei*. Nature 309:282–284
Myler PJ, Allen AL, Agabian N, Stuart K (1985) Antigenic variation in clones of *Trypanosoma brucei* grown in immune-deficient mice. Infect Immun 47:684–690
Narayanan MS, Rudenko G (2013) TDP1 is an HMG chromatin protein facilitating RNA polymerase I transcription in African trypanosomes. Nucleic Acids Res 41:2981–2992. doi:10.1093/nar/gks1469
Narayanan MS, Kushwaha M, Ersfeld K et al (2011) NLP is a novel transcription regulator involved in VSG expression site control in *Trypanosoma brucei*. Nucleic Acids Res 39:2018–2031. doi:10.1093/nar/gkq950
Natesan SKA, Peacock L, Matthews K et al (2007) Activation of endocytosis as an adaptation to the mammalian host by trypanosomes. Eukaryot Cell 6:2029–2037. doi:10.1128/EC.00213-07
Navarro M, Gull K (2001) A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. Nature 414:759–763. doi:10.1038/414759a
Navarro M, Cross GA, Wirtz E (1999) *Trypanosoma brucei* variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. EMBO J 18:2265–2272. doi:10.1093/emboj/18.8.2265
Navarro M, Peñate X, Landeira D (2007) Nuclear architecture underlying gene expression in *Trypanosoma brucei*. Trends Microbiol 15:263–270. doi:10.1016/j.tim.2007.04.004
Nguyen TN, Nguyen BN, Lee JH et al (2012) Characterization of a novel class I transcription factor A (CITFA) subunit that is indispensable for transcription by the multifunctional RNA polymerase I of *Trypanosoma brucei*. Eukaryot Cell 11:1573–1581. doi:10.1128/EC.00250-12
Nguyen TN, Müller LSM, Park SH et al (2014) Promoter occupancy of the basal class I transcription factor A differs strongly between active and silent VSG expression sites in *Trypanosoma brucei*. Nucleic Acids Res 42:3164–3176. doi:10.1093/nar/gkt1301
O'Beirne C, Lowry CM, Voorheis HP (1998) Both IgM and IgG anti-VSG antibodies initiate a cycle of aggregation-disaggregation of bloodstream forms of *Trypanosoma brucei* without damage to the parasite. Mol Biochem Parasitol 91:165–193
Overath P, Engstler M (2004) Endocytosis, membrane recycling and sorting of GPI-anchored proteins: *Trypanosoma brucei* as a model system. Mol Microbiol 53:735–744. doi:10.1111/j.1365-2958.2004.04224.x
Pal A, Hall BS, Jeffries TR, Field MC (2003) Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in *Trypanosoma brucei*. Biochem J 374:443–451. doi:10.1042/BJ20030469

Pays E, Vanhamme L, Berberof M (1994) Genetic controls for the expression of surface antigens in African trypanosomes. Annu Rev Microbiol 48:25–52. doi:10.1146/annurev.mi.48.100194.000325

Peng G-H, Chen S (2007) Crx activates opsin transcription by recruiting HAT-containing co-activators and promoting histone acetylation. Hum Mol Genet 16:2433–2452. doi:10.1093/hmg/ddm200

Pillus L, Rine J (1989) Epigenetic inheritance of transcriptional states in *S. cerevisiae*. Cell 59:637–647

Povelones ML, Gluenz E, Dembek M et al (2012) Histone H1 plays a role in heterochromatin formation and VSG expression site silencing in *Trypanosoma brucei*. PLoS Pathog 8, e1003010. doi:10.1371/journal.ppat.1003010

Radwanska M, Guirnalda P, De Trez C et al (2008) Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses. PLoS Pathog 4, e1000078. doi:10.1371/journal.ppat.1000078

Recker M, Nee S, Bull PC et al (2004) Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. Nature 429:555–558. doi:10.1038/nature02486

Rister J, Desplan C, Vasiliauskas D (2013) Establishing and maintaining gene expression patterns: insights from sensory receptor patterning. Development 140:493–503. doi:10.1242/dev.079095

Robinson NP, Burman N, Melville SE, Barry JD (1999) Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. Mol Cell Biol 19:5839–5846

Robinson NP, McCulloch R, Conway C et al (2002) Inactivation of Mre11 does not affect VSG gene duplication mediated by homologous recombination in *Trypanosoma brucei*. J Biol Chem 277:26185–26193. doi:10.1074/jbc.M203205200

Rodriguez I (2013) Singular expression of olfactory receptor genes. Cell 155:274–277. doi:10.1016/j.cell.2013.09.032

Ross R, Thomson D (1910) A case of sleeping sickness studied by precise enumerative methods: regular periodical increase of the parasites disclosed. Proc R Soc Lond B 82:411–415. doi:10.1098/rspb.1910.0035

Roth C, Bringaud F, Layden RE et al (1989) Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. Proc Natl Acad Sci USA 86:9375–9379

Rudenko G, Blundell PA, Dirks-Mulder A et al (1995) A ribosomal DNA promoter replacing the promoter of a telomeric VSG gene expression site can be efficiently switched on and off in *T. brucei*. Cell 83:547–553

Schleifer KW, Filutowicz H, Schopf LR, Mansfield JM (1993) Characterization of T helper cell responses to the trypanosome variant surface glycoprotein. J Immunol 150:2910–2919

Seed JR, Edwards R, Sechelski J (1984) The ecology of antigenic variation. J Protozool 31:48–53

Seyfang A, Mecke D, Duszenko M (1990) Degradation, recycling, and shedding of *Trypanosoma brucei* variant surface glycoprotein. J Protozool 37:546–552

Shea C, Glass DJ, Parangi S, Van der Ploeg LH (1986) Variant surface glycoprotein gene expression site switches in *Trypanosoma brucei*. J Biol Chem 261:6056–6063

Siegel TN, Hekstra DR, Kemp LE et al (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. Genes Dev 23:1063–1076. doi:10.1101/gad.1790409

Stanhope-Baker P, Hudson KM, Shaffer AL et al (1996) Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. Cell 85:887–897

Stanne TM, Rudenko G (2010) Active VSG expression sites in *Trypanosoma brucei* are depleted of nucleosomes. Eukaryot Cell 9:136–147. doi:10.1128/EC.00281-09

Tiengwe C, Marcello L, Farr H et al (2012) Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. PLoS One 7, e32674. doi:10.1371/journal.pone.0032674

Turner CM (1997) The rate of antigenic variation in fly-transmitted and syringe-passaged infections of *Trypanosoma brucei*. FEMS Microbiol Lett 153:227–231

Turner CM, Barry JD (1989) High frequency of antigenic variation in *Trypanosoma brucei* rhodesiense infections. Parasitology 99(Pt 1):67–75

Van der Ploeg LH, Valerio D, De Lange T et al (1982) An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucleic Acids Res 10:5905–5923

Vasquez J-J, Hon C-C, Vanselow JT et al (2014) Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. Nucleic Acids Res 42:3623–3637. doi:10.1093/nar/gkt1386

Vettermann C, Schlissel MS (2010) Allelic exclusion of immunoglobulin genes: models and mechanisms. Immunol Rev 237:22–42. doi:10.1111/j.1600-065X.2010.00935.x

Volz JC, Bártfai R, Petter M et al (2012) PfSET10, a *Plasmodium falciparum* methyltransferase, maintains the active var gene in a poised state during parasite division. Cell Host Microbe 11:7–18. doi:10.1016/j.chom.2011.11.011

Wang Q-P, Kawahara T, Horn D (2010) Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. Mol Microbiol 77:1237–1245. doi:10.1111/j.1365-2958.2010.07284.x

Wickstead B, Ersfeld K, Gull K (2004) The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes. Genome Res 14: 1014–1024. doi:10.1101/gr.2227704

Yang X, Figueiredo LM, Espinal A et al (2009) RAP1 is essential for silencing telomeric variant surface glycoprotein genes in *Trypanosoma brucei*. Cell 137:99–109. doi:10.1016/j.cell.2009.01.037

Zomerdijk JC, Ouellette M, ten Asbroek AL et al (1990) The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. EMBO J 9:2791–2801

Zomerdijk JC, Kieft R, Borst P (1991) Efficient production of functional mRNA mediated by RNA polymerase I in *Trypanosoma brucei*. Nature 353:772–775. doi:10.1038/353772a0

Antigenic Variation in *Plasmodium falciparum*

Michaela Petter and Michael F. Duffy

Abstract *Plasmodium falciparum* is the protozoan parasite that causes most malaria-associated morbidity and mortality in humans with over 500,000 deaths annually. The disease symptoms are associated with repeated cycles of invasion and asexual multiplication inside red blood cells of the parasite. Partial, non-sterile immunity to *P. falciparum* malaria develops only after repeated infections and continuous exposure. The successful evasion of the human immune system relies on the large repertoire of antigenically diverse parasite proteins displayed on the red blood cell surface and on the merozoite membrane where they are exposed to the human immune system. Expression switching of these polymorphic proteins between asexual parasite generations provides an efficient mechanism to adapt to the changing environment in the host and to maintain chronic infection. This chapter discusses antigenic diversity and variation in the malaria parasite and our current understanding of the molecular mechanisms that direct the expression of these proteins.

Abbreviations

AMA-1	Apical membrane antigen 1
ATS	Acidic terminal sequence
CD36	Cluster of differentiation 36
CIDR	Cysteine-rich interdomain region
CM	Cerebral malaria
COPII	Coat protein complex II
DBL	Duffy binding-like
DC	Domain cassettes
dN/dS	Non-synonymous/synonymous polymorphism rates (dN/dS)
DNA	Deoxyribonucleic acid
EBA	Erythrocyte-binding antigen
EBL	Erythrocyte-binding-like

M. Petter • M.F. Duffy (✉)
Department of Medicine Royal Melbourne Hospital, Peter Doherty Institute, University of Melbourne, 792 Elizabeth Street, Melbourne, VIC 3010, Australia
e-mail: mpetter@unimelb.edu.au; mduffy@unimelb.edu.au

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_3

EPCR	Endothelial protein C receptor
ER	Endoplasmic reticulum
ETRAMPs	Early transcribed membrane proteins
exp1	Exported protein 1
GPI	Glycosylphosphatidylinositol
H2A.Z	Histone 2A.Z
H2B.Z	Histone 2B.Z
H3K36me3	Trimethylated histone H3 lysine 36
HB	Homology blocks
HSP	Heat shock protein
HT	Host-targeting signal
ICAM-1	Intercellular adhesion molecule 1
IE	Infected erythrocyte
IgG	Immunoglobulin G
KAHRP	Knob-associated histidine-rich protein
LANCL1	Lantibiotic synthetase component C-like 1
MAHRP2	Membrane-associated histidine-rich protein 2
MC	Maurer's clefts
MSP	Merozoite surface protein
MSRPs	Merozoite surface-related proteins
ncRNA	Non-coding RNA
PECAM1	Platelet endothelial cell adhesion molecule 1
PEXEL	*Plasmodium* export element
PfEMP1	*Plasmodium falciparum* erythrocyte membrane protein 1
PfEMP3	*Plasmodium falciparum* erythrocyte membrane protein 3
PfHP1	*P. falciparum* heterochromatin protein 1
PfPTP1	PfEMP1 trafficking protein 1
PfRh	*Plasmodium falciparum* reticulocyte-binding homologue
PfRNase II	*Plasmodium falciparum* ribonuclease II
PfSBP1	*Plasmodium falciparum* skeleton-binding protein 1
PfSIP2	*Plasmodium falciparum* SPE2-interacting protein
PTEX	*Plasmodium* translocon complex
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
RBC	Red blood cell
RBPs	Reticulocyte-binding proteins
RESA	Ring-infected erythrocyte surface protein
rif	Repetitive interspersed family
RNA	Ribonucleic acid
SPE2	Subtelomeric var promoter element 2
stevor	Subtelomeric open reading frame
TARE	Telomere-associated repeat element
TRX2	Thioredoxin-2
VSA	Variant surface antigen

1 Malaria Disease

Malaria is caused by infection with protozoan parasites of the genus *Plasmodium* and is transmitted by the bite of a female *Anopheles* mosquito. According to the 2014 World Malaria Report, presently 3.3 billion people (40 % of the world's population) in more than 100 countries live at risk of acquiring a malaria infection. An estimated 198 million cases occurred in 2013, more than 500,000 of which resulted in death. Although globally many tropical and subtropical regions are affected, the greater part of the burden rests on the sub-Saharan African countries, where 90 % of the malaria deaths occur (Arrowsmith et al. 2012).

Five species of *Plasmodium* can cause disease in humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and the primate species *P. knowlesi* (Singh et al. 2004). Of these, *P. falciparum* is responsible for most of the malaria-associated morbidity and mortality worldwide. The majority of infections are uncomplicated and cause fever, headache, muscle pain and other influenza-like symptoms. However a substantial number of patients suffer from complications potentially leading to fatality. Mainly children under the age of five are affected by severe disease and account for 78 % of malaria deaths (Arrowsmith et al. 2012).

In children, manifestations of the syndrome of severe malaria include cerebral malaria, severe anaemia, severe respiratory distress, renal failure, hypoglycaemia and pulmonary oedema, appearing alone or in combinations (Miller et al. 2013). After repeated infections with *P. falciparum*, individuals in malaria-endemic regions gradually develop semi-immunity, resulting in protection from clinical symptoms in adults (Bull et al. 1998). However, women become highly susceptible to the disease again when they get pregnant, especially in first or second pregnancies. Malaria substantially contributes to maternal death, stillbirth and miscarriage, as well as to complications like maternal anaemia and low birth weight babies (Umbers et al. 2011).

2 *Plasmodium* Life Cycle

*Plasmodium*parasites exhibit a complicated life cycle consisting of a sexual and an asexual phase. The infection begins with a bite by an infected female *Anopheles* mosquito. Together with the insect's saliva, *Plasmodium* sporozoites are injected into the subcutaneous tissues of the human host. The initial steps during infection have been documented in animal models by in vivo imaging techniques using transgenic fluorescent parasites [reviewed in Menard et al. (2013)]. Upon injection, sporozoites first breach the blood vessels to access the circulation from where they are rapidly transported to the liver (Amino et al. 2006). Here, sporozoites first transmigrate through Kupffer cells and several hepatocytes before they finally infect a hepatocyte and build up a parasitophorous vacuole (PV) (Mota et al. 2002). The developing hepatic schizont differentiates into thousands of

merozoites, which travel to the liver sinusoids in vesicles extruding from the infected hepatocyte, called merosomes (Sturm et al. 2006). It is presently unclear how merozoites are finally released from these structures; however, once freed into the bloodstream, they rapidly invade erythrocytes. Within 48 h, *P. falciparum* parasites multiply asexually through schizogony, giving rise to a new generation of merozoites. Upon release from the host cell, these initiate repeated cycles of red blood cell (RBC) infection that is responsible for the acute symptoms and pathology of the disease.

Some of the merozoites invading new RBCs differentiate into sexual forms which are called gametocytes. These are the transmissive stages, which, after ingestion by a blood-feeding mosquito, develop into female macrogametocytes and male microgametocytes. In the stomach of the insect, male and female gametes form and fuse to build a motile diploid zygote called ookinete. The ookinete penetrates the mosquito midgut wall and differentiates into an oocyst. When mature, this gives rise to a number of sporozoites capable of migrating into the salivary gland (Frischknecht et al. 2004), from which they are discharged into another human host during a blood meal.

3 Mechanisms of Immune Evasion

3.1 Antigenic Variation

Severe malaria is mainly a disease of young children and pregnant women. The protection of older children and adults in holoendemic areas is commonly understood as the result of slowly acquired immunity, which first shields from susceptibility to severe symptoms, and following continued exposure mediates protection from clinical disease (Bruce-Chwatt 1963; McGregor 1974).

Clinical immunity to malaria is developed only after repeated infections, because the parasite has evolved mechanisms to efficiently evade the host immune response. One strategy is the expression of variable antigens at the surface of the different life cycle stages that are exposed to host immunity. Antigenic variation describes the process of changing the proteins exposed to and recognised by the host immune system, helping the parasite to evade immune clearance and to establish long-lasting infections. This is achieved by switching the expression from one multigene family member to another during the course of an infection. Thus, the host is confronted with an ever-changing opponent, explaining why sterile immunity against malaria is rarely acquired. Gene products of multigene families are important candidate antigens for the development of an antimalaria vaccine because they are responsible for establishing chronic infections and because they elicit the non-sterile, protective immunity (Chen 2007; Hviid 2007). Two mechanisms contribute to antigenic diversity: (1) the presence of polymorphic alleles in

the parasite population and (2) the presence of multicopy gene families encoding variant surface antigens (VSA) (Ferreira et al. 2004).

3.2 Sequestration

Gene products of the *var*, *rif* and *stevor* multigene families, termed *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1), RIFIN (Fernandez et al. 1999; Helmby et al. 1993) and STEVOR (Niang et al. 2014), respectively, have been implicated in a second important immune evasion strategy, which is the capacity of infected erythrocytes to cytoadhere to the vascular endothelium and thereby sequester in the microvasculature of various organs, allowing the parasite to leave the circulation and to avoid immune clearance during passage through the spleen. Other cytoadhesive phenomena described for the parasite include the formation of rosettes together with uninfected erythrocytes or autoagglutination with other parasitised RBCs, bridged by platelets (Carlson et al. 1990; Pain et al. 2001; Roberts et al. 2000; Rowe et al. 1997; Udomsangpetch et al. 1989). It is generally accepted that the main physiological function of sequestration, rosetting and autoagglutination is the avoidance of splenic clearance, although these phenomena also significantly contribute to the pathology of the disease (Bray and Sinden 1979; Fremount and Rossan 1974; Langreth and Peterson 1985).

4 Immune Evasion Strategies of Intraerythrocytic Stages

4.1 Building a Home Inside the Host Cell

After invasion of the RBC, the parasite first undergoes a series of dramatic morphological changes (Gruring et al. 2011). It then develops within 48 h from the initial ring stage into a trophozoite and subsequently into the schizont stage, which harbours approximately 16–32 progeny merozoites. During the ring stage, which lasts about 24 h, the parasite shows no significant change in size and is still found in the circulation (Bannister et al. 2000, 2004). The rather inconspicuous ring stage is thought to serve as a lag phase during which host cell modifications are installed that are necessary for the export and surface expression of parasite proteins involved in immune evasion, reshaping the interface between the infected cell and the host (Spielmann et al. 2006). These proteins mediate what is possibly the most remarkable feature differentiating trophozoites and schizonts from the circulating ring stages, namely, their ability to sequester in the microvasculature of various organs.

The parasitophorous vacuole membrane (PVM) in which the parasite becomes engulfed during invasion of an erythrocyte originates from the erythrocyte

membrane, but is modified by the pathogen (Ward et al. 1993; Atkinson et al. 1988). Major erythrocyte proteins such as band 3, glycophorin A or cytoskeletal proteins are absent from the PVM, whilst detergent-resistant membrane raft proteins are recruited to the vacuole (Atkinson et al. 1988; Lauer et al. 2000; Murphy et al. 2004; Nagao et al. 2002). Moreover, parasite proteins such as early transcribed membrane proteins (ETRAMPs) or exported proteins 1 and 2 (Exp1 and Exp2) are inserted (Fischer et al. 1998; Gunther et al. 1991; Spielmann et al. 2003). From the PVM, a network of tubular structures designated as the tubovesicular network extrudes into the erythrocyte cytoplasm (Elmendorf and Haldar 1993, 1994). Whilst this network is proposed to serve the parasite for nutrient uptake (Lauer et al. 1997, 2000), a second system apparent as cisterna-shaped membranous vesicles, called Maurer's clefts (MCs), was described as constituting a secretory organelle shuttling proteins from the parasite to the host erythrocyte surface (Haeggstrom et al. 2004, 2007; Hinterberg et al. 1994; Petter et al. 2007; Wickert et al. 2003; Winter et al. 2005). Whether the two systems are actually constituents of the same or of different membrane assemblies remains controversial (Lanzer et al. 2006; Marti et al. 2005).

There is evidence suggesting that MCs are anchored to the host erythrocyte skeleton, since these structures remain associated with the host cell ghost after schizont rupture and merozoite release. One component thought to be mediating this interaction is the *P. falciparum* skeleton-binding protein 1 (PfSBP1) binding to the host protein lantibiotic synthetase component C-like protein (LANCL1) (Blisnick et al. 2000, 2005). However, infected erythrocyte (IE) morphology changed little when PfSBP1 was knocked out, challenging its role as a structural protein. More importantly, the knockout parasites exhibited a significant constraint in their ability to cytoadhere, identifying PfSBP1 as an important player involved in virulence mechanisms of the malaria parasite (Cooke et al. 2006). Candidates for mediating the docking of MCs to the host cell membrane are electron-dense tethers which contain the membrane-associated histidine-rich protein 2 (MAHRP2) (Pachlatko et al. 2010; McMillan et al. 2013).

During the second half of intraerythrocytic maturation, hundreds of electron-dense knob-like structures begin decorating the IE surface (Gruenberg et al. 1983). The main proteinaceous component of these knobs is knob-associated histidine-rich protein (KAHRP) which interacts with the host cell cytoskeleton component spectrin (Pei et al. 2005). The knobs represent contact points between the IE and other cells they adhere to (Oh et al. 2000; Rug et al. 2006; Waller et al. 1999; Wickham et al. 2001). Loss of KAHRP leads to knobless parasitized cells and a marked reduction in cytoadherence (Biggs et al. 1989; Crabb et al. 1997; Pologe and Ravetch 1986). Similar results were obtained after targeted deletion of a second major knob component, termed *Plasmodium falciparum* erythrocyte membrane protein 3 (PfEMP3) (Waterkeyn et al. 2000).

4.2 Modification of the Host Cell Membrane

Living inside a PV within a metabolically silent cell poses a dilemma to the malaria parasite. It cannot hitchhike on any existing protein transport machinery present in the host cell, as other pathogens do, but needs to install its own trafficking machinery. It is known that hydrophobic amino-terminal or recessed signal peptides present in many of the proteins encoded in the plasmodial genome can mediate secretion across the parasite membrane into the PV, but the events allowing passage of proteins beyond the PVM into the erythrocyte cytoplasm and to the erythrocyte membrane are just being uncovered. Two studies have independently reported the presence of a functional motif that is required for the export of proteins across the PVM, designated *Plasmodium* export element (PEXEL) (Marti et al. 2004) or host-targeting (HT) signal (Hiller et al. 2004) with the consensus sequence R/KxLxQ/E. The consequently predicted "exportome" includes both soluble proteins present in the erythrocyte cytosol and membrane-bound proteins associated with the MC and the erythrocyte membrane (Marti et al. 2005). The endoplasmic reticulum (ER) resident aspartic protease plasmepsin V is critical for protein export and processes the PEXEL (Russo et al. 2010; Boddey et al. 2010). The PVM contains the *Plasmodium* translocon complex (PTEX), which mediates the crossing of exported parasite proteins into the host cell cytoplasm. The complex is anchored in the PVM by Exp2 and contains the essential chaperone heat shock protein 101 (HSP101) (Beck et al. 2014; Elsworth et al. 2014), non-essential thioredoxin-2 (TRX2) (Matthews et al. 2013) and at least two other essential components termed PTEX150 and PTEX88 (de Koning-Ward et al. 2009).

How membrane-associated proteins are transported through the erythrocyte cytoplasm remains controversial. Two major models have been suggested. The first proposes lateral diffusion of membrane proteins along a continuous MC network (Wickert et al. 2003, 2004); however, this has been challenged by the fact that MC biogenesis is completed before export of many proteins begins (Gruring et al. 2011). The other model favours a vesicular pathway, dividing the intraerythrocytic membrane compartment into functionally distinct modules allowing the assembly, for example, of the cytoadherence complex consisting of PfEMP1 and KAHRP at certain sites (Haeggstrom et al. 2007; Wickham et al. 2001). In support of this, *Plasmodium* homologues of several components of the vesicle-mediated eukaryotic export machinery such as elements of the coat protein complex II (COPII) have been found in association with the MC, showing parallels between mechanisms involved in protein trafficking between endoplasmic reticulum and Golgi apparatus and trafficking in the erythrocyte cytoplasm (Adisa et al. 2001; Albano et al. 1999). High-resolution imaging techniques have been employed to further decipher this question and showed that MCs are indeed not linked to the PVM and the erythrocyte membrane (McMillan et al. 2013; Gruring et al. 2011). Electron tomography identified novel vesicular intermediate trafficking compartments (Hanssen et al. 2010) and the discovery of mobile HSP40/HSP70x-chaperone containing so-called J-dots in the host cell cytosol further supports a

model whereby proteins are trafficked to and from the MC in vesicular structures (Kulzer et al. 2010, 2012).

4.3 Variant Surface Antigens of the Infected Erythrocyte

4.3.1 PfEMP1, Adhesion and Malaria Disease

PfEMP1 is the immunodominant, variant antigen of the intraerythrocytic stage of *P. falciparum*. PfEMP1 is exported from the parasite and expressed in clusters at the knob-like structures on the surface of the IE where it mediates adhesion to various host receptors expressed on erythrocytes, platelets, endothelium lining the microvasculature, placental syncytiotrophoblasts and extracellular matrix within the placental intervillous space. The pleiotropic adhesion phenotypes mediated by PfEMP1 provide a diverse array of potential binding sites that allow the mature IE to sequester away from the circulation and thus avoid destruction in the spleen. PfEMP1 is anchored at the erythrocyte surface by interactions between its conserved, intracellular, acidic (C) terminal sequence (ATS) and various host cytoskeleton proteins and parasite proteins including KAHRP that together form the knob-like structures. Concentration of PfEMP1 at the knobs is required for effective adhesion of the IE to host receptors (Crabb et al. 1997; Cooke et al. 1994) and, in the absence of selective pressure exerted by splenic clearance, parasites during in vitro culture frequently delete KAHRP with no deleterious effect on parasite growth. Parasites isolated from splenectomised patients also have decreased PfEMP1 expression (Bachmann et al. 2009). The adhesion of parasites to host receptors is a major virulence determinant of *P. falciparum* (White et al. 2013) and indeed is a defining characteristic that separates it from the other four species of *Plasmodium* that infect humans with far lower mortality. The actual mechanisms of pathogenesis deriving from *P. falciparum* sequestration remain uncertain, but vascular obstruction, decreased perfusion, direct stimulation of immunopathology via binding to immune signalling receptors and indirectly through concentration of parasites in organs affected by immunopathology are all proposed to play a role.

Although the relative contribution of sequestration to pathogenesis remains contentious, sequestration in multiple organs including the brain, kidney, liver, heart, lung and skin has been observed at post-mortem of patients who died of severe malaria (Pongponratn et al. 2003; Silamut et al. 1999; Nguansangiam et al. 2007; Prommano et al. 2005; MacPherson et al. 1985), and sequestration in the placenta is the defining characteristic of parasites that cause malaria in pregnancy (Bray and Sinden 1979). Cerebral malaria (CM) is the malaria syndrome with the highest mortality with a case fatality rate of up to 18 %. The key histopathological feature of fatal human CM is sequestration of IEs in brain microvasculature with or without immune cells (Taylor et al. 2004) which results in a massive 28- to 51-fold enrichment of IEs in the brain compared to the circulation (MacPherson et al. 1985; Pongponratn et al. 2003; Silamut

et al. 1999). Although sequestration of IEs in the brain is necessary for CM, it is not by itself sufficient, and host inflammatory responses (Combes et al. 2004a; Hunt and Grau 2003; van der Heyde et al. 2006; Weinberg et al. 2008; Weiser et al. 2007) [reviewed in Shikani et al. (2012)] and parasite toxins (Pamplona et al. 2007) also contribute to its pathogenesis. Sequestered IEs impair the integrity of the blood-brain barrier in vitro (Tripathi et al. 2007) and are associated with regions of blood-brain barrier disruption post-mortem (Dorovini-Zis et al. 2011). Through dysregulation of inflammatory cytokines (Shikani et al. 2012) and vasomodulators such as angiopoietin-2 (Yeo et al. 2008), sequestered IEs can indirectly cause vasculopathy, but they may also directly induce apoptosis of the brain endothelial cells (Siau et al. 2007).

var Genes

PfEMP1s are encoded by the *var* multigene family (Smith et al. 1995; Baruch et al. 1995; Su et al. 1995). There are approximately 60 copies of *var* genes per parasite genome (Gardner et al. 2002b; Rask et al. 2010). The *var* genes are different within and between parasite clones leading to tremendous diversity (Rask et al. 2010). *Var* genes are defined by their upstream sequence, into groups A, B, C and E (Gardner et al. 2002a; Lavstsen et al. 2003). Groups A, B and C comprise approximately 20 %, 60 % and 20 %, respectively, of the genomic repertoire of *var* genes (Rask et al. 2010) (Fig. 1a); group E consists of a single *var* gene called *var2csa* which encodes adhesion to the placental receptor chondroitin sulphate A and is probably only involved in malaria during pregnancy (Salanti et al. 2003). *Var* genes are located in clusters, mostly at subtelomeric sites but also at some chromosome-internal sites (Gardner et al. 2002a).

Parasites primarily transcribe a single *var* gene at one time, typically early in the 48 h intraerythrocytic life cycle. Thus individual infected erythrocytes express only a single antigenic variant of PfEMP1 on their surface. However, parasites switch between the *var* genes they transcribe and the switched expression is heritable leading to variegated expression of *var* genes within a clonal population. By switching to expression of a novel variant *var* gene, *P. falciparum* can avoid the immunity acquired to the previously expressed PfEMP1 and thus maintain a chronic infection through the process of clonal antigenic variation (Biggs et al. 1991).

Phylogeny of PfEMP1 Domains

PfEMP1s contain combinations of Duffy binding-like (DBL) domains and cysteine-rich interdomain regions (CIDRs) (Smith et al. 2000b), subtypes of which adhere to different host receptors (Fig. 1a) [reviewed in Rowe et al. (2009) and Smith et al. (2013)]. PfEMP1 DBL domains are members of a domain family first described in erythrocyte-binding proteins that are involved in the invasion of uninfected erythrocytes by multiple species of *Plasmodium* (Adams et al. 1992; Chitnis and Miller 1994). CIDR domains also have subtypes with different adhesion specificities (Robinson et al. 2003; Turner et al. 2013). DBL and CIDR domains are defined by regions of semi-conserved sequence into DBLα, β, γ, δ, ε, ζ and x types and CIDRα, β, γ and δ types (Smith et al. 2000b; Lavstsen et al. 2003;

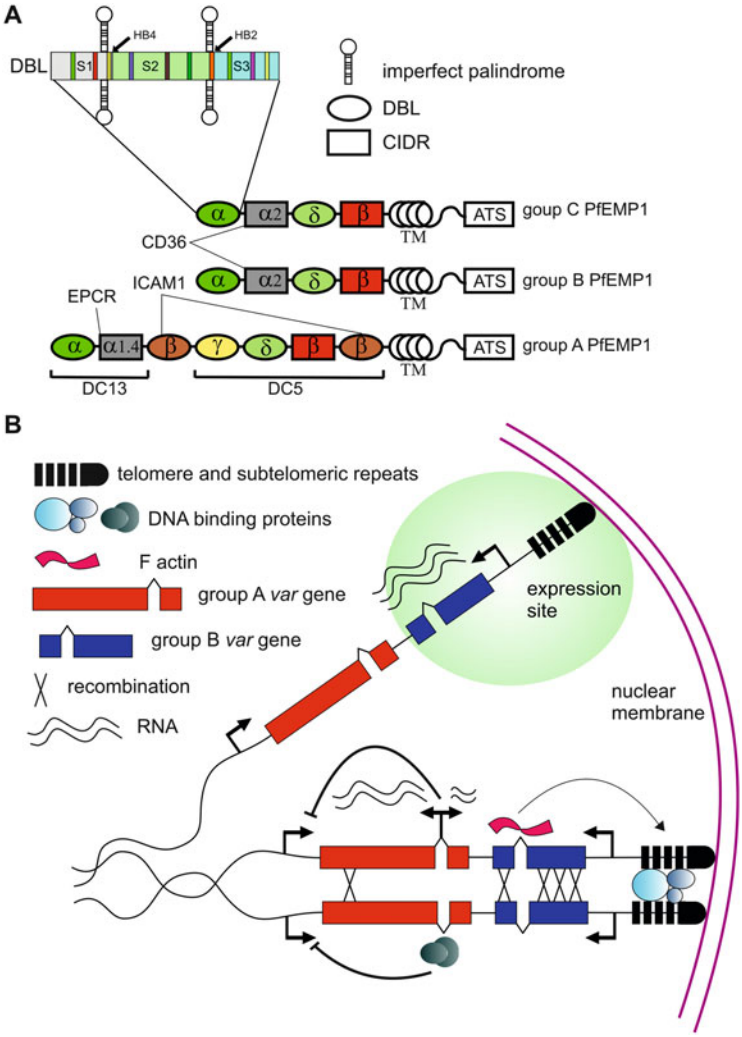


Fig. 1 *P. falciparum var* gene structure, sites of recombination and regulation of expression. (**a**) Domain organisation of PfEMP1 variants. A representative DBL domain with the three subdomains S1, S2 and S3 indicated; the thin coloured bars represent homology blocks (HBs). HB4 and HB2 near the boundaries of subdomain S2 are frequently close to imperfect palindromic repeats that promote recombination. Combinations of DBL and CIDR domains form the N-terminal extracellular region of PfEMP1 followed by the transmembrane domain (TM) and the intracellular conserved acidic terminal sequence (ATS). Different domains can bind different host receptors, e.g., CIDRα1.4 can bind EPCR, CIDRα2 can bind CD36, and DBLβ can bind ICAM1. Conserved arrangements of DBLs and CIDRs are called domain cassettes (DCs). Group A PfEMP1s are larger and more conserved than groups B and C. Groups A and B *var* genes are subtelomeric, with group B closest to the telomere and transcribed towards the centromere whilst group A *var* genes are transcribed in the opposite direction. (**b**) Nuclear organisation of *var* genes contributes to their transcriptional regulation and to preferential intra-group *var* gene recombination. Subtelomeric and chromosome-internal (not shown) *var* genes are all recruited to clusters at the nuclear periphery by a mechanism that involves actin interacting indirectly with the *var* intron. The clusters are maintained by unidentified proteins that bind the subtelomeric repeat sequences.

Rask et al. 2010). Multiple phylogenetic sub-classifications of domains have been identified within these broader types (Rask et al. 2010). Broad division of all *var* genes into four contiguous components revealed that some domains were invariably arranged in a conserved order, e.g. component 1 consists of a semi-conserved N-terminal sequence-DBLα-CIDR arrangement and is found in 95 % of *var* genes (Rask et al. 2010), whilst other domains also associated in arrangements but at a lower frequency. Eighteen contiguous arrangements of domain subtypes were also conserved both within and across component boundaries; these were referred to as domain cassettes (DCs) (Fig. 1a). Crystal structures have elucidated, conserved, critical DBL and CIDR residues for adhesion (Singh et al. 2006; Higgins 2008; Klein et al. 2008), whilst human antibodies have been shown to target the variable sequence of DBL domains (Barfod et al. 2007).

The Roles of PfEMP1 Domains and *var* Gene Groups in Disease and Immunity

Naturally acquired immunity to malaria is non-sterile and is dependent on acquisition of antibodies to a range of PfEMP1 variants (Marsh et al. 1989; Newbold et al. 1992; Chan et al. 2012b). Parasites that cause severe disease seem to express a conserved subset of variant antigens on the IE surface that are encountered earlier in life and that are thus more widely recognised by sera from semi-immune children than parasites causing mild disease (Bull et al. 2000; Nielsen et al. 2002). This probably explains why immunity to severe malaria is acquired much more rapidly than immunity to mild malaria (Marsh et al. 1989). Conserved variants that caused severe disease may have scope as vaccine candidates.

The expression of particular DBLα sequences in severe malaria suggests severe disease may be caused by a restricted subset of PfEMP1s (Normark et al. 2007; Warimwe et al. 2009). Increased expression of group A and B *var* genes has been associated with clinical malaria in Papua New Guinea (PNG) (Kaestli et al. 2006; Falk et al. 2009) and severe malaria in Africa (Rottmann et al. 2006). The binding of infected erythrocytes to uninfected erythrocytes is called rosetting, and in Africa but not Asia, this phenotype is associated with severe disease (Rowe et al. 2009) and increased expression of group A *var* genes (Kaestli et al. 2006; Normark et al. 2007; Warimwe et al. 2012). Cerebral malaria in Africa was associated with increased expression of group A (Kyriacou et al. 2006; Warimwe et al. 2009, 2012)

Fig. 1 (continued) The proximity of the *var* genes in tightly bound clusters has been proposed to both promote recombination within *var* groups and impede recombination between the opposite facing group A and B *var* genes. Group A *var* genes recombine less frequently than group B and group C *var* genes. The *var* genes at the perinuclear clusters are maintained in a transcriptionally silent state in facultative heterochromatin. Cis sequences play an important role in maintaining *var* gene silencing and the promoter activity of the *var* intron is critical for proper *var* gene silencing; the *var* intron binds the same unidentified proteins as the *var* promoter suggesting that the two sequences loop to interact. To be transcribed, a single *var* gene leaves the cluster and moves to an expression site where it is maintained in a transcriptionally active, euchromatic state over multiple generations

or group B (Kalmbach et al. 2010) *var* genes. Group A var genes were also expressed by parasites recognised by antibodies from semi-immune children (Jensen et al. 2004). Group A and B *var* genes dominated during early infection of a naive individual (Wang et al. 2009) and more individuals including younger children have antibodies to group A PfEMP1s than group B or C (Cham et al. 2009). In summary, there is considerable evidence indicating that group A and B *var* genes are preferentially expressed by parasites that infect non- or semi-immune individuals and that cause severe malaria disease.

Some group A and B PfEMP1s can bind to the endothelial receptor, intercellular adhesion molecule 1 (ICAM-1), by their DBLβ domains (Smith et al. 2000a; Howell et al. 2008; Oleinikov et al. 2009). Adhesion to ICAM-1 has been associated with cerebral malaria in one (Ochola et al. 2011) but not other studies (Newbold et al. 1997; Rogerson et al. 1999). In adults with cerebral malaria, ICAM-1 expression was upregulated in brain endothelium and, together with E-selectin, co-localised with sequestered IEs (Turner et al. 1994). Binding to ICAM-1 causes signalling and increased endothelial cell adhesion molecule expression (Tripathi et al. 2006) thus potentiating further adhesion of IEs, immune cells and platelets.

Recently IEs expressing PfEMP1 domain DCs 8 and 13 were shown to bind brain endothelial cells (Avril et al. 2012; Claessens et al. 2012) via endothelial protein C receptor (EPCR) which also bound IEs from African children with severe malaria (Turner et al. 2013). DC 8 PfEMP1s were widely expressed at high levels in African severe malaria cases but DC 13 and DC 5 PfEMP1s were also expressed at elevated levels in some children with severe disease (Lavstsen et al. 2012) DCs 5 and 13 are primarily from group A *var* genes and DC 8 is primarily group B (Rask et al. 2010). DC 8 and DC 13 PfEMP1s bind EPCR via their CIDRα1.1–1.8 domains (with the exclusion of CIDRα1.2 and 1.3) (Turner et al. 2013; Lau et al. 2015) (Fig. 1). Sera from African children with clinical malaria recognise DC 8- and DC 13-expressing parasites at higher levels than non-DC 8- and DC 13-expressing parasites, but it is unclear whether severe malaria specifically induces antibodies to DC 8 and 13 (Avril et al. 2012; Claessens et al. 2012). Curiously EPCR has also been reported as depleted at sites of IE sequestration and endothelial pathology in the brain (Moxon et al. 2013).

High levels of DC 5 *var* gene expression were also detected in isolates causing severe disease, but only together with expression of either DC 8 or 13 *var* genes (Lavstsen et al. 2012). Parasites expressing DC 5 PfEMP1s bind platelet endothelial cell adhesion molecule 1 (PECAM1) (Berger et al. 2013). Brain endothelium can be induced to express PECAM1 by elevated levels of platelet microparticles in cerebral malaria (Combes et al. 2004b); however, parasite adhesion to PECAM1 is not restricted to severe malaria (Heddini et al. 2001). Semi-immune children have antibodies that bind erythrocytes infected with parasites expressing a DC 5 PfEMP1 (Jensen et al. 2004) and antibodies reactive with DC 5 PfEMP1 also correlated with protection from malaria episodes (Magistrado et al. 2007). Children from areas of moderate and high transmission acquired antibodies to a DC 5 PfEMP1 domain more rapidly than to 47 other PfEMP1 domains (Cham et al. 2009), suggesting that DC 5 *var* genes were expressed by parasites in the earliest infections. Thus several

promising candidates have emerged as the restricted population of PfEMP1s responsible for severe malaria, but the relative contribution of groups A and B and DCs 8, 13 and 5 remains unclear.

4.3.2 *Rif* and *Stevor* Genes

The *rif* and *stevor* gene families are mainly found in the subtelomeric regions of all 14 *P. falciparum* chromosomes. Here, they are clustered together with the *var* genes and are arranged in repeats of several genes either on the same or opposing strands (Gardner et al. 2002a). About 185 *rif* genes and 66 *stevor* genes are encoded in the genome of *P. falciparum*. The gene products, designated RIFINs and STEVORs, share a common architecture. They are organised in two exons, the first of which is predicted to code for a signal peptide, whilst the second represents the main body of the protein, consisting of semi-conserved as well as highly variable parts (Cheng et al. 1998; Kyes et al. 1999). RIFINs have been subgrouped into A-type and B-type RIFINs which differ by a conserved 25 amino acid motif in the semi-conserved domain that is only present in A-type RIFINs (Joannin et al. 2008).

The initial proposed RIFIN topology was two transmembrane domains flanking the polymorphic middle section and exposing it in a loop at the IE surface (Cheng et al. 1998). Recently this has been challenged for A-type RIFINs through the use of better bioinformatic prediction tools which detect only a single transmembrane domain (Petter et al. 2008; Bultrini et al. 2009). In support of a single transmembrane domain topology for STEVORs, antibodies to the semi-conserved N-terminal region of STEVORs label the surface of the infected erythrocyte, indicating that only the short highly positively charged C-terminus is located inside the erythrocyte (Niang et al. 2009).

Based on their similar organisation, RIFINs and STEVORs are postulated to belong to a larger superfamily, extending over species boundaries (Janssen et al. 2004). Thus, a superfamily of *Plasmodium* interspersed repeats (*pir*) was defined including similarly organised protein families present in many other *Plasmodium* species, for example, in *P. vivax* (*vir*), as well as in several rodent and primate malaria parasites (*kir*, *bir*, *cir*, *yir*). Homology between the different members of this superfamily is based on topographic conservation as well as on the presence of several short conserved motifs, indicating a common evolutionary ancestry (Janssen et al. 2004). *Rif* and *stevor* gene copies are expanded in *P. reichenowi*, a close relative of *P. falciparum* which infects chimpanzees, which indicates that they might be linked to host specificity (Otto et al. 2014). STEVORs have recently been functionally linked to host cell invasion, pointing to a possible mechanism by which these proteins could mediate host specificity (Niang et al. 2014).

Both RIFINs and STEVORs contain a PEXEL/HT motif (Hiller et al. 2004; Marti et al. 2004), which labels them as exported proteins. This characteristic and their large number and hypervariability support their suggested role in antigenic variation. Both RIFINs and STEVORs require processing by plasmepsin V for

export across the PVM (Boddey et al. 2013). There are differences in the mechanisms facilitating their trafficking to the MC, as STEVORs, along with PfEMP1, require the MC protein PfEMP1 trafficking protein 1 (PfPTP1), whereas RIFIN proteins reach MCs even in the absence of PfPTP1 (Rug et al. 2014). Differences in trafficking have been observed for different RIFIN variants, as certain B-type variants seem not to be exported into the erythrocyte (Petter et al. 2007).

The first report on *rif* genes dates back to the year 1988, when Weber and colleagues identified a repetitive interspersed element in the *P. falciparum* genome, which they first thought to be a transposable element (Weber 1988). The *P. falciparum* genome sequence revealed a total of 149 gene copies coding for proteins with a predicted molecular weight of 30–45 kDa (Gardner et al. 2002a). The *rif* and *stevor* repertoires are largely diverse between three different *P. falciparum* strains, although a small number of strain-transcendent variants exist (Claessens et al. 2010; Blythe et al. 2009). Northern blot analyses showed that *rif* transcription peaks around 18–24 h post-invasion in asexual parasites, coincident with the decrease of *var* gene transcription (Kyes et al. 1999, 2000; Bachmann et al. 2012), followed by upregulation of *stevor* gene expression around 26–30 h post-invasion (Kaviratne et al. 2002). A study measuring VSA transcripts in fresh isolates compared to lab-adapted strains recapitulated this pattern and monitored a second peak in transcription for both A-type *rif* and *stevor* genes in early ring stage parasites. In addition, a rapid loss in global *rif* or *stevor* transcript abundance during culture adaptation was observed (Bachmann et al. 2009). Both STEVORs and RIFINs are expressed in a clonal fashion and are subject to switches in gene expression (Fernandez et al. 1999; Lavazec et al. 2007), supporting a role in antigenic variation. However, RIFINs of A-type and B-type can be expressed in a single cell simultaneously (Petter et al. 2007).

Being similar in size to a previously reported panel of proteins identified in strongly rosetting parasite strains, RIFINs were first referred to as rosettins (Helmby et al. 1993). RIFINs were also reported to be expressed at higher levels on the surface of rosetting than on non-rosetting parasites (Fernandez et al. 1999). However, surface trypsinisation experiments and the identification of PfEMP1 as the parasite ligand for rosetting challenged a direct involvement of RIFINs in rosette formation (Fernandez et al. 1999; Kyes et al. 1999; Rowe et al. 1997). Interestingly, there is direct evidence for an involvement of STEVOR in rosetting, and this seems to be mediated by binding to glycophorin C (Niang et al. 2014). Moreover, STEVOR expression influences the mechanical properties of the erythrocyte membrane in asexual- and sexual-stage parasites, making them more rigid which possibly enhances sequestration of immature gametocytes and IE (Sanyal et al. 2012; Tiburcio et al. 2012). A role of RIFINs and STEVORs in sequestration as a mechanism of immune evasion was further supported by a case report of a patient who after splenectomy experienced a malaria relapse with an expansion of parasites that had lost transcription of PfEMP1, STEVOR and A-type RIFINs (Bachmann et al. 2009).

Several studies have documented anti-RIFIN and anti-STEVOR immune responses (Schreiber et al. 2008), showing that anti-RIFIN antibodies are associated

with a stable response over time and with rapid clearance of parasites from the circulation (Abdel-Latif et al. 2002, 2003, 2004). Dissection of the immune response according to IgG subclasses revealed that anti-RIFIN IgG2 antibodies occur predominantly in cerebral malaria patients, indicative of important functions in malaria pathology (Schreiber et al. 2006). In volunteer infections, antibodies to RIFINs are rapidly acquired, supporting the idea that their variability serves a function in immune evasion (Turner et al. 2011). However, it appears that PfEMP1 is the major immune target on the red blood cell surface and RIFINs and STEVORs comprise minor epitopes (Chan et al. 2012a).

5 Immune Evasion Strategies of Merozoites

Plasmodium parasites are classified as members of the phylum Apicomplexa because their invasive stages, including merozoites, sporozoites and ookinetes, are hallmarked by a unique collection of organelles found at their anterior end. In merozoites, these include a pair of pear-shaped membrane-bound rhoptries, several smaller micronemes which are attached to the rhoptry duct and a number of vesicles described as dense granules that are distributed in the cytoplasm (Cowman et al. 2012). This apical complex plays a pivotal role during host cell invasion during which the contents are sequentially released. Many of the invasion proteins are exposed to the host immune system during this process, making the parasite vulnerable to complement-mediated lysis or opsonic phagocytosis and exerting diversifying selection pressure on these proteins which are prime targets of immunity (Persson et al. 2013; Richards et al. 2013).

Antibodies to several merozoite proteins have been shown to interfere with invasion and to protect from disease, thus making these antigens promising vaccine candidates. It is puzzling, though, how the parasite goes through several rounds of reinvasion without being cleared by neutralising antibodies elicited in preceding infections. One possible reason is the extremely short reinvasion time, estimated to be less than 60 s (Gilson and Crabb 2009). Another explanation is that many merozoite proteins such as the merozoite surface proteins (MSP) or erythrocyte-binding-like (EBL) proteins exist in several alleles or copies in the genome, showing a high degree of polymorphism (Holder et al. 1999; Khan et al. 2001).

5.1 Merozoite Surface Proteins

The initial contact with a RBC is mediated by MSPs covering the membrane of the merozoite as a dense coat. The best characterised is the essential MSP-1, which undergoes extensive proteolytic processing around the time of merozoite release (McBride and Heidrich 1987; Heidrich et al. 1983). The products, including a 42 kDa glycosylphosphatidylinositol (GPI)-anchored fragment, build a complex

together with other MSPs (Kauth et al. 2003; McBride and Heidrich 1987; Heidrich et al. 1983), and members of this complex are sequentially processed in a highly regulated cascade (Boyle et al. 2014). So far no host receptor for MSP attachment has been unambiguously identified; thus the molecular mechanisms of this initial contact still remain unclear.

Given the exposure of MSPs on the merozoite membrane where they provide the first contact to the host defence, it is unsurprising that they are highly polymorphic and that their genes bear signatures of balancing selection (Amambua-Ngwa et al. 2012). MSP-1 and MSP-2 are both GPI-anchored proteins that occur in two alleles with multiple variants each. Despite their variability, these proteins also contain highly conserved functional domains which were studied extensively as vaccine targets. However, clinical phase IIb trials with MSP-1 or MSP-2 showed no protection (Ogutu et al. 2009; Genton et al. 2002). Other small MSP families present in clusters in the *P. falciparum* genome include the MSP-3/MSP-6 (including MSPDBL1 and 2) and the MSP-7 [including merozoite surface-related proteins, MSRPs (Kadekoppala et al. 2010)] families. There is presently only limited evidence for clonal expression of members of these families (Rovira-Graells et al. 2012), but MSP-7 members seem to have redundant functions as deletion of each MSRP—except MSP-7 itself—shows no obvious phenotype (Kadekoppala et al. 2010). An MSP-3-like gene showed a strong signature of balancing selection in a population genomic screen, but displayed low-level expression in a clonally variant manner (Amambua-Ngwa et al. 2012).

5.2 *Proteins of the Rhoptries and Micronemes*

Rhoptry and microneme proteins are discharged early during invasion and appear to be responsible for several processes, including reorientation, formation of a tight junction, membrane invagination and formation of the PV. Several studies have illustrated the strict coordination of these events that seem to rely to some extent on the spatial organisation of proteins within the organelles (Counihan et al. 2013; Proellocks et al. 2010; Singh et al. 2010; Zuccala et al. 2012; Riglar et al. 2011). Known micronemal proteins implicated in these processes through interaction with host cell receptors and other parasite proteins include the single-copy and highly polymorphic protein apical membrane antigen 1 (AMA-1), as well as a number of paralogous proteins belonging to the erythrocyte-binding-like (EBL) protein family (EBA-175, EBA-181/JESEBL, EBA-140/BAEBL, EBL-1) (Cowman et al. 2012). EBL proteins contain a Duffy-binding-like (DBL) domain which has been shown to mediate host receptor binding (Tolia et al. 2005). For EBA-175, EBL-1 and EBA-140, glycophorins A, B and C, respectively, have been identified as host cell receptors (Mayer et al. 2009; Lobo et al. 2003; Maier et al. 2003; Sim et al. 1994). EBL family proteins have redundant functions in invasion and enable the parasite to switch between alternative invasion pathways. Studies on double knockout parasite lines provided evidence that the variation in EBL family member

expression contributes to phenotypic variation and immune evasion (Persson et al. 2013). Interestingly, EBA-175 but not EBA-140 showed an excess of non-synonymous polymorphisms indicating strong diversifying pressure on EBA-175 in a population in Nigeria (Baum et al. 2003). Conversely, in South America, stronger diversifying pressure led to greater sequence variability in EBA-140 than EBA-175 (Yalcindag et al. 2014). In Melanesia, a deletion in the EBA-140 receptor glycophorin C occurs with high frequency, demonstrating a counter-adaptation process in the human population (Maier et al. 2003).

Several rhoptry proteins, among them the *Plasmodium falciparum* reticulocyte-binding-like homologues (PfRhs), are also present as small multigene families in the genome (Gardner et al. 2002a). PfRh proteins have homologues in other *Plasmodium* species, including the reticulocyte-binding proteins (RBPs) in *P. vivax* and Py235 family in *P. yoelii* (Galinski et al. 1992; Preiser et al. 2002). Like micronemal EBL proteins, PfRhs seem to engage erythrocyte surface receptors, including complement receptor 1 for PfRh4 and CD147/basigin for PfRh5 (Crosnier et al. 2011; Rayner et al. 2000; Tham et al. 2010). PfRh proteins also show redundant functions during erythrocyte invasion based on different host cell receptor interactions (Duraisingh et al. 2003), with the exception of the essential and conserved PfRh5 (Baum et al. 2009). Immunisation with PfRh5 showed cross-protective reactivity against heterologous strains in a monkey model, suggesting it may be a promising vaccine target (Douglas et al. 2015).

Expressing such a large array of erythrotropic proteins is thought to equip the parasite with the ability to quickly switch the invasion pathway to circumvent polymorphisms in invasion receptors such as blood group antigens in different host individuals. Interestingly, PfRh and EBL family proteins appear to be functionally connected with each other in this process, as, for example, disruption of EBA-181 affects invasion through the Rh2a/b-dependent pathway (Lopaticki et al. 2011). Upregulation of Rh4 after deletion of EBA-175 further supports interdependency of the two classes of red blood cell receptor ligands in the invasion process (Stubbs et al. 2005).

5.3 Dense Granule Proteins

Once the merozoite has entered the host, dense granules move to the pellicle, release their contents into the PV and move into finger-like channels of the PVM (Torii et al. 1989; Trager et al. 1992). Two dense granule proteins have so far been identified to localise to the erythrocyte membrane shortly after invasion (Trager et al. 1992; Aikawa et al. 1990). There is evidence suggesting that one of them, the ring-infected erythrocyte surface protein (RESA), is responsible for the increased rigidity of the IE at febrile temperatures by interaction with spectrin at the membrane cytoskeleton and moreover suppresses further invasion (Foley et al. 1991; Mills et al. 2007; Pei et al. 2007).

Dense granule proteins thus appear to be crucial in the early modifications of the host cell. There are multiple RESA-like proteins encoded in the *P. falciparum* genome, but their function is presently unclear.

5.4 Subtelomeric Gene Families in Merozoites

Functionally less defined is the role of RIFIN and STEVOR variants that are expressed in different compartments within merozoites (Petter et al. 2007; Khattab et al. 2008; Khattab and Meri 2011; Blythe et al. 2008). Transcription of STEVORs and RIFINs undergoes two peaks during intraerythrocytic development which supports their role in various life cycle stages (Bachmann et al. 2012). At least some STEVOR variants are released from internal organelles and exposed at the merozoite membrane (Khattab and Meri 2011; Niang et al. 2014). STEVOR seems to have a dual function in red blood cell adhesion on the infected erythrocyte surface as well as on the merozoite membrane and has been reported to use glycophorin C as its host cell receptor (Niang et al. 2014). Antibodies targeting STEVOR are able to block invasion, providing evidence for a critical role in host cell attachment (Niang et al. 2014).

Another multigene family coding for proteins termed SURFINs has been implicated in antigenic variation of *P. falciparum* merozoites (Winter et al. 2005; Mphande et al. 2008). However, the functional relevance of these proteins in merozoites remains elusive.

6 Evolution of Diversity in the *var* Multigene Family

6.1 var *Gene Recombination*

The foundation of *P.falciparum*immune evasion is the antigenic diversity generated by polymorphism within the *var*multigene family. The polymorphisms within *var* genes correlate within a segment indicating that genes consist of mosaics of semi-conserved blocks that are exchanged by recombination (Bockhorst et al. 2007). Analysing *var* genes at the subdomain level identified conserved subdomain arrangements S1, S2 and S3 in DBL domains (Fig. 1a) and M1, M2 and M3 in CIDR domains and a large number of smaller, conserved homology blocks (HBs) (Smith et al. 2000b; Rask et al. 2010). These subdomains had different phylogenetic relationships to those of their cognate entire DBL domains, as did the arrangements of homology blocks, indicating that recombination occurred within the DBL domains (Rask et al. 2010). The semi-conserved homology blocks contain conserved amino acids required for maintaining the protein's structure (Kraemer et al. 2007; Rask et al. 2010), and the expression of the same, few homology blocks

within the DBLα was associated with severe malaria in Kenya and Mali (Rorick et al. 2013) indicating that they represent functionally constrained sequences possibly involved in virulent cytoadhesion phenotypes.

The majority of recombinations between *var* genes conserve the reading frame of the rearranged *var* gene; they initiate within short regions of homology and involve multiple crossovers between the same two *var* genes and are largely between domains of the same type (Claessens et al. 2014; Kraemer et al. 2007). The majority of recombinations in *P. falciparum* overall are in or near *var* genes. In two longitudinal studies of multiple clones, 98 recombinations were identified, 30 were in *var* gene-coding sequences, and the majority of the remnants were in other neighbouring members of variant, multigene families or in non-coding sequence close to *var* genes (Claessens et al. 2014; Bopp et al. 2013). *P. falciparum* clinical isolates of limited genetic diversity from a single region also had high levels of recombination primarily within and around *var* gene clusters (Dharia et al. 2010). Thus increased rates of recombination of variant genes appear to have evolved as a strategy to drive diversity in *P. falciparum* variant, multigene families.

Recombination between *var* genes appears to be partially restricted to occur largely within *var* gene groups (Kraemer et al. 2007; Claessens et al. 2014) (Fig. 1). This may reflect the evolution of *var* gene groups under different selective pressures. For example, domains within larger, multidomain *var* genes are more conserved than in shorter *var* genes, and group A *var* genes are in general larger than group B and C *var* genes (Buckee and Recker 2012) (Fig. 1a). Segregation of *var* gene groups has functional consequences, e.g. it preserves the placental adhesion phenotype of the unique variant *var2csa* and restricts adhesion to CD36 to group B and C PfEMP1s due to the CIDR1α domain that is not shared with group A (Robinson et al. 2003). Modelling suggests that fitness is optimal with a genomic repertoire including both more conserved, larger genes which have an advantage through optimal conserved adhesion phenotypes and shorter more diverse *var* genes that have an advantage in escaping immunity in semi-immune hosts (Buckee and Recker 2012). This is consistent with the way these functional groups appear to shape the course of natural infections as patterns of antibody acquisition in children indicate that PfEMP1s encoded by group A and B *var* genes elicit antibodies before PfEMP1s encoded by group C *var* genes (Cham et al. 2009). Interestingly the group A and B *var* genes that are expressed more frequently in severe disease are also transcriptionally silenced more rapidly than the group C *var* genes (Frank et al. 2007). Thus parasites expressing group C *var* genes eventually dominate a parasite population in the absence of selection pressures, which would be consistent with their preferential expression in chronic infections of immune hosts.

6.2 *Mechanisms of Recombination Between* var *Genes*

The location and orientation of the *var* genes is believed to facilitate recombination between different *var* genes within the same group. Group B *var* genes are located at the telomeric proximal ends of subtelomeric clusters of *var* genes and are transcribed towards the centromere (Fig. 1). Group A *var* genes are located at the centromere proximal end of subtelomeric *var* gene clusters and are transcribed towards the telomere. Group C *var* genes are located in chromosome-internal clusters (Gardner et al. 2002a). The *var* genes are maintained in close proximity through the tethering of telomeres at the nuclear periphery in several polychromosomal clusters in asexual blood stages (Fig. 1b) and in a single bouquet-like arrangement in sexual-stage gametocytes (Freitas-Junior et al. 2000). Physical proximity would maximise the opportunities for recombination between *var* genes in *trans*. The subtelomeric sites are generally recombinogenic due to the large number of telomere-associated repeat elements (TAREs) that are present and that could facilitate strand invasion following double strand break; however, the semi-conserved *var* sequences themselves are capable of initiating recombination (Freitas-Junior et al. 2000). The opposing orientation of the group A and B *var* genes has been proposed to inhibit recombination between group A and B *var* genes that are tightly constrained within a cluster of telomeres (Kraemer and Smith 2006) (Fig. 1b) that persists through S and M phase (Arnot et al. 2011).

Initial studies indicated that recombination between *var* genes occurred during meiosis (Freitas-Junior et al. 2000). We since showed that ectopic recombination between *var* genes also occurred during mitosis (Duffy et al. 2009) and large sequencing studies have subsequently shown that these mitotic rearrangements are not uncommon (Bopp et al. 2013; Claessens et al. 2014). Indeed the frequency of asexual stage *var* gene rearrangements (0.2 % per generation) was proposed to explain the majority of *var* gene recombinations (Claessens et al. 2014). Whether the mechanisms of recombination in asexual- and sexual-stage parasites differ is contentious (Claessens et al. 2014; Sander et al. 2014). However, the largest survey of mitotic and meiotic recombinations suggests they occur by the same mechanism (Claessens et al. 2014). These recombinations can be reciprocal or gene conversions and probably occur via homologous recombination following double strand break as the parasite possesses the requisite enzymes for this pathway but lacks the enzymes for non-homologous end joining (Gardner et al. 2002a; Gopalakrishnan and Kumar 2013).

In addition to their conserved DNA motifs, *var* genes have other specific recombinogenic characteristics. Imperfect palindromes capable of forming DNA secondary structures with low folding free energy are enriched at the boundaries of structural domains in PfEMP1 and in intergenic regions prone to recombination (Sander et al. 2014) (Fig. 1a). These sequences can induce recombination in yeast and are proposed to promote recombination in *var* genes through interfering with replication leading to template switching (Sander et al. 2014).

6.3 var *Gene Evolution*

Identifying the selection pressures exerted on *var* gene sequences is difficult because their tremendous diversity confounds definition of genuine alleles for comparison. However, useful data has been extracted from the comparison of *var2csa* gene alleles. This gene that encodes placental adhesion is atypically conserved between isolates and is present as a recognisable allele in 1 or 2 copies per genome. Overall non-synonymous/synonymous polymorphism rates (dN/dS) are high in *var2csa* genes indicating that balancing or diversifying selection is driven by immune pressure; however, there is a considerable variation in the dN/dS ratio across regions of *var2csa* indicating that some regions are less exposed to immune pressure and polymorphism is probably functionally constrained to maintain the adhesion phenotype (Trimnell et al. 2006; Bockhorst et al. 2007; Bordbar et al. 2014).

Recombination has driven diversity of *var* genes such that the limits of diversity in DBLα tags have not been detected globally nor in African sites (Barry et al. 2007; Chen et al. 2011). However, *var* diversity did plateau in samples from Brazil and Papua New Guinea which may also have had lower genotype diversity (Barry et al. 2007; Chen et al. 2011). Modelling the saturated *var* gene diversity from Papua New Guinea suggests that host immune pressure constrains *var* allelic recombination to maintain divergent antigenic strains, these being most divergent in areas of lower transmission with less opportunities for sexual recombination (Artzy-Randrup et al. 2012).

The location of the chromosome-internal positioned *var* gene clusters, the number of *var* genes and the most frequent PfEMP1 domain organisation and the most conserved DBLα sequence are all conserved between the subgenus *Laverania* parasites *P. falciparum* and *P. reichenowi* (Otto et al. 2014; Zilversmit et al. 2013). Balancing selection is probably responsible for the surprising conservation of DBLα homology blocks and larger sequence blocks between these species despite the high rate of recombination within *var* genes (Zilversmit et al. 2013). Frequency-dependent balancing selection may have prevented loss of ancient variants as their decreasing prevalence in the population would increase their fitness due to decreased immune recognition (Zilversmit et al. 2013).

7 Regulation of Variant Antigen Expression

7.1 *Clonal Antigenic Variation and* var *Genes*

The incredible diversity of the *var* gene family is carefully husbanded by the parasite which employs allelic exclusive expression of only one or a few *var* genes at any one time. The parasite avoids acquired immunity by switching between the expressed *var* genes in a process of clonal antigenic variation. This

process is regulated at the transcriptional level by epigenetic mechanisms (Scherf et al. 1998) that depend on DNA packaging into chromatin. The single active *var* gene is maintained in a transcriptionally active, relaxed, euchromatic state. In contrast, the clusters of transcriptionally silent *var* genes are maintained in non-contiguous islands of condensed, facultative heterochromatin. *P. falciparum* heterochromatin is restricted to these relatively small regions that are also enriched in other multigene families which have clonally variant expression and are involved in host-parasite interactions (Rovira-Graells et al. 2012). It is unusual that the parasite has reserved heterochromatin for silencing of redundant contingency genes rather than for maintaining silencing of genes that are expressed in other developmental stages.

The condensed, heterochromatic state of the nucleosomes in which the silent *var* genes are packaged is propagated by enzymatic modification of histones, some of which bind chromatin structural proteins (Lopez-Rubio et al. 2009; Duraisingh et al. 2005; Freitas-Junior et al. 2005; Flueck et al. 2009). Silencing of *var* genes requires trimethylation of histone H3 lysine 36 (H3K36me3) at *var* promoters by the histone methyltransferase PfSETvs (Jiang et al. 2013) and the activities of the histone deacetylases PfSir2A, PfSir2B and PfHDA2 (Duraisingh et al. 2005; Tonkin et al. 2009; Coleman et al. 2014). Unsurprisingly all of these enzymes also silence other genes. PfSir2A can deacetylate H3K9ac and H4K16ac (French et al. 2008) and both of these acetylations are present at the active *var* gene (Freitas-Junior et al. 2005). PfSir2A is also required for the binding of the chromatin structural protein PfOrc1 to *var* gene promoters and subtelomeric sequences (Deshmukh et al. 2012). It is probable that PfSir2A and PfOrc1 recruit each other to propagate the spread of heterochromatin. Similarly *P. falciparum* heterochromatin protein 1 (PfHP1) binds H3K9me3 and is required for propagation of trimethylation of H3K9 and thus heterochromatin formation at *var* loci and also for *var* gene silencing (Brancucci et al. 2014). However, the histone methyltransferase responsible for conferring the silencing mark H3K9me3 has not yet been identified.

The active *var* gene is transcribed only for the first 16 h of the 48 h life cycle; it is then transiently repressed until the following life cycle. The active *var* gene promoter is depleted of H3K36me3, H3K9me3 and PfHP1 but instead enriched in acetylated H3K9 (H3K9ac), H3K4me3 (Lopez-Rubio et al. 2007) and the interacting histone variants PfH2A.Z and Pf H2B.Z (Petter et al. 2011, 2013). In other eukaryotes, H2A.Z is exchanged post-replication into nucleosomes in regulatory sequences. The role of H2A.Z is contentious and differs between species, and it has been shown to antagonise the spread of subtelomeric heterochromatin and also variously to be associated with active gene expression, gene poising and repression of inducible genes [reviewed in Duffy et al. (2013)]. In *P. falciparum*, both PfH2A.Z and PfH2B.Z are removed from the promoters of both silent *var* genes and transiently repressed *var* genes (Petter et al. 2011, 2013). However, Pf H2A.Z persists at promoters of transiently repressed *var* genes in the absence of PfSir2A, hinting at a role for Pf H2A.Z in forming a barrier to PfSir2A-dependent heterochromatin (Petter et al. 2011). H3K4 is dimethylated as the gene is transiently repressed (Lopez-Rubio et al. 2007) and H3K4me2 is proposed to mark the *var*

gene as poised for activation in the next life cycle. The histone methyl transferase PfSET10 catalyses di- and trimethylation of H3K4 and co-localises with the transiently repressed *var* gene; thus it probably poises *var* genes (Volz et al. 2012).

Upstream sequence containing a *var* gene promoter and a motif that binds an unidentified protein are, together with a *var* intron, sufficient to integrate a reporter gene into the programme of *var* gene allelic exclusive expression (Voss et al. 2006; Dzikowski et al. 2006; Brancucci et al. 2012). The *var* intron represses the upstream *var* promoter in a manner that is dependent on promoter activity within the *var* intron itself (Calderwood et al. 2003; Deitsch et al. 2001) (Fig. 1b); interestingly, the *var* intron is also enriched in the histone variants Pf HA.Z and Pf H2B.z which define *cis* regulatory sequences (Petter et al. 2011, 2013). In the absence of a *var* intron or surrogate sequence with promoter activity, the *var* upstream promoter escapes silencing allowing simultaneous transcription of multiple *var* genes (Frank et al. 2006).

Some mechanisms by which sequence elements function in *var* regulatory sequences have been identified. The subtelomeric *var* promoter element 2 (SPE2) sequence element upstream of the group B var genes binds the transcription repressor (PfSIP2) (Voss et al. 2007) and the *var* upstream promoter and the *var* intron both possess insulator-like elements that bind the same nuclear proteins and interact in silenced *var* genes (Avraham et al. 2012) (Fig. 1b). These observations are suggestive of dynamic chromatin architecture regulating *var* gene expression. Indeed in two chromosome conformation capture next generation sequencing (Hi-C) studies, major inter- and intra-chromosomal interactions were between the regions containing the *var* virulence genes (Lemieux et al. 2013; Ay et al. 2014). Subtelomeric and chromosome-internal *var* gene clusters interacted with each other, and the clusters were present at the boundaries of chromatin compartments suggesting that the *var* gene clusters were major components defining nuclear chromatin architecture (Ay et al. 2014).

The inter-chromosomal interactions between clusters of silent *var* genes can be visualised as several foci each containing chromosome-internal *var* gene clusters or multiple telomeric regions that are tethered at the nuclear periphery (Lopez-Rubio et al. 2009) via telomeric sequences (Figueiredo et al. 2002) (Fig. 1). The *var* genes are recruited to the perinuclear polychromosomal clusters by F actin that indirectly binds the *var* introns (Zhang et al. 2011), and the polychromosomal clusters are held together by interactions between subtelomeric repetitive sequences and proteins (Marty et al. 2006; O'Donnell et al. 2002) (Fig. 1b). This nuclear organisation extends to non-*var* genes that are also resident in transcriptionally silent heterochromatin and are also tethered at the nuclear periphery (Lopez-Rubio et al. 2009). The tethered clusters of silent *var* genes appear to define nuclear regions of transcriptional repression as even genes in a distant genomic location or on *trans* chromosomes have decreased expression if they are in spatial proximity to the telomeric sequences (Ay et al. 2014). When activated, a *var* gene is found separate from the polychromosomal clusters containing silent *var* genes and instead is located in a separate perinuclear expression site (Lopez-Rubio et al. 2009; Duraisingh et al. 2005) (Fig. 1b). In general, genes with similar expression states

appear to co-localise within the nucleus (Ay et al. 2014) which is consistent with the occasional occupancy of the *var* gene expression site by more than one *var* gene (Joergensen et al. 2010; Brolin et al. 2009) and also transcribed rifin genes (Howitt et al. 2009).

How the dynamic nuclear architecture of *P. falciparum* regulates *var* genes is unknown but an ordered process has been elucidated for the erythrocyte invasion gene *PfRh4*, which is expressed in a phase-variable manner (Coleman et al. 2012). Silent *PfRh4* is also present in subtelomeric heterochromatin. *PfRh4* occasionally leaves perinuclear, inter-chromosomal clusters of silent heterochromatin and enters a separate expression site. However, it is only after entering the expression site that the heterochromatic *PfRh4* locus is remodelled into transcriptionally active euchromatin by H3K9 demethylation and acetylation. Thus the nuclear mobility of *PfRh4* determines its potential rate of activation but does not lead to activation per se (Coleman et al. 2012); it seems probable that a similar mechanism functions in *var* gene regulation.

An additional level of *var* gene regulation is exerted by non-coding RNAs (ncRNAs). The *var* intron's bidirectional promoter activity leads to transcription of sense and anti-sense ncRNAs (Epp et al. 2009). The *var* anti-sense ncRNAs associate with chromatin and may be retained at the telomeric clusters (Epp et al. 2009). Both the anti-sense ncRNAs and short RNAs transcribed from the upstream promoter of group A *var* genes are degraded by a unique nuclease, PfRNase II, that co-localises with the group A *var* genes (Zhang et al. 2014). Modification of PfRNase II leads to increased group A *var* gene expression and allows multiple group A *var* genes to escape silencing, implicating the activity of PfRNase II in the maintenance of allelic exclusive expression. Long ncRNAs are also expressed from subtelomeric repeat sequences that are adjacent to *var* genes and that contain the SPE2 sequences that also bind repressors upstream of *var* genes (Broadbent et al. 2011). However, no role for these sequences in *var* gene regulation has yet been identified.

Clearly a complex network of *cis* sequences, chromatin structural proteins, transcription factors and repressors, RNAs and three-dimensional nuclear architecture interact to achieve the allelic exclusive expression and *var* gene switching that result in effective clonal antigenic variation in *P. falciparum*. Individual *var* gene transcription units differ in their propensity for activation and this is presumed to relate to promoter sequence but also genomic location (Frank et al. 2007). The sequence outside of the *var* gene promoter also impacts on a *var* gene propensity to be activated (Duffy et al. 2009), presumably through its ability to maintain a chromatin state or participate in three-dimensional nuclear structures. Thus the process of recombination not only drives the diversity of the *var* gene repertoire but can also directly influence the process of antigenic variation through modifying the rates at which individual *var* genes are activated and silenced (Duffy et al. 2009).

7.2 Small VSA Gene Regulation

Like *var*gene regulation, the regulation of the small VSA underlies epigenetic mechanisms. *Stevor* and *rif* genes are present in the heterochromatic regions of the genome (Lopez-Rubio et al. 2009; Flueck et al. 2009; Salcedo-Amaya et al. 2009), and active *rif* genes carry the histone mark H3K9ac at the promoter (Cabral et al. 2012). However, Sir2A deletion resulted in a shift in the expression of *rif* genes away from the trophozoite stage towards the ring or schizont stage, rather than a group-specific upregulation as seen for *var* genes (Cabral et al. 2012). Only a subset of *rif* genes orientated head to head with a group A *var* gene were upregulated (Duraisingh et al. 2005). Surprisingly, Sir2B knockout parasites had lower levels of *rif* transcription (Tonkin et al. 2009). These studies thus identify important differences in the regulation of heterochromatic *var* and *rif* genes in close proximity. In fact, *var, stevor* and *rif* genes seem to generally be regulated independently from each other (Tham et al. 2007; Sharp et al. 2006), even in cases where they share the same 5'UTR (Tham et al. 2007). As an exception, one report found co-regulation of a *var* and its adjacent *rif* gene in parasites selected for surface expression of severe malaria antigen by panning on serum antibodies from children who experienced severe disease (Wang et al. 2008); however, it is unclear whether this is a result of the selection process or a result of mechanistic co-regulation. It was suggested that *var*, *rif* and *stevor* promoters share a common titratable activation factor, as episomal overexpression of the promoter from one family resulted in global downregulation of members of all three families (Howitt et al. 2009); however, this observation was not reproduced in a separate, similar study (Witmer et al. 2012). Although evidence so far is limited in support of their coactivation, certain *var* and *rif* genes might share a common mechanism to achieve silencing. Detailed mapping of transcriptionally important sequence elements in the promoter of a *rif* gene identified two silencing motifs, one of which shares some sequence similarity with the group B *var* SIP2-binding site (Tham et al. 2007).

7.3 Conclusion

As an intracellular pathogen, *P. falciparum* avoids direct exposure to the host's immune system for most of its erythrocytic life cycle. Nonetheless, a considerable portion of its genome is devoted to the diversification of redundant members of multigene families that are exposed to host immunity. In the two processes described above, erythrocyte invasion and sequestration of the infected erythrocyte, the parasite has also employed clonally variant expression of redundant members of the multigene families mediating these processes in order to evade immunity. By combining recombination/mutation with clonally variant expression, the parasite has evolved an elegant mechanism for evading host immunity during the protracted periods that it must maintain an infection whilst awaiting transmission. The

challenge to the parasite is to so disguise the functional elements of its cytoadhesive ligands and invasion proteins that they only poorly elicit immunity. The challenge for the research community is to identify motifs or structural epitopes that the parasite must conserve for function and which are sufficiently exposed to be targeted by immunity. Though unsuccessful so far, the rapid, natural acquisition of immunity to severe malaria disease gives hope that such epitopes may yet be identified and employed as vaccines protective against disease or blood-stage infection.

References

Abdel-Latif MS, Khattab A, Lindenthal C, Kremsner PG, Klinkert MQ (2002) Recognition of variant Rifin antigens by human antibodies induced during natural *Plasmodium falciparum* infections. Infect Immun 70(12):7013–7021

Abdel-Latif MS, Dietz K, Issifou S, Kremsner PG, Klinkert MQ (2003) Antibodies to *Plasmodium falciparum* rifin proteins are associated with rapid parasite clearance and asymptomatic infections. Infect Immun 71(11):6229–6233

Abdel-Latif MS, Cabrera G, Kohler C, Kremsner PG, Luty AJ (2004) Antibodies to rifin: a component of naturally acquired responses to *Plasmodium falciparum* variant surface antigens on infected erythrocytes. Am J Trop Med Hyg 71(2):179–186, 71/2/179 [pii]

Adams JH, Sim BKL, Dolan SA, Fang X, Kaslow DC, Miller LH (1992) A family of erythrocyte binding proteins of malaria parasites. Proc Natl Acad Sci USA 89:7085–7089

Adisa A, Albano FR, Reeder J, Foley M, Tilley L (2001) Evidence for a role for a *Plasmodium falciparum* homologue of Sec31p in the export of proteins to the surface of malaria parasite-infected erythrocytes. J Cell Sci 114(Pt 18):3377–3386

Aikawa M, Torii M, Sjolander A, Berzins K, Perlmann P, Miller LH (1990) Pf155/RESA antigen is localized in dense granules of *Plasmodium falciparum* merozoites. Exp Parasitol 71 (3):326–329

Albano FR, Berman A, La Greca N, Hibbs AR, Wickham M, Foley M, Tilley L (1999) A homologue of Sar1p localises to a novel trafficking pathway in malaria-infected erythrocytes. Eur J Cell Biol 78(7):453–462

Amambua-Ngwa A, Tetteh KK, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, Cheeseman IH, Newbold CI, Holder AA, Knuepfer E, Janha O, Jallow M, Campino S, Macinnis B, Kwiatkowski DP, Conway DJ (2012) Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. PLoS Genet 8(11), e1002992

Amino R, Thiberge S, Martin B, Celli S, Shorte S, Frischknecht F, Menard R (2006) Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. Nat Med 12(2):220–224

Arnot DE, Ronander E, Bengtsson DC (2011) The progression of the intra-erythrocytic cell cycle of *Plasmodium falciparum* and the role of the centriolar plaques in asynchronous mitotic division during schizogony. Int J Parasitol 41(1):71–80. doi:10.1016/j.ijpara.2010.07.012, S0020-7519(10)00286-9 [pii]

Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M (2012) Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov 11(5):384–400. doi:10.1038/nrd3674, nrd3674 [pii]

Artzy-Randrup Y, Rorick MM, Day K, Chen D, Dobson AP, Pascual M (2012) Population structuring of multi-copy, antigen-encoding genes in *Plasmodium falciparum*. Elife 1, e00093. doi:10.7554/eLife.00093

Atkinson CT, Aikawa M, Perry G, Fujino T, Bennett V, Davidson EA, Howard RJ (1988) Ultrastructural localization of erythrocyte cytoskeletal and integral membrane proteins in *Plasmodium falciparum*-infected erythrocytes. Eur J Cell Biol 45(2):192–199

Avraham I, Schreier J, Dzikowski R (2012) Insulator-like pairing elements regulate silencing and mutually exclusive expression in the malaria parasite *Plasmodium falciparum*. Proc Natl Acad Sci USA 109(52):E3678–E3686. doi:10.1073/pnas.1214572109, 1214572109 [pii]

Avril M, Tripathi AK, Brazier AJ, Andisi C, Janes JH, Soma VL, Sullivan DJ Jr, Bull PC, Stins MF, Smith JD (2012) A restricted subset of var genes mediates adherence of *Plasmodium falciparum*-infected erythrocytes to brain endothelial cells. Proc Natl Acad Sci USA 109(26): E1782–E1790. doi:10.1073/pnas.1120534109, 1120534109 [pii]

Ay F, Bunnik EM, Varoquaux N, Bol SM, Prudhomme J, Vert JP, Noble WS, Le Roch KG (2014) Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. Genome Res 24 (6):974–988. doi:10.1101/gr.169417.113

Bachmann A, Esser C, Petter M, Predehl S, von Kalckreuth V, Schmiedel S, Bruchhaus I, Tannich E (2009) Absence of erythrocyte sequestration and lack of multicopy gene family expression in *Plasmodium falciparum* from a splenectomized malaria patient. PLoS One 4(10), e7459. doi:10.1371/journal.pone.0007459

Bachmann A, Petter M, Tilly A-K, Biller L, Uliczka KA, Duffy MF, Tannich E, Bruchhaus I (2012) Temporal expression and localization patterns of variant surface antigens in clinical *Plasmodium falciparum* isolates during erythrocyte schizogony. PloS one 7(11), e49540. doi:10.1371/journal.pone.0049540

Bannister LH, Hopkins JM, Fowler RE, Krishna S, Mitchell GH (2000) A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages. Parasitol Today 16 (10):427–433

Bannister LH, Hopkins JM, Margos G, Dluzewski AR, Mitchell GH (2004) Three-dimensional ultrastructure of the ring stage of *Plasmodium falciparum*: evidence for export pathways. Microsc Microanal 10(5):551–562

Barfod L, Bernasconi NL, Dahlback M, Jarrossay D, Andersen PH, Salanti A, Ofori MF, Turner L, Resende M, Nielsen MA, Theander TG, Sallusto F, Lanzavecchia A, Hviid L (2007) Human pregnancy-associated malaria-specific B cells target polymorphic, conformational epitopes in VAR2CSA. Mol Microbiol 63(2):335–347

Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, Brown SM, McVean GA, Day KP (2007) Population genomics of the immune evasion (*var*) genes of *Plasmodium falciparum*. PLoS Pathog 3(3), e34

Baruch DI, Pasloske BL, Singh HB, Bi X, Ma XC, Feldman M, Taraschi TF, Howard RJ (1995) Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell 82:77–87

Baum J, Thomas AW, Conway DJ (2003) Evidence for diversifying selection on erythrocyte-binding antigens of *Plasmodium falciparum* and P. vivax. Genetics 163(4):1327–1336

Baum J, Chen L, Healer J, Lopaticki S, Boyle M, Triglia T, Ehlgen F, Ralph SA, Beeson JG, Cowman AF (2009) Reticulocyte-binding protein homologue 5 – an essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum*. Int J Parasitol 39(3):371–380

Beck JR, Muralidharan V, Oksman A, Goldberg DE (2014) PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. Nature 511(7511):592–595

Berger SS, Turner L, Wang CW, Petersen JE, Kraft M, Lusingu JP, Mmbando B, Marquard AM, Bengtsson DB, Hviid L, Nielsen MA, Theander TG, Lavstsen T (2013) *Plasmodium falciparum* expressing domain cassette 5 type PfEMP1 (DC5-PfEMP1) bind PECAM1. PLoS One 8(7), e69117. doi:10.1371/journal.pone.0069117, PONE-D-13-01460 [pii]

Biggs BA, Kemp DJ, Brown GV (1989) Subtelomeric chromosome deletions in field isolates of *Plasmodium falciparum* and their relationship to loss of cytoadherence in vitro. Proc Natl Acad Sci USA 86(7):2428–2432

Biggs BA, Goozé L, Wycherley K, Wollish W, Southwell B, Leech JH, Brown GV (1991) Antigenic variation in *Plasmodium falciparum*. Proc Natl Acad Sci USA 88(20):9171–9174

Blisnick T, Morales Betoulle ME, Barale JC, Uzureau P, Berry L, Desroses S, Fujioka H, Mattei D, Braun Breton C (2000) Pfsbp1, a Maurer's cleft *Plasmodium falciparum* protein, is associated with the erythrocyte skeleton. Mol Biochem Parasitol 111(1):107–121

Blisnick T, Vincensini L, Barale JC, Namane A, Braun Breton C (2005) LANCL1, an erythrocyte protein recruited to the Maurer's clefts during *Plasmodium falciparum* development. Mol Biochem Parasitol 141(1):39–47

Blythe JE, Yam XY, Kuss C, Bozdech Z, Holder AA, Marsh K, Langhorne J, Preiser PR (2008) *Plasmodium falciparum* STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. Infect Immun 76(7):3329–3336

Blythe JE, Niang M, Marsh K, Holder AA, Langhorne J, Preiser PR (2009) Characterization of the repertoire diversity of the *Plasmodium falciparum* stevor multigene family in laboratory and field isolates. Malar J 8:140

Bockhorst J, Lu F, Janes JH, Keebler J, Gamain B, Awadalla P, Su XZ, Samudrala R, Jojic N, Smith JD (2007) Structural polymorphism and diversifying selection on the pregnancy malaria vaccine candidate VAR2CSA. Mol Biochem Parasitol 155(2):103–112. doi:10.1016/j.molbiopara.2007.06.007

Boddey JA, Hodder AN, Gunther S, Gilson PR, Patsiouras H, Kapp EA, Pearce JA, de Koning-Ward TF, Simpson RJ, Crabb BS, Cowman AF (2010) An aspartyl protease directs malaria effector proteins to the host cell. Nature 463(7281):627–631. doi:10.1038/nature08728, nature08728 [pii]

Boddey JA, Carvalho TG, Hodder AN, Sargeant TJ, Sleebs BE, Marapana D, Lopaticki S, Nebl T, Cowman AF (2013) Role of plasmepsin V in export of diverse protein families from the *Plasmodium falciparum* exportome. Traffic 14(5):532–550

Bopp SE, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, McCormack S, Plouffe D, McNamara CW, Walker JR, Fidock DA, Denchi EL, Winzeler EA (2013) Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. PLoS Genet 9(2), e1003293. doi:10.1371/journal.pgen.1003293

Bordbar B, Tuikue Ndam N, Renard E, Jafari-Guemouri S, Tavul L, Jennison C, Gnidehou S, Tahar R, Gamboa D, Bendezu J, Menard D, Barry AE, Deloron P, Sabbagh A (2014) Genetic diversity of VAR2CSA ID1-DBL2Xb in worldwide *Plasmodium falciparum* populations: impact on vaccine design for placental malaria. Infect Genet Evol 25:81–92. doi:10.1016/j.meegid.2014.04.010

Boyle MJ, Langer C, Chan JA, Hodder AN, Coppel RL, Anders RF, Beeson JG (2014) Sequential processing of merozoite surface proteins during and after erythrocyte invasion by *Plasmodium falciparum*. Infect Immun 82(3):924–936

Brancucci NM, Witmer K, Schmid CD, Flueck C, Voss TS (2012) Identification of a cis-acting DNA-protein interaction implicated in singular var gene choice in *Plasmodium falciparum*. Cell Microbiol 14(12):1836–1848. doi:10.1111/cmi.12004

Brancucci NM, Bertschi NL, Zhu L, Niederwieser I, Chin WH, Wampfler R, Freymond C, Rottmann M, Felger I, Bozdech Z, Voss TS (2014) Heterochromatin protein 1 secures survival and transmission of malaria parasites. Cell Host Microbe 16(2):165–176. doi:10.1016/j.chom.2014.07.004

Bray RS, Sinden RE (1979) The sequestration of *Plasmodium falciparum* infected erythrocytes in the placenta. Trans R Soc Trop Med Hyg 73(6):716–719

Broadbent KM, Park D, Wolf AR, Van Tyne D, Sims JS, Ribacke U, Volkman S, Duraisingh M, Wirth D, Sabeti PC, Rinn JL (2011) A global transcriptional analysis of *Plasmodium falciparum* malaria reveals a novel family of telomere-associated lncRNAs. Genome Biol 12(6):R56. doi:10.1186/gb-2011-12-6-r56, gb-2011-12-6-r56 [pii]

Brolin KJ, Ribacke U, Nilsson S, Ankarklev J, Moll K, Wahlgren M, Chen Q (2009) Simultaneous transcription of duplicated var2csa gene copies in individual *Plasmodium falciparum* parasites. Genome Biol 10(10):R117. doi:10.1186/gb-2009-10-10-r117, gb-2009-10-10-r117 [pii]

Bruce-Chwatt LJ (1963) A longitudinal survey of natural malaria infection in a group of west African adults. West Afr Med J 12:199–217

Buckee CO, Recker M (2012) Evolution of the multi-domain structures of virulence genes in the human malaria parasite, *Plasmodium falciparum*. PLoS Comput Biol 8(4), e1002451. doi:10.1371/journal.pcbi.1002451

Bull PC, Lowe BS, Kortok M, Molyneux CS, Newbold CI, Marsh K (1998) Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med 4 (3):358–360

Bull PC, Kortok M, Kai O, Ndungu F, Ross A, Lowe BS, Newbold CI, Marsh K (2000) *Plasmodium falciparum*-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. J Infect Dis 182(1):252–259

Bultrini E, Brick K, Mukherjee S, Zhang Y, Silvestrini F, Alano P, Pizzi E (2009) Revisiting the *Plasmodium falciparum* RIFIN family: from comparative genomics to 3D-model prediction. BMC Genomics 10:445

Cabral FJ, Fotoran WL, Wunderlich G (2012) Dynamic activation and repression of the *Plasmodium falciparum* rif gene family and their relation to chromatin modification. PLoS One 7(1), e29881

Calderwood MS, Gannoun-Zaki L, Wellems TE, Deitsch KW (2003) *Plasmodium falciparum var* genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. J Biol Chem 278(36):34125–34132

Carlson J, Helmby H, Hill AV, Brewster D, Greenwood BM, Wahlgren M (1990) Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. Lancet 336(8729):1457–1460

Cham GK, Turner L, Lusingu J, Vestergaard L, Mmbando BP, Kurtis JD, Jensen AT, Salanti A, Lavstsen T, Theander TG (2009) Sequential, ordered acquisition of antibodies to *Plasmodium falciparum* erythrocyte membrane protein 1 domains. J Immunol 183(5):3356–3363. doi:10.4049/jimmunol.0901331, jimmunol.0901331 [pii]

Chan J-A, Howell KB, Reiling L, Ataide R, Mackintosh CL, Fowkes FJI, Petter M, Chesson JM, Langer C, Warimwe GM, Duffy MF, Rogerson SJ, Bull PC, Cowman AF, Marsh K, Beeson JG (2012a) Targets of antibodies against Plasmodium falciparum-infected erythrocytes in malaria immunity. J Clin Investig 122(9):3227–3238. doi:10.1172/jci62182

Chan JA, Howell KB, Reiling L, Ataide R, Mackintosh CL, Fowkes FJ, Petter M, Chesson JM, Langer C, Warimwe GM, Duffy MF, Rogerson SJ, Bull PC, Cowman AF, Marsh K, Beeson JG (2012b) Targets of antibodies against *Plasmodium falciparum*-infected erythrocytes in malaria immunity. J Clin Invest 122(9):3227–3238. doi:10.1172/JCI62182, 62182 [pii]

Chen Q (2007) The naturally acquired immunity in severe malaria and its implication for a PfEMP-1 based vaccine. Microbes Infect 9(6):777–783

Chen DS, Barry AE, Leliwa-Sytek A, Smith TA, Peterson I, Brown SM, Migot-Nabias F, Deloron P, Kortok MM, Marsh K, Daily JP, Ndiaye D, Sarr O, Mboup S, Day KP (2011) A molecular epidemiological study of var gene diversity to characterize the reservoir of *Plasmodium falciparum* in humans in Africa. PLoS One 6(2), e16629. doi:10.1371/journal.pone.0016629

Cheng Q, Cloonan N, Fischer K, Thompson J, Waine G, Lanzer M, Saul A (1998) stevor and rif are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. Mol Biochem Parasitol 97(1–2):161–176, S0166-6851(98)00144-3 [pii]

Chitnis CE, Miller LH (1994) Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. J Exp Med 180:497–506

Claessens A, Ghumra A, Gupta AP, Mok S, Bozdech Z, Rowe JA (2010) Design of a variant surface antigen-supplemented microarray chip for whole transcriptome analysis of multiple

Plasmodium falciparum cytoadherent strains, and identification of strain-transcendent rif and stevor genes. Malar J 10:180

Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, Bull PC, Mok S, Gupta AP, Wang CW, Turner L, Arman M, Raza A, Bozdech Z, Rowe JA (2012) A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brain endothelial cells. Proc Natl Acad Sci USA 109(26):E1772–E1781. doi:10.1073/pnas.1120461109, 1120461109 [pii]

Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, Kwiatkowski D (2014) Generation of antigenic diversity in *Plasmodium falciparum* by structured rearrangement of Var genes during mitosis. PLoS Genet 10(12), e1004812. doi:10.1371/journal.pgen.1004812

Coleman BI, Ribacke U, Manary M, Bei AK, Winzeler EA, Wirth DF, Duraisingh MT (2012) Nuclear repositioning precedes promoter accessibility and is linked to the switching frequency of a *Plasmodium falciparum* invasion gene. Cell Host Microbe 12(6):739–750. doi:10.1016/j.chom.2012.11.004, S1931-3128(12)00394-0 [pii]

Coleman BI, Skillman KM, Jiang RH, Childs LM, Altenhofen LM, Ganter M, Leung Y, Goldowitz I, Kafsack BF, Marti M, Llinas M, Buckee CO, Duraisingh MT (2014) A *Plasmodium falciparum* histone deacetylase regulates antigenic variation and gametocyte conversion. Cell Host Microbe 16(2):177–186. doi:10.1016/j.chom.2014.06.014

Combes V, Rosenkranz AR, Redard M, Pizzolato G, Lepidi H, Vestweber D, Mayadas TN, Grau GE (2004a) Pathogenic role of P-selectin in experimental cerebral malaria: importance of the endothelial compartment. Am J Pathol 164(3):781–786

Combes V, Taylor TE, Juhan-Vague I, Mege JL, Mwenechanya J, Tembo M, Grau GE, Molyneux ME (2004b) Circulating endothelial microparticles in Malawian children with severe *falciparum* malaria complicated with coma. JAMA 291(21):2542–2544

Cooke BM, Berendt AR, Craig AG, MacGregor J, Newbold CI, Nash GB (1994) Rolling and stationary cytoadhesion of red blood cells parasitized by *Plasmodium falciparum*: separate roles for ICAM-1, CD36 and thrombospondin. Br J Haematol 87:162–170

Cooke BM, Buckingham DW, Glenister FK, Fernandez KM, Bannister LH, Marti M, Mohandas N, Coppel RL (2006) A Maurer's cleft-associated protein is essential for expression of the major malaria virulence antigen on the surface of infected red blood cells. J Cell Biol 172 (6):899–908

Counihan NA, Kalanon M, Coppel RL, de Koning-Ward TF (2013) *Plasmodium* rhoptry proteins: why order is important. Trends Parasitol 29(5):228–236

Cowman AF, Berry D, Baum J (2012) The cellular and molecular basis for malaria parasite invasion of the human red blood cell. J Cell Biol 198(6):961–971. doi:10.1083/jcb.201206112, jcb.201206112 [pii]

Crabb BS, Cooke BM, Reeder JC, Waller RF, Caruana SR, Davern KM, Wickham ME, Brown GV, Coppel RL, Cowman AF (1997) Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell 89(2):287–296

Crosnier C, Bustamante LY, Bartholdson SJ, Bei AK, Theron M, Uchikawa M, Mboup S, Ndir O, Kwiatkowski DP, Duraisingh MT, Rayner JC, Wright GJ (2011) Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. Nature 480(7378):534–537

de Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, Sanders PR, Lundie RJ, Maier AG, Cowman AF, Crabb BS (2009) A newly discovered protein export machine in malaria parasites. Nature 459(7249):945–949. doi:10.1038/nature08104, nature08104 [pii]

Deitsch KW, Calderwood MS, Wellems TE (2001) Cooperative silencing elements in *var* genes. Nature 412:875–876

Deshmukh AS, Srivastava S, Herrmann S, Gupta A, Mitra P, Gilberger TW, Dhar SK (2012) The role of N-terminus of *Plasmodium falciparum* ORC1 in telomeric localization and var gene silencing. Nucleic Acids Res 40(12):5313–5331. doi:10.1093/nar/gks202, gks202 [pii]

Dharia NV, Plouffe D, Bopp SE, Gonzalez-Paez GE, Lucas C, Salas C, Soberon V, Bursulaya B, Kochel TJ, Bacon DJ, Winzeler EA (2010) Genome scanning of Amazonian *Plasmodium*

falciparum shows subtelomeric instability and clindamycin-resistant parasites. Genome Res 20 (11):1534–1544. doi:10.1101/gr.105163.110

Dorovini-Zis K, Schmidt K, Huynh H, Fu W, Whitten RO, Milner D, Kamiza S, Molyneux M, Taylor TE (2011) The neuropathology of fatal cerebral malaria in Malawian children. Am J Pathol 178(5):2146–2158. doi:10.1016/j.ajpath.2011.01.016

Douglas AD, Baldeviano GC, Lucas CM, Lugo-Roman LA, Crosnier C, Bartholdson SJ, Diouf A, Miura K, Lambert LE, Ventocilla JA, Leiva KP, Milne KH, Illingworth JJ, Spencer AJ, Hjerrild KA, Alanine DG, Turner AV, Moorhead JT, Edgel KA, Wu Y, Long CA, Wright GJ, Lescano AG, Draper SJ (2015) A PfRH5-based vaccine is efficacious against heterologous strain blood-stage *Plasmodium falciparum* infection in *Aotus* monkeys. Cell Host Microbe 17 (1):130–139

Duffy MF, Byrne TJ, Carret C, Ivens A, Brown GV (2009) Ectopic recombination of a malaria var gene during mitosis associated with an altered var switch rate. J Mol Biol 389(3):453–469. doi:10.1016/j.jmb.2009.04.032, S0022-2836(09)00476-8 [pii]

Duffy MF, Selvarajah SA, Josling GA, Petter M (2013) Epigenetic regulation of the *Plasmodium falciparum* genome. Brief Funct Genomics 13(3):203–216. doi:10.1093/bfgp/elt047, elt047 [pii]

Duraisingh MT, Triglia T, Ralph SA, Rayner JC, Barnwell JW, McFadden GI, Cowman AF (2003) Phenotypic variation of *Plasmodium falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes. EMBO J 22(5):1047–1057

Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, Freitas-Junior LH, Scherf A, Crabb BS, Cowman AF (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. Cell 121(1):13–24

Dzikowski R, Frank M, Deitsch K (2006) Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. PLoS Pathog 2(3), e22

Elmendorf HG, Haldar K (1993) Secretory transport in *Plasmodium*. Parasitol Today 9(3):98–102

Elmendorf HG, Haldar K (1994) Plasmodium falciparum exports the Golgi marker sphingomyelin synthase into a tubovesicular network in the cytoplasm of mature erythrocytes. J Cell Biol 124 (4):449–462

Elsworth B, Matthews K, Nie CQ, Kalanon M, Charnaud SC, Sanders PR, Chisholm SA, Counihan NA, Shaw PJ, Pino P, Chan JA, Azevedo MF, Rogerson SJ, Beeson JG, Crabb BS, Gilson PR, de Koning-Ward TF (2014) PTEX is an essential nexus for protein export in malaria parasites. Nature 511(7511):587–591. doi:10.1038/nature13555

Epp C, Li F, Howitt CA, Chookajorn T, Deitsch KW (2009) Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite *Plasmodium falciparum*. RNA 15(1):116–127. doi:10.1261/rna.1080109, rna.1080109 [pii]

Falk N, Kaestli M, Qi W, Ott M, Baea K, Cortes A, Beck HP (2009) Analysis of *Plasmodium falciparum var* genes expressed in children from Papua New Guinea. J Infect Dis 200 (3):347–356. doi:10.1086/600071

Fernandez V, Hommel M, Chen Q, Hagblom P, Wahlgren M (1999) Small, clonally variant antigens expressed on the surface of the *Plasmodium falciparum*-infected erythrocyte are encoded by the rif gene family and are the target of human immune responses. J Exp Med 190(10):1393–1404

Ferreira MU, da Silva Nunes M, Wunderlich G (2004) Antigenic diversity and immune evasion by malaria parasites. Clin Diagn Lab Immunol 11(6):987–995

Figueiredo LM, Freitas-Junior LH, Bottius E, Olivo-Marin JC, Scherf A (2002) A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation. EMBO J 21(4):815–824

Fischer K, Marti T, Rick B, Johnson D, Benting J, Baumeister S, Helmbrecht C, Lanzer M, Lingelbach K (1998) Characterization and cloning of the gene encoding the vacuolar membrane protein EXP-2 from Plasmodium falciparum. Mol Biochem Parasitol 92(1):47–57

Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, Ehlgen F, Ralph SA, Cowman AF, Bozdech Z, Stunnenberg HG, Voss TS (2009) *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog 5(9), e1000569. doi:10.1371/journal.ppat.1000569

Foley M, Tilley L, Sawyer WH, Anders RF (1991) The ring-infected erythrocyte surface antigen of *Plasmodium falciparum* associates with spectrin in the erythrocyte membrane. Mol Biochem Parasitol 46(1):137–147

Frank M, Dzikowski R, Costantini D, Amulic B, Berdougo E, Deitsch K (2006) Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite *Plasmodium falciparum*. J Biol Chem 281(15):9942–9952

Frank M, Dzikowski R, Amulic B, Deitsch K (2007) Variable switching rates of malaria virulence genes are associated with chromosomal position. Mol Microbiol 64(6):1486–1498

Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U, Wellems TE, Scherf A (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *Plasmodium falciparum*. Nature 407(6807):1018–1022

Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP, Mancio-Silva L, Leal-Silvestre RJ, Gontijo AM, Shorte S, Scherf A (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell 121(1):25–36

Fremount HN, Rossan RN (1974) The sites of sequestration of the Uganda-Palo Alto strain of *Plasmodium falciparum*-infected red blood cells in the squirrel monkey, Saimiri sciureus. J Parasitol 60(3):534–536

French JB, Cen Y, Sauve AA (2008) *Plasmodium falciparum* Sir2 is an NAD + -dependent deacetylase and an acetyllysine-dependent and acetyllysine-independent NAD+ glycohydrolase. Biochemistry 47(38):10227–10239. doi:10.1021/bi800767t

Frischknecht F, Baldacci P, Martin B, Zimmer C, Thiberge S, Olivo-Marin JC, Shorte SL, Menard R (2004) Imaging movement of malaria parasites during transmission by *Anopheles* mosquitoes. Cell Microbiol 6(7):687–694

Galinski MR, Medina CC, Ingravallo P, Barnwell JW (1992) A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. Cell 69(7):1213–1226

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002a) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419(6906):498–511. doi:10.1038/nature01097, nature01097 [pii]

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Elsen JA, Rutherford K, Saizberg SL, Craig A, Kyes S, Chan M-S, Nene V, Shallom SJ, Suh B, Peterson J, Angluoll S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Valdya AB, Martin DMA, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LD, Subramanlan GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002b) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419:498–511

Genton B, Betuela I, Felger I, Al-Yaman F, Anders RF, Saul A, Rare L, Baisor M, Lorry K, Brown GV, Pye D, Irving DO, Smith TA, Beck HP, Alpers MP (2002) A recombinant blood-stage malaria vaccine reduces *Plasmodium falciparum* density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea. J Infect Dis 185(6):820–827

Gilson PR, Crabb BS (2009) Morphology and kinetics of the three distinct phases of red blood cell invasion by *Plasmodium falciparum* merozoites. Int J Parasitol 39(1):91–96. doi:10.1016/j.ijpara.2008.09.007, S0020-7519(08)00378-0 [pii]

Gopalakrishnan AM, Kumar N (2013) Opposing roles for two molecular forms of replication protein A in Rad51-Rad54-mediated DNA recombination in *Plasmodium falciparum*. MBio 4 (3):e00252-13. doi:10.1128/mBio.00252-13

Gruenberg J, Allred DR, Sherman IW (1983) Scanning electron microscope-analysis of the protrusions (knobs) present on the surface of *Plasmodium falciparum*-infected erythrocytes. J Cell Biol 97(3):795–802

Gruring C, Heiber A, Kruse F, Ungefehr J, Gilberger TW, Spielmann T (2011) Development and host cell modifications of *Plasmodium falciparum* blood stages in four dimensions. Nat Commun 2:165

Gunther K, Tummler M, Arnold HH, Ridley R, Goman M, Scaife JG, Lingelbach K (1991) An exported protein of *Plasmodium falciparum* is synthesized as an integral membrane protein. Mol Biochem Parasitol 46(1):149–157

Haeggstrom M, Kironde F, Berzins K, Chen Q, Wahlgren M, Fernandez V (2004) Common trafficking pathway for variant antigens destined for the surface of the *Plasmodium falciparum*-infected erythrocyte. Mol Biochem Parasitol 133(1):1–14, S0166685103002391 [pii]

Haeggstrom M, von Euler A, Kironde F, Fernandez V, Wahlgren M (2007) Characterization of Maurer's clefts in *Plasmodium falciparum*-infected erythrocytes. Am J Trop Med Hyg 76 (1):27–32, 76/1/27 [pii]

Hanssen E, Carlton P, Deed S, Klonis N, Sedat J, DeRisi J, Tilley L (2010) Whole cell imaging reveals novel modular features of the exomembrane system of the malaria parasite, *Plasmodium falciparum*. Int J Parasitol 40(1):123–134. doi:10.1016/j.ijpara.2009.09.004, S0020-7519 (09)00371-3 [pii]

Heddini A, Chen Q, Obiero J, Kai O, Fernandez V, Marsh K, Muller WA, Wahlgren M (2001) Binding of *Plasmodium falciparum*-infected erythrocytes to soluble platelet endothelial cell adhesion molecule-1 (PECAM-1/CD31): frequent recognition by clinical isolates. Am J Trop Med Hyg 65(1):47–51

Heidrich HG, Strych W, Mrema JE (1983) Identification of surface and internal antigens from spontaneously released *Plasmodium falciparum* merozoites by radio-iodination and metabolic labelling. Z Parasitenkd 69(6):715–725

Helmby H, Cavelier L, Pettersson U, Wahlgren M (1993) Rosetting *Plasmodium falciparum*-infected erythrocytes express unique strain-specific antigens on their surface. Infect Immun 61 (1):284–288

Higgins MK (2008) The structure of a chondroitin sulfate-binding domain important in placental malaria. J Biol Chem 283(32):21842–21846. doi:10.1074/jbc.C800086200

Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, Lopez-Estrano C, Haldar K (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science 306(5703):1934–1937

Hinterberg K, Scherf A, Gysin J, Toyoshima T, Aikawa M, Mazie JC, da Silva LP, Mattei D (1994) *Plasmodium falciparum*: the Pf332 antigen is secreted from the parasite by a brefeldin A-dependent pathway and is translocated to the erythrocyte membrane via the Maurer's clefts. Exp Parasitol 79(3):279–291

Holder AA, Guevara Patino JA, Uthaipibull C, Syed SE, Ling IT, Scott-Finnigan T, Blackman MJ (1999) Merozoite surface protein 1, immune evasion, and vaccines against asexual blood stage malaria. Parassitologia 41(1–3):409–414

Howell DP, Levin EA, Springer AL, Kraemer SM, Phippard DJ, Schief WR, Smith JD (2008) Mapping a common interaction site used by *Plasmodium falciparum* Duffy binding-like domains to bind diverse host receptors. Mol Microbiol 67(1):78–87

Howitt CA, Wilinski D, Llinas M, Templeton TJ, Dzikowski R, Deitsch KW (2009) Clonally variant gene families in *Plasmodium falciparum* share a common activation factor. Mol Microbiol 73(6):1171–1185. doi:10.1111/j.1365-2958.2009.06846.x, MMI6846 [pii]

Hunt NH, Grau GE (2003) Cytokines: accelerators and brakes in the pathogenesis of cerebral malaria. Trends Immunol 24(9):491–499

Hviid L (2007) Development of vaccines against *Plasmodium falciparum* malaria: taking lessons from naturally acquired protective immunity. Microbes Infect 9(6):772–776

Janssen CS, Phillips RS, Turner CM, Barrett MP (2004) *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasites. Nucleic Acids Res 32(19):5712–5720. doi:10.1093/nar/gkh907, 32/19/5712 [pii]

Jensen AT, Magistrado P, Sharp S, Joergensen L, Lavstsen T, Chiucchiuini A, Salanti A, Vestergaard LS, Lusingu JP, Hermsen R, Sauerwein R, Christensen J, Nielsen MA, Hviid L, Sutherland C, Staalsoe T, Theander TG (2004) *Plasmodium falciparum* associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A *var* genes. J Exp Med 199(9):1179–1190

Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K, Yang W, Turner L, Lavstsen T, Theander TG, Peng W, Wei G, Jing Q, Wakabayashi Y, Bansal A, Luo Y, Ribeiro JM, Scherf A, Aravind L, Zhu J, Zhao K, Miller LH (2013) PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. Nature 499(7457):223–227. doi:10.1038/nature12361, nature12361 [pii]

Joannin N, Abhiman S, Sonnhammer EL, Wahlgren M (2008) Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. BMC Genomics 9:19. doi:10.1186/1471-2164-9-19, 1471-2164-9-19 [pii]

Joergensen L, Bengtsson DC, Bengtsson A, Ronander E, Berger SS, Turner L, Dalgaard MB, Cham GK, Victor ME, Lavstsen T, Theander TG, Arnot DE, Jensen AT (2010) Surface co-expression of two different PfEMP1 antigens on single *Plasmodium falciparum*-infected erythrocytes facilitates binding to ICAM1 and PECAM1. PLoS Pathog 6(9), e1001083. doi:10.1371/journal.ppat.1001083

Kadekoppala M, Ogun SA, Howell S, Gunaratne RS, Holder AA (2010) Systematic genetic analysis of the *Plasmodium falciparum* MSP7-like family reveals differences in protein expression, location, and importance in asexual growth of the blood-stage parasite. Eukaryot cell 9(7):1064–1074

Kaestli M, Cockburn IA, Cortes A, Baea K, Rowe JA, Beck HP (2006) Virulence of malaria is associated with differential expression of *Plasmodium falciparum var* gene subgroups in a case-control study. J Infect Dis 193(11):1567–1574

Kalmbach Y, Rottmann M, Kombila M, Kremsner PG, Beck HP, Kun JF (2010) Differential var gene expression in children with malaria and antidromic effects on host gene expression. J Infect Dis 202(2):313–317. doi:10.1086/653586

Kauth CW, Epp C, Bujard H, Lutz R (2003) The merozoite surface protein 1 complex of human malaria parasite *Plasmodium falciparum*: interactions and arrangements of subunits. J Biol Chem 278(25):22257–22264

Kaviratne M, Khan SM, Jarra W, Preiser PR (2002) Small variant STEVOR antigen is uniquely located within Maurer's clefts in *Plasmodium falciparum*-infected red blood cells. Eukaryot cell 1(6):926–935

Khan SM, Jarra W, Preiser PR (2001) The 235 kDa rhoptry protein of *Plasmodium (yoelii) yoelii*: function at the junction. Mol Biochem Parasitol 117(1):1–10

Khattab A, Meri S (2011) Exposure of the *Plasmodium falciparum* clonally variant STEVOR proteins on the merozoite surface. Malar J 10:58

Khattab A, Bonow I, Schreiber N, Petter M, Schmetz C, Klinkert MQ (2008) *Plasmodium falciparum* variant STEVOR antigens are expressed in merozoites and possibly associated with erythrocyte invasion. Malar J 7:137. doi:10.1186/1475-2875-7-137, 1475-2875-7-137 [pii]

Klein MM, Gittis AG, Su HP, Makobongo MO, Moore JM, Singh S, Miller LH, Garboczi DN (2008) The cysteine-rich interdomain region from the highly variable *Plasmodium falciparum* erythrocyte membrane protein-1 exhibits a conserved structure. PLoS Pathog 4(9), e1000147. doi:10.1371/journal.ppat.1000147

Kraemer SM, Smith JD (2006) A family affair: var genes, PfEMP1 binding, and malaria disease. Curr Opin Microbiol 9(4):374–380

Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, Christodoulou Z, Smith LM, Wang W, Levin E, Newbold CI, Myler PJ, Smith JD (2007) Patterns of gene recombination shape *var* gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. BMC Genomics 8:45

Kulzer S, Rug M, Brinkmann K, Cannon P, Cowman A, Lingelbach K, Blatch GL, Maier AG, Przyborski JM (2010) Parasite-encoded Hsp40 proteins define novel mobile structures in the cytosol of the *P. falciparum*-infected erythrocyte. Cell Microbiol 12(10):1398–1420

Kulzer S, Charnaud S, Dagan T, Riedel J, Mandal P, Pesce ER, Blatch GL, Crabb BS, Gilson PR, Przyborski JM (2012) *Plasmodium falciparum*-encoded exported hsp70/hsp40 chaperone/co-chaperone complexes within the host erythrocyte. Cell Microbiol 14(11):1784–1795. doi:10.1111/j.1462-5822.2012.01840.x

Kyes SA, Rowe JA, Kriek N, Newbold CI (1999) Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. Proc Natl Acad Sci USA 96(16):9333–9338

Kyes S, Pinches R, Newbold C (2000) A simple RNA analysis method shows var and rif multigene family expression patterns in *Plasmodium falciparum*. Mol Biochem Parasitol 105 (2):311–315, S0166-6851(99)00193-0 [pii]

Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, Thera MA, Kone AK, Doumbo OK, Plowe CV, Rowe JA (2006) Differential *var* gene transcription in *Plasmodium falciparum* isolates from patients with cerebral malaria compared to hyperparasitaemia. Mol Biochem Parasitol 150(2):211–218

Langreth SG, Peterson E (1985) Pathogenicity, stability, and immunogenicity of a knobless clone of *Plasmodium falciparum* in Colombian owl monkeys. Infect Immun 47(3):760–766

Lanzer M, Wickert H, Krohne G, Vincensini L, Braun Breton C (2006) Maurer's clefts: a novel multi-functional organelle in the cytoplasm of *Plasmodium falciparum*-infected erythrocytes. Int J Parasitol 36(1):23–36

Lau CK, Turner L, Jespersen JS, Lowe ED, Petersen B, Wang CW, Petersen JE, Lusingu J, Theander TG, Lavstsen T, Higgins MK (2015) Structural conservation despite huge sequence diversity allows EPCR binding by the PfEMP1 family implicated in severe childhood malaria. Cell Host Microbe 17(1):118–129. doi:10.1016/j.chom.2014.11.007

Lauer SA, Rathod PK, Ghori N, Haldar K (1997) A membrane network for nutrient import in red cells infected with the malaria parasite. Science 276(5315):1122–1125

Lauer S, VanWye J, Harrison T, McManus H, Samuel BU, Hiller NL, Mohandas N, Haldar K (2000) Vacuolar uptake of host components, and a role for cholesterol and sphingomyelin in malarial infection. EMBO J 19(14):3556–3564

Lavazec C, Sanyal S, Templeton TJ (2007) Expression switching in the stevor and Pfmc-2TM superfamilies in *Plasmodium falciparum*. Mol Microbiol 64(6):1621–1634

Lavstsen T, Salanti A, Jensen AT, Arnot DE, Theander TG (2003) Sub-grouping of *Plasmodium falciparum* 3D7 *var* genes based on sequence analysis of coding and non-coding regions. Malar J 2(1):27

Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, Wang CW, Berger SS, Baraka V, Marquard AM, Seguin-Orlando A, Willerslev E, Gilbert MT, Lusingu J, Theander TG (2012) Plasmodium falciparum erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. Proc Natl Acad Sci USA 109(26):E1791–E1800. doi:10.1073/pnas.1120455109, 1120455109 [pii]

Lemieux JE, Kyes SA, Otto TD, Feller AI, Eastman RT, Pinches RA, Berriman M, Su XZ, Newbold CI (2013) Genome-wide profiling of chromosome interactions in *Plasmodium falciparum* characterizes nuclear architecture and reconfigurations associated with antigenic variation. Mol Microbiol. doi:10.1111/mmi.12381

Lobo CA, Rodriguez M, Reid M, Lustigman S (2003) Glycophorin C is the receptor for the *Plasmodium falciparum* erythrocyte binding ligand PfEBP-2 (baebl). Blood 101 (11):4628–4631

Lopaticki S, Maier AG, Thompson J, Wilson DW, Tham WH, Triglia T, Gout A, Speed TP, Beeson JG, Healer J, Cowman AF (2011) Reticulocyte and erythrocyte binding-like proteins function cooperatively in invasion of human erythrocytes by malaria parasites. Infect Immun 79(3):1107–1117

Lopez-Rubio JJ, Riviere L, Scherf A (2007) Shared epigenetic mechanisms control virulence factors in protozoan parasites. Curr Opin Microbiol 10(6):560–568

Lopez-Rubio JJ, Mancio-Silva L, Scherf A (2009) Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 5(2):179–190. doi:10.1016/j.chom.2008.12.012, S1931-3128(09)00030-4 [pii]

MacPherson GG, Warrell MJ, White NJ, Looareesuwan S, Warrell DA (1985) Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration. Am J Pathol 119:385–401

Magistrado PA, Lusingu J, Vestergaard LS, Lemnge M, Lavstsen T, Turner L, Hviid L, Jensen AT, Theander TG (2007) Immunoglobulin G antibody reactivity to a group A Plasmodium falciparum erythrocyte membrane protein 1 and protection from P. falciparum malaria. Infect Immun 75(5):2415–2420. doi:10.1128/IAI.00951-06, IAI.00951-06 [pii]

Maier AG, Duraisingh MT, Reeder JC, Patel SS, Kazura JW, Zimmerman PA, Cowman AF (2003) *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. Nat Med 9(1):87–92

Marsh K, Otoo L, Hayes RJ, Carson DC, Greenwood BM (1989) Antibodies to blood stage antigens of *Plasmodium falciparum* in rural Gambians and their relation to protection against infection. Trans R Soc Trop Med Hyg 83(3):293–303

Marti M, Good RT, Rug M, Knuepfer E, Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 306(5703):1930–1933

Marti M, Baum J, Rug M, Tilley L, Cowman AF (2005) Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. J Cell Biol 171(4):587–592

Marty AJ, Thompson JK, Duffy MF, Voss TS, Cowman AF, Crabb BS (2006) Evidence that *Plasmodium falciparum* chromosome end clusters are cross-linked by protein and are the sites of both virulence gene silencing and activation. Mol Microbiol 62(1):72–83

Matthews K, Kalanon M, Chisholm SA, Sturm A, Goodman CD, Dixon MW, Sanders PR, Nebl T, Fraser F, Haase S, McFadden GI, Gilson PR, Crabb BS, de Koning-Ward TF (2013) The *Plasmodium* translocon of exported proteins (PTEX) component thioredoxin-2 is important for maintaining normal blood-stage growth. Mol Microbiol 89(6):1167–1186

Mayer DC, Cofie J, Jiang L, Hartl DL, Tracy E, Kabat J, Mendoza LH, Miller LH (2009) Glycophorin B is the erythrocyte receptor of Plasmodium falciparum erythrocyte-binding ligand, EBL-1. Proc Natl Acad Sci USA 106(13):5348–5352

McBride JS, Heidrich HG (1987) Fragments of the polymorphic Mr 185,000 glycoprotein from the surface of isolated *Plasmodium falciparum* merozoites form an antigenic complex. Mol Biochem Parasitol 23(1):71–84

McGregor IA (1974) Mechanisms of acquired immunity and epidemiological patterns of antibody responses in malaria in man. Bull World Health Organ 50(3-4):259–266

McMillan PJ, Millet C, Batinovic S, Maiorca M, Hanssen E, Kenny S, Muhle RA, Melcher M, Fidock DA, Smith JD, Dixon MW, Tilley L (2013) Spatial and temporal mapping of the PfEMP1 export pathway in *Plasmodium falciparum*. Cell Microbiol 15(8):1401–1418. doi:10.1111/cmi.12125

Menard R, Tavares J, Cockburn I, Markus M, Zavala F, Amino R (2013) Looking under the skin: the first steps in malarial infection and immunity. Nat Rev Microbiol 11(10):701–712. doi:10.1038/nrmicro3111, nrmicro3111 [pii]

Miller LH, Ackerman HC, Su XZ, Wellems TE (2013) Malaria biology and disease pathogenesis: insights for new treatments. Nat Med 19(2):156–167. doi:10.1038/nm.3073, m.3073 [pii]

Mills JP, Diez-Silva M, Quinn DJ, Dao M, Lang MJ, Tan KS, Lim CT, Milon G, David PH, Mercereau-Puijalon O, Bonnefoy S, Suresh S (2007) Effect of plasmodial RESA protein on

deformability of human red blood cells harboring *Plasmodium falciparum*. Proc Natl Acad Sci USA 104(22):9213–9217

Mota MM, Hafalla JC, Rodriguez A (2002) Migration through host cells activates *Plasmodium* sporozoites for infection. Nat Med 8(11):1318–1322

Moxon CA, Wassmer SC, Milner DA Jr, Chisala NV, Taylor TE, Seydel KB, Molyneux ME, Faragher B, Esmon CT, Downey C, Toh CH, Craig AG, Heyderman RS (2013) Loss of endothelial protein C receptors links coagulation and inflammation to parasite sequestration in cerebral malaria in African children. Blood 122(5):842–851. doi:10.1182/blood-2013-03-490219

Mphande FA, Ribacke U, Kaneko O, Kironde F, Winter G, Wahlgren M (2008) SURFIN4.1, a schizont-merozoite associated protein in the SURFIN family of *Plasmodium falciparum*. Malar J 7:116

Murphy SC, Samuel BU, Harrison T, Speicher KD, Speicher DW, Reid ME, Prohaska R, Low PS, Tanner MJ, Mohandas N, Haldar K (2004) Erythrocyte detergent-resistant membrane proteins: their characterization and selective uptake during malarial infection. Blood 103(5):1920–1928

Nagao E, Seydel KB, Dvorak JA (2002) Detergent-resistant erythrocyte membrane rafts are modified by a *Plasmodium falciparum* infection. Exp Parasitol 102(1):57–59

Newbold CI, Pinches R, Roberts DJ, Marsh K (1992) *Plasmodium falciparum*: the human agglutinating antibody response to the infected red cell surface is predominantly variant specific. Exp Parasitol 75(3):281–292

Newbold C, Warn P, Black G, Berendt A, Craig A, Snow B, Msobo M, Peshu N, Marsh K (1997) Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. Am J Trop Med Hyg 57:389–398

Nguansangiam S, Day NP, Hien TT, Mai NT, Chaisri U, Riganti M, Dondorp AM, Lee SJ, Phu NH, Turner GD, White NJ, Ferguson DJ, Pongponratn E (2007) A quantitative ultrastructural study of renal pathology in fatal *Plasmodium falciparum* malaria. Trop Med Int Health 12 (9):1037–1050. doi:10.1111/j.1365-3156.2007.01881.x, TMI1881 [pii]

Niang M, Yan Yam X, Preiser PR (2009) The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. PLoS Pathog 5(2), e1000307. doi:10.1371/journal.ppat.1000307

Niang M, Bei AK, Madnani KG, Pelly S, Dankwa S, Kanjee U, Gunalan K, Amaladoss A, Yeo KP, Bob NS, Malleret B, Duraisingh MT, Preiser PR (2014) STEVOR is a *Plasmodium falciparum* erythrocyte binding protein that mediates merozoite invasion and rosetting. Cell Host Microbe 16(1):81–93

Nielsen MA, Staalsoe T, Kurtzhals JA, Goka BQ, Dodoo D, Alifrangis M, Theander TG, Akanmori BD, Hviid L (2002) *Plasmodium falciparum* variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. J Immunol 168(7):3444–3450

Normark J, Nilsson D, Ribacke U, Winter G, Moll K, Wheelock CE, Bayarugaba J, Kironde F, Egwang TG, Chen Q, Andersson B, Wahlgren M (2007) PfEMP1-DBL1alpha amino acid motifs in severe disease states of *Plasmodium falciparum* malaria. Proc Natl Acad Sci USA 104(40):15835–15840

O'Donnell RA, Freitas-Junior LH, Preiser PR, Williamson DH, Duraisingh M, McElwain TF, Scherf A, Cowman AF, Crabb BS (2002) A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of *Plasmodium falciparum* chromosomes. EMBO J 21(5):1231–1239

Ochola LB, Siddondo BR, Ocholla H, Nkya S, Kimani EN, Williams TN, Makale JO, Liljander A, Urban BC, Bull PC, Szestak T, Marsh K, Craig AG (2011) Specific receptor usage in *Plasmodium falciparum* cytoadherence is associated with disease outcome. PLoS One 6(3), e14741. doi:10.1371/journal.pone.0014741

Ogutu BR, Apollo OJ, McKinney D, Okoth W, Siangla J, Dubovsky F, Tucker K, Waitumbi JN, Diggs C, Wittes J, Malkin E, Leach A, Soisson LA, Milman JB, Otieno L, Holland CA, Polhemus M, Remich SA, Ockenhouse CF, Cohen J, Ballou WR, Martin SK, Angov E, Stewart

VA, Lyon JA, Heppner DG, Withers MR (2009) Blood stage malaria vaccine eliciting high antigen-specific antibody concentrations confers no protection to young children in Western Kenya. PLoS One 4(3), e4708

Oh SS, Voigt S, Fisher D, Yi SJ, LeRoy PJ, Derick LH, Liu S, Chishti AH (2000) *Plasmodium falciparum* erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. Mol Biochem Parasitol 108 (2):237–247

Oleinikov AV, Amos E, Frye IT, Rossnagle E, Mutabingwa TK, Fried M, Duffy PE (2009) High throughput functional assays of the variant antigen PfEMP1 reveal a single domain in the 3D7 *Plasmodium falciparum* genome that binds ICAM1 with high affinity and is targeted by naturally acquired neutralizing antibodies. PLoS Pathog 5(4), e1000386. doi:10.1371/journal.ppat.1000386

Otto TD, Rayner JC, Bohme U, Pain A, Spottiswoode N, Sanders M, Quail M, Ollomo B, Renaud F, Thomas AW, Prugnolle F, Conway DJ, Newbold C, Berriman M (2014) Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. Nat Commun 5:4754. doi:10.1038/ncomms5754

Pachlatko E, Rusch S, Muller A, Hemphill A, Tilley L, Hanssen E, Beck HP (2010) MAHRP2, an exported protein of *Plasmodium falciparum*, is an essential component of Maurer's cleft tethers. Mol Microbiol 77(5):1136–1152

Pain A, Ferguson DJ, Kai O, Urban BC, Lowe B, Marsh K, Roberts DJ (2001) Platelet-mediated clumping of *Plasmodium falciparum*-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. Proc Natl Acad Sci USA 98(4):1805–1810

Pamplona A, Ferreira A, Balla J, Jeney V, Balla G, Epiphanio S, Chora A, Rodrigues CD, Gregoire IP, Cunha-Rodrigues M, Portugal S, Soares MP, Mota MM (2007) Heme oxygenase-1 and carbon monoxide suppress the pathogenesis of experimental cerebral malaria. Nat Med 13 (6):703–710

Pei X, An X, Guo X, Tarnawski M, Coppel R, Mohandas N (2005) Structural and functional studies of interaction between *Plasmodium falciparum* knob-associated histidine-rich protein (KAHRP) and erythrocyte spectrin. J Biol Chem 280(35):31166–31171

Pei X, Guo X, Coppel R, Bhattacharjee S, Haldar K, Gratzer W, Mohandas N, An X (2007) The ring-infected erythrocyte surface antigen (RESA) of *Plasmodium falciparum* stabilizes spectrin tetramers and suppresses further invasion. Blood 110(3):1036–1042

Persson KE, Fowkes FJ, McCallum FJ, Gicheru N, Reiling L, Richards JS, Wilson DW, Lopaticki S, Cowman AF, Marsh K, Beeson JG (2013) Erythrocyte-binding antigens of Plasmodium falciparum are targets of human inhibitory antibodies and function to evade naturally acquired immunity. J Immunol 191(2):785–794

Petter M, Haeggstroem M, Khattab A, Fernandez V, Klinkert M-Q, Wahlgren M (2007) Variant proteins of the *Plasmodium falciparum* RIFIN family show distinct subcellular localization and developmental expression patterns. Mol Biochem Parasitol 156(1):51–61. doi:10.1016/j.molbiopara.2007.07.011

Petter M, Bonow I, Klinkert MQ (2008) Diverse expression patterns of subgroups of the rif multigene family during *Plasmodium falciparum* gametocytogenesis. PLoS One 3(11), e3779. doi:10.1371/journal.pone.0003779

Petter M, Lee CC, Byrne TJ, Boysen KE, Volz J, Ralph SA, Cowman AF, Brown GV, Duffy MF (2011) Expression of *P. falciparum* var genes involves exchange of the histone variant H2A.Z at the promoter. PLoS Pathog 7(2), e1001292. doi:10.1371/journal.ppat.1001292

Petter M, Selvarajah SA, Lee CC, Chin WH, Gupta AP, Bozdech Z, Brown GV, Duffy MF (2013) H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite *Plasmodium falciparum*. Mol Microbiol 87(6):1167–82. doi:10.1111/mmi.12154

Pologe LG, Ravetch JV (1986) A chromosomal rearrangement in a *P. falciparum* histidine-rich protein gene is associated with the knobless phenotype. Nature 322(6078):474–477

Pongponratn E, Turner GD, Day NP, Phu NH, Simpson JA, Stepniewska K, Mai NT, Viriyavejakul P, Looareesuwan S, Hien TT, Ferguson DJ, White NJ (2003) An ultrastructural study of the brain in fatal *Plasmodium falciparum* malaria. Am J Trop Med Hyg 69(4):345–359

Preiser PR, Khan S, Costa FT, Jarra W, Belnoue E, Ogun S, Holder AA, Voza T, Landau I, Snounou G, Renia L (2002) Stage-specific transcription of distinct repertoires of a multigene family during *Plasmodium* life cycle. Science 295(5553):342–345

Proellocks NI, Coppel RL, Waller KL (2010) Dissecting the apicomplexan rhoptry neck proteins. Trends Parasitol 26(6):297–304

Prommano O, Chaisri U, Turner GD, Wilairatana P, Ferguson DJ, Viriyavejakul P, White NJ, Pongponratn E (2005) A quantitative ultrastructural study of the liver and the spleen in fatal falciparum malaria. Southeast Asian J Trop Med Public Health 36(6):1359–1370

Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T (2010) *Plasmodium falciparum* erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. PLoS Comput Biol 6(9). doi:10.1371/journal.pcbi.1000933

Rayner JC, Galinski MR, Ingravallo P, Barnwell JW (2000) Two Plasmodium falciparum genes express merozoite proteins that are related to Plasmodium vivax and *Plasmodium yoelii* adhesive proteins involved in host cell selection and invasion. Proc Natl Acad Sci USA 97 (17):9648–9653

Richards JS, Arumugam TU, Reiling L, Healer J, Hodder AN, Fowkes FJ, Cross N, Langer C, Takeo S, Uboldi AD, Thompson JK, Gilson PR, Coppel RL, Siba PM, King CL, Torii M, Chitnis CE, Narum DL, Mueller I, Crabb BS, Cowman AF, Tsuboi T, Beeson JG (2013) Identification and prioritization of merozoite antigens as targets of protective human immunity to *Plasmodium falciparum* malaria for vaccine and biomarker development. J Immunol 191 (2):795–809

Riglar DT, Richard D, Wilson DW, Boyle MJ, Dekiwadia C, Turnbull L, Angrisano F, Marapana DS, Rogers KL, Whitchurch CB, Beeson JG, Cowman AF, Ralph SA, Baum J (2011) Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. Cell Host Microbe 9(1):9–20

Roberts DJ, Pain A, Kai O, Kortok M, Marsh K (2000) Autoagglutination of malaria-infected red blood cells and malaria severity. Lancet 355(9213):1427–1428

Robinson BA, Welch TL, Smith JD (2003) Widespread functional specialization of *Plasmodium falciparum* erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. Mol Microbiol 47(5):1265–1278

Rogerson SJ, Tembenu R, Dobano C, Plitt S, Taylor TE, Molyneux ME (1999) Cytoadherence characteristics of *Plasmodium falciparum*-infected erythrocytes from Malawian children with severe and uncomplicated malaria. Am J Trop Med Hyg 61:467–472

Rorick MM, Rask TS, Baskerville EB, Day KP, Pascual M (2013) Homology blocks of *Plasmodium falciparum* var genes and clinically distinct forms of severe malaria in a local population. BMC Microbiol 13:244. doi:10.1186/1471-2180-13-244

Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen AT, Muller D, Theander T, Beck HP (2006) Differential expression of *var* gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. Infect Immun 74(7):3904–3911

Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, Ribas de Pouplana L, Preiser PR, Bozdech Z, Cortes A (2012) Transcriptional variation in the malaria parasite *Plasmodium falciparum*. Genome Res 22(5):925–938. doi:10.1101/gr.129692.111, gr.129692.111 [pii]

Rowe JA, Moulds JM, Newbold CI, Miller LH (1997) P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. Nature 388 (6639):292–295

Rowe JA, Claessens A, Corrigan RA, Arman M (2009) Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. Expert Rev Mol Med 11, e16. doi:10.1017/S1462399409001082

Rug M, Prescott SW, Fernandez KM, Cooke BM, Cowman AF (2006) The role of KAHRP domains in knob formation and cytoadherence of infected human erythrocytes. Blood 108 (1):370–378

Rug M, Cyrklaff M, Mikkonen A, Lemgruber L, Kuelzer S, Sanchez CP, Thompson J, Hanssen E, O'Neill M, Langer C, Lanzer M, Frischknecht F, Maier AG, Cowman AF (2014) Export of virulence proteins by malaria-infected erythrocytes involves remodeling of host actin cytoskeleton. Blood 124(23):3459–3468. doi:10.1182/blood-2014-06-583054

Russo I, Babbitt S, Muralidharan V, Butler T, Oksman A, Goldberg DE (2010) Plasmepsin V licenses *Plasmodium* proteins for export into the host erythrocyte. Nature 463(7281):632–636, S0020-7519(06)00397-3 [pii]

Salanti A, Staalsoe T, Lavstsen T, Jensen AT, Sowa MP, Arnot DE, Hviid L, Theander TG (2003) Selective upregulation of a single distinctly structured *var* gene in chondroitin sulphate A-adhering *Plasmodium falciparum* involved in pregnancy-associated malaria. Mol Microbiol 49(1):179–191

Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AM, Cohen AM, Janssen-Megens EM, van de Vegte-Bolmer M, Selzer RR, Iniguez AL, Green RD, Sauerwein RW, Jensen ON, Stunnenberg HG (2009) Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. Proc Natl Acad Sci USA 106 (24):9655–9660. doi:10.1073/pnas.0902515106, 0902515106 [pii]

Sander AF, Lavstsen T, Rask TS, Lisby M, Salanti A, Fordyce SL, Jespersen JS, Carter R, Deitsch KW, Theander TG, Pedersen AG, Arnot DE (2014) DNA secondary structures are associated with recombination in major *Plasmodium falciparum* variable surface antigen gene families. Nucleic Acids Res 42(4):2270–2281. doi:10.1093/nar/gkt1174

Sanyal S, Egee S, Bouyer G, Perrot S, Safeukui I, Bischoff E, Buffet P, Deitsch KW, Mercereau-Puijalon O, David PH, Templeton TJ, Lavazec C (2012) *Plasmodium falciparum* STEVOR proteins impact erythrocyte mechanical properties. Blood 119(2):e1–e8

Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, Gysin J, Lanzer M (1998) Antigenic variation in malaria: *in situ* switching, relaxed and mutally exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. EMBO J 17:5418–5426

Schreiber N, Brattig N, Evans J, Tsiri A, Horstmann RD, May J, Klinkert MQ (2006) Cerebral malaria is associated with IgG2 and IgG4 antibody responses to recombinant *Plasmodium falciparum* RIFIN antigen. Microbes Infect 8(5):1269–1276. doi:10.1016/j.micinf.2005.12.007, S1286-4579(06)00017-7 [pii]

Schreiber N, Khattab A, Petter M, Marks F, Adjei S, Kobbe R, May J, Klinkert MQ (2008) Expression of *Plasmodium falciparum* 3D7 STEVOR proteins for evaluation of antibody responses following malaria infections in naive infants. Parasitology 135(2):155–167. doi:10.1017/S0031182007003794, S0031182007003794 [pii]

Sharp S, Lavstsen T, Fivelman QL, Saeed M, McRobert L, Templeton TJ, Jensen AT, Baker DA, Theander TG, Sutherland CJ (2006) Programmed transcription of the var gene family, but not of stevor, in *Plasmodium falciparum* gametocytes. Eukaryot cell 5(8):1206–1214

Shikani HJ, Freeman BD, Lisanti MP, Weiss LM, Tanowitz HB, Desruisseaux MS (2012) Cerebral malaria: we have come a long way. Am J Pathol 181(5):1484–1492. doi:10.1016/j.ajpath.2012.08.010

Siau A, Toure FS, Ouwe-Missi-Oukem-Boyer O, Ciceron L, Mahmoudi N, Vaquero C, Froissard P, Bisvigou U, Bisser S, Coppee JY, Bischoff E, David PH, Mazier D (2007) Whole-transcriptome analysis of *Plasmodium falciparum* field isolates: identification of new pathogenicity factors. J Infect Dis 196(11):1603–1612

Silamut K, Phu NH, Whitty C, Turner GDH, Louwrier K, Mai NTH, Simpson JA, Hien TT, White NJ (1999) A quantitative analysis of the microvascular sequestration of malaria parasites in the human brain. Am J Pathol 155:395–410

Sim BK, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH (1994) Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. Science 264(5167):1941–1944

Singh B, Kim Sung L, Matusop A, Radhakrishnan A, Shamsul SS, Cox-Singh J, Thomas A, Conway DJ (2004) A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. Lancet 363(9414):1017–1024

Singh SK, Hora R, Belrhali H, Chitnis CE, Sharma A (2006) Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. Nature 439(7077):741–744. doi:10.1038/nature04443

Singh S, Alam MM, Pal-Bhowmick I, Brzostowski JA, Chitnis CE (2010) Distinct external signals trigger sequential release of apical organelles during erythrocyte invasion by malaria parasites. PLoS Pathog 6(2), e1000746

Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, Peterson DS, Pinches R, Newbold CI, Miller LH (1995) Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell 82(1):101–110

Smith JD, Craig AG, Kriek N, Hudson-Taylor D, Kyes S, Fagen T, Pinches R, Baruch DI, Newbold CI, Miller LH (2000a) Identification of a *Plasmodium falciparum* intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. Proc Natl Acad Sci USA 97:1766–1771

Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH (2000b) Classification of adhesive domains in the *Plasmodium falciparum* erythrocyte membrane protein 1 family. Mol Biochem Parasitol 110:293–310

Smith JD, Rowe JA, Higgins MK, Lavstsen T (2013) Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. Cell Microbiol 15(12):1976–1983. doi:10.1111/cmi.12183

Spielmann T, Fergusen DJ, Beck HP (2003) etramps, a new Plasmodium falciparum gene family coding for developmentally regulated and highly charged membrane proteins located at the parasite-host cell interface. Mol Biol Cell 14(4):1529–1544

Spielmann T, Hawthorne PL, Dixon MW, Hannemann M, Klotz K, Kemp DJ, Klonis N, Tilley L, Trenholme KR, Gardiner DL (2006) A cluster of ring stage-specific genes linked to a locus implicated in cytoadherence in *Plasmodium falciparum* codes for PEXEL-negative and PEXEL-positive proteins exported into the host cell. Mol Biol Cell 17(8):3613–3624

Stubbs J, Simpson KM, Triglia T, Plouffe D, Tonkin CJ, Duraisingh MT, Maier AG, Winzeler EA, Cowman AF (2005) Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. Science 309(5739):1384–1387

Sturm A, Amino R, van de Sand C, Regen T, Retzlaff S, Rennenberg A, Krueger A, Pollok JM, Menard R, Heussler VT (2006) Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. Science 313(5791):1287–1290

Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, Ravetch JA, Wellems TE (1995) The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell 82(1):89–100

Taylor TE, Fu WJ, Carr RA, Whitten RO, Mueller JG, Fosiko NG, Lewallen S, Liomba NG, Molyneux ME (2004) Differentiating the pathologies of cerebral malaria by postmortem parasite counts. Nat Med 10(2):143–145

Tham WH, Payne PD, Brown GV, Rogerson SJ (2007) Identification of basic transcriptional elements required for rif gene transcription. Int J Parasitol 37(6):605–615. doi:10.1016/j.ijpara.2006.11.006

Tham WH, Wilson DW, Lopaticki S, Schmidt CQ, Tetteh-Quarcoo PB, Barlow PN, Richard D, Corbin JE, Beeson JG, Cowman AF (2010) Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum* PfRh4 invasion ligand. Proc Natl Acad Sci USA 107(40):17327–17332

Tiburcio M, Niang M, Deplaine G, Perrot S, Bischoff E, Ndour PA, Silvestrini F, Khattab A, Milon G, David PH, Hardeman M, Vernick KD, Sauerwein RW, Preiser PR, Mercereau-Puijalon O, Buffet P, Alano P, Lavazec C (2012) A switch in infected erythrocyte

deformability at the maturation and blood circulation of *Plasmodium falciparum* transmission stages. Blood 119(24):e172–e180

Tolia NH, Enemark EJ, Sim BK, Joshua-Tor L (2005) Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. Cell 122(2):183–193

Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, Hommel M, Duffy MF, Silva LM, Scherf A, Ivens A, Speed TP, Beeson JG, Cowman AF (2009) Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. PLoS Biol 7(4), e84. doi:10.1371/journal.pbio.1000084, 08-PLBI-RA-5273 [pii]

Torii M, Adams JH, Miller LH, Aikawa M (1989) Release of merozoite dense granules during erythrocyte invasion by *Plasmodium knowlesi*. Infect Immun 57(10):3230–3233

Trager W, Rozario C, Shio H, Williams J, Perkins ME (1992) Transfer of a dense granule protein of *Plasmodium falciparum* to the membrane of ring stages and isolation of dense granules. Infect Immun 60(11):4656–4661

Trimnell AR, Kraemer SM, Mukherjee S, Phippard DJ, Janes JH, Flamoe E, Su XZ, Awadalla P, Smith JD (2006) Global genetic diversity and evolution of var genes associated with placental and severe childhood malaria. Mol Biochem Parasitol 148(2):169–180

Tripathi AK, Sullivan DJ, Stins MF (2006) *Plasmodium falciparum*-infected erythrocytes increase intercellular adhesion molecule 1 expression on brain endothelium through NF-kappaB. Infect Immun 74(6):3262–3270. doi:10.1128/IAI.01625-05

Tripathi AK, Sullivan DJ, Stins MF (2007) *Plasmodium falciparum*-infected erythrocytes decrease the integrity of human blood-brain barrier endothelial cell monolayers. J Infect Dis 195 (7):942–950

Turner GDH, Morrison H, Jones M, Davis TME, Looareesuwan S, Buley ID, Gatter KC, Newbold CI, Pukritayakamee S, Nagachinta B, White NJ, Berendt AR (1994) An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. Am J Pathol 145:1057–1069

Turner L, Wang CW, Lavstsen T, Mwakalinga SB, Sauerwein RW, Hermsen CC, Theander TG (2011) Antibodies against PfEMP1, RIFIN, MSP3 and GLURP are acquired during controlled Plasmodium falciparum malaria infections in naive volunteers. PLoS One 6(12), e29025

Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JE, Avril M, Brazier AJ, Freeth J, Jespersen JS, Nielsen MA, Magistrado P, Lusingu J, Smith JD, Higgins MK, Theander TG (2013) Severe malaria is associated with parasite binding to endothelial protein C receptor. Nature 498 (7455):502–505. doi:10.1038/nature12216, nature12216 [pii]

Udomsangpetch R, Wahlin B, Carlson J, Berzins K, Torii M, Aikawa M, Perlmann P, Wahlgren M (1989) *Plasmodium falciparum*-infected erythrocytes form spontaneous erythrocyte rosettes. J Exp Med 169(5):1835–1840

Umbers AJ, Boeuf P, Clapham C, Stanisic DI, Baiwog F, Mueller I, Siba P, King CL, Beeson JG, Glazier J, Rogerson SJ (2011) Placental malaria-associated inflammation disturbs the insulin-like growth factor axis of fetal growth regulation. J Infect Dis 203(4):561–569

van der Heyde HC, Nolan J, Combes V, Gramaglia I, Grau GE (2006) A unified hypothesis for the genesis of cerebral malaria: sequestration, inflammation and hemostasis leading to microcirculatory dysfunction. Trends Parasitol 22(11):503–508

Volz JC, Bartfai R, Petter M, Langer C, Josling GA, Tsuboi T, Schwach F, Baum J, Rayner JC, Stunnenberg HG, Duffy MF, Cowman AF (2012) PfSET10, a *Plasmodium falciparum* Methyltransferase, Maintains the Active var Gene in a Poised State during Parasite Division. Cell Host Microbe 11(1):7–18. doi:10.1016/j.chom.2011.11.011, S1931-3128(11)00403-3 [pii]

Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, Beeson JG, Reeder JC, Crabb BS, Cowman AF (2006) A *var* gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. Nature 439(7079):1004–1009

Voss TS, Tonkin CJ, Marty AJ, Thompson JK, Healer J, Crabb BS, Cowman AF (2007) Alterations in local chromatin environment are involved in silencing and activation of subtelomeric *var* genes in *Plasmodium falciparum*. Mol Microbiol 66(1):139–150

Waller KL, Cooke BM, Nunomura W, Mohandas N, Coppel RL (1999) Mapping the binding domains involved in the interaction between the *Plasmodium falciparum* knob-associated histidine-rich protein (KAHRP) and the cytoadherence ligand P. falciparum erythrocyte membrane protein 1 (PfEMP1). J Biol Chem 274(34):23808–23813

Wang CW, Magistrado PA, Nielsen MA, Theander TG, Lavstsen T (2008) Preferential transcription of conserved rif genes in two phenotypically distinct *Plasmodium falciparum* parasite lines. Int J Parasitol. doi:10.1016/j.ijpara.2008.11.014, S0020-7519(08)00480-3 [pii]

Wang CW, Hermsen CC, Sauerwein RW, Arnot DE, Theander TG, Lavstsen T (2009) The *Plasmodium falciparum* var gene transcription strategy at the onset of blood stage infection in a human volunteer. Parasitol Int 58(4):478–480. doi:10.1016/j.parint.2009.07.004, S1383-5769(09)00076-2 [pii]

Ward GE, Miller LH, Dvorak JA (1993) The origin of parasitophorous vacuole membrane lipids in malaria-infected erythrocytes. J Cell Sci 106(Pt 1):237–248

Warimwe GM, Keane TM, Fegan G, Musyoki JN, Newton CR, Pain A, Berriman M, Marsh K, Bull PC (2009) *Plasmodium falciparum* var gene expression is modified by host immunity. Proc Natl Acad Sci USA 106(51):21801–21806. doi:10.1073/pnas.0907590106, 0907590106 [pii]

Warimwe GM, Fegan G, Musyoki JN, Newton CR, Opiyo M, Githinji G, Andisi C, Menza F, Kitsao B, Marsh K, Bull PC (2012) Prognostic indicators of life-threatening malaria are associated with distinct parasite variant antigen profiles. Sci Transl Med 4(129):129ra145. doi:10.1126/scitranslmed.3003247, 4/129/129ra45 [pii]

Waterkeyn JG, Wickham ME, Davern KM, Cooke BM, Coppel RL, Reeder JC, Culvenor JG, Waller RF, Cowman AF (2000) Targeted mutagenesis of *Plasmodium falciparum* erythrocyte membrane protein 3 (PfEMP3) disrupts cytoadherence of malaria-infected red blood cells. EMBO J 19(12):2813–2823

Weber JL (1988) Interspersed repetitive DNA from *Plasmodium falciparum*. Mol Biochem Parasitol 29(2-3):117–124

Weinberg JB, Lopansri BK, Mwaikambo E, Granger DL (2008) Arginine, nitric oxide, carbon monoxide, and endothelial function in severe malaria. Curr Opin Infect Dis 21(5):468–475

Weiser S, Miu J, Ball HJ, Hunt NH (2007) Interferon-gamma synergises with tumour necrosis factor and lymphotoxin-alpha to enhance the mRNA and protein expression of adhesion molecules in mouse brain endothelial cells. Cytokine 37(1):84–91

White NJ, Turner GD, Day NP, Dondorp AM (2013) Lethal malaria: Marchiafava and Bignami were right. J Infect Dis 208(2):192–198. doi:10.1093/infdis/jit116

Wickert H, Wissing F, Andrews KT, Stich A, Krohne G, Lanzer M (2003) Evidence for trafficking of PfEMP1 to the surface of *P. falciparum*-infected erythrocytes via a complex membrane network. Eur J Cell Biol 82(6):271–284

Wickert H, Gottler W, Krohne G, Lanzer M (2004) Maurer's cleft organization in the cytoplasm of infected erythrocytes: new insights from three-dimensional reconstruction of serial ultrathin sections. Eur J Cell Biol 83(10):567–582

Wickham ME, Rug M, Ralph SA, Klonis N, McFadden GI, Tilley L, Cowman AF (2001) Trafficking and assembly of the cytoadherence complex in Plasmodium falciparum-infected human erythrocytes. EMBO J 20(20):5636–5649

Winter G, Kawai S, Haeggstrom M, Kaneko O, von Euler A, Kawazu S, Palm D, Fernandez V, Wahlgren M (2005) SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. J Exp Med 201(11):1853–1863. doi:10.1084/jem.20041392, jem.20041392 [pii]

Witmer K, Schmid CD, Brancucci NM, Luah YH, Preiser PR, Bozdech Z, Voss TS (2012) Analysis of subtelomeric virulence gene families in *Plasmodium falciparum* by comparative transcriptional profiling. Mol Microbiol 84(2):243–259. doi:10.1111/j.1365-2958.2012.08019.x

Yalcindag E, Rougeron V, Elguero E, Arnathau C, Durand P, Brisse S, Diancourt L, Aubouy A, Becquart P, D'Alessandro U, Fontenille D, Gamboa D, Maestre A, Menard D, Musset L, Noya O, Veron V, Wide A, Carme B, Legrand E, Chevillon C, Ayala FJ, Renaud F, Prugnolle F (2014) Patterns of selection on *Plasmodium falciparum* erythrocyte-binding antigens after the colonization of the New World. Mol Ecol 23(8):1979–1993

Yeo TW, Lampah DA, Gitawati R, Tjitra E, Kenangalem E, Piera K, Price RN, Duffull SB, Celermajer DS, Anstey NM (2008) Angiopoietin-2 is associated with decreased endothelial nitric oxide and poor clinical outcome in severe *falciparum* malaria. Proc Natl Acad Sci USA 105(44):17097–17102. doi:10.1073/pnas.0805782105

Zhang Q, Huang Y, Zhang Y, Fang X, Claes A, Duchateau M, Namane A, Lopez-Rubio JJ, Pan W, Scherf A (2011) A critical role of perinuclear filamentous actin in spatial repositioning and mutually exclusive expression of virulence genes in malaria parasites. Cell Host Microbe 10 (5):451–463. doi:10.1016/j.chom.2011.09.013, S1931-3128(11)00332-5 [pii]

Zhang Q, Siegel TN, Martins RM, Wang F, Cao J, Gao Q, Cheng X, Jiang L, Hon CC, Scheidig-Benatar C, Sakamoto H, Turner L, Jensen AT, Claes A, Guizetti J, Malmquist NA, Scherf A (2014) Exonuclease-mediated degradation of nascent RNA silences genes linked to severe malaria. Nature 513(7518):431–435. doi:10.1038/nature13468

Zilversmit MM, Chase EK, Chen DS, Awadalla P, Day KP, McVean G (2013) Hypervariable antigen genes in malaria have ancient roots. BMC Evol Biol 13:110. doi:10.1186/1471-2148-13-110

Zuccala ES, Gout AM, Dekiwadia C, Marapana DS, Angrisano F, Turnbull L, Riglar DT, Rogers KL, Whitchurch CB, Ralph SA, Speed TP, Baum J (2012) Subcompartmentalisation of proteins in the rhoptries correlates with ordered events of erythrocyte invasion by the blood stage malaria parasite. PLoS One 7(9), e46160

Polymorphic Mucin-Like Proteins in *Schistosoma mansoni*, a Variable Antigen and a Key Component of the Compatibility Between the Schistosome and Its Snail Host

Benjamin Gourbal, André Théron, Christoph Grunau, David Duval, and Guillaume Mitta

Abstract The arms race between vertebrate hosts and parasites has led to diversification systems able to generate huge repertoires of immune recognition receptors and antigenic variants. Until recently, the invertebrate immunity was considered to be poorly specific, and consequently, antigenic variability was not expected to be high for their respective parasites. In the present chapter, we show how the study of the interaction between the snail *Biomphalaria glabrata* and its parasite *Schistosome mansoni* has shaken this paradigm. We show that the fate of the interaction between the snail and its parasite is at least partly the result of the concordance of highly variable repertoires of immune recognition receptors in the snail and corresponding antigenic variants in the parasite. We call these antigenic variants of the schistosome *Schistosoma mansoni* polymorphic mucins (*Sm*PoMucs). We show that their high level of diversification is the result of a complex cascade of mechanisms, thus presenting evidence for antigenic variation in a parasite infecting an invertebrate species.

1 Introduction

The comprehensive understanding of host–parasite interactions represents a major challenge in evolutionary biology. Because parasites are responsible for substantial deleterious effects, they represent a major driving force for the evolution of their hosts. In parallel, parasites have to coevolve with their host to avoid elimination. This adaptation of the Red Queen hypothesis (Van Valen 1974) to host–parasite systems predicts that an arms race will lead to the evolution of mechanisms that generate diversity and maintain polymorphism of molecules that play a key role in

B. Gourbal • A. Théron • C. Grunau • D. Duval • G. Mitta (✉)
CNRS, Ecologie et Evolution des Interactions, UMR 5244, Université de Perpignan
Via Domitia, 58 avenue Paul Alduy, Perpignan 66860, France
e-mail: benjamin.gourbal@univ-perp.fr; Theron@univ-perp.fr;
christoph.grunau@univ-perp.fr; david.duval@univ-perp.fr; mitta@univ-perp.fr

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_4

host–parasite interplay (Combes 2000; Ebert 2008; Gupta et al. 1996). In vertebrate hosts, the most striking example is the exceptional diversity of antigen-specific receptors of the adaptive immune system of jawed vertebrates. This system relies on somatic gene rearrangement and hypermutation (Di Noia and Neuberger 2007; Tonegawa 1983). It is the ultimate evolutionary outcome of an immune system that is able to recognize all antigens that may arise in a host. In the pathogenic bacteria and viruses of vertebrates, a variety of active and passive mechanisms permitting evasion and/or suppression of the host's immune response were evidenced, and antigenic variation is a widespread strategy (Finlay and McFadden 2006). These different strategies will allow preventing recognition, presentation of the antigens, or blocking immune effector mechanisms. In the case of invertebrate hosts and their parasites, the picture was believed to be completely different since the prevailing view was that invertebrates have no acquired adaptive immunity. Their immune system was considered until recently as innate and "nonspecific" (Medzhitov and Janeway 1997). The detection of parasites by invertebrates was thought to rely exclusively on invariable germline-encoded pattern recognition receptors (PRRs) that recognize pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs). Despite this, the presumably rudimentary immune system of invertebrates has been remarkably successful in evolution. But how was this evolutionary success possible without an adaptive immune system? One reason could be that the adaptive immunity was not advantageous enough for invertebrates considering its fitness cost. This could be due to the different evolutionary K and r selection strategies used by vertebrates and invertebrates, respectively (Pianka 1970), or specific defense strategies using "danger"/"damage" signals (Pradeu and Cooper 2012) helping invertebrates to mount an immune response when recognition failed to initiate immunity. Another possible explanation was that the diversification of invertebrate immune PRRs has been underestimated. Numerous works developed during this last decade have provided evidence for diverse putative immune PRR sequences in jawless vertebrates [lamprey (Pancer et al. 2004)], protochordates [*Amphioxus* (Cannon et al. 2004)], echinoderms [sea urchin (Pancer 2000; Buckley and Rast 2012)], insects (Dong et al. 2006, 2012; Watson et al. 2005), and mollusks [*Biomphalaria glabrata* (Zhang et al. 2004)]. Most of these molecules contain domains known to be shared by vertebrate immune receptors (e.g., IgSF domains). The diversification of these molecules occurs at the individual level (duplications and somatic processes) and population level (polymorphism), and their role in anti-infectious defense was clearly demonstrated for some of them (see Ghosh et al. (2011) for review). These results support the view that invertebrate immunity could be more specific than previously thought. Some species like *B. glabrata* (a Lophotrochozoa) or *Anopheles gambiae* (an Ecdysozoa) are able to express more diversity than their genome contains using probably somatic recombination/hypermutation events (Zhang et al. 2004) or by targeted alternative splicing mechanism (Dong et al. 2006, 2012), respectively. Some of the genes encoding these immune recognition molecules are highly instable; they display haplotype variability, and all these polymorphism levels increase the immune repertoire at the population level (see Ghosh et al. (2011) for review). This means that the immune

recognition molecules encoding genes from invertebrates are submitted to similar selective pressures than those of vertebrates. This is not surprising considering that these immune PRRs are probably exposed to the same antigen diversity as those of vertebrate species. If this holds true, this situation would be the result of diversification processes occurring in the antigen counterpart of their corresponding pathogens. In this context, we sought to verify if antigens of parasites or of specific parasite stages that interact with invertebrate hosts would be subjected to diversification processes. One lock preventing this verification was that no antigen interacting with these new immune recognition receptors from invertebrates had been identified. This was our first objective when we started to study these questions on our biological model, the interaction between *B. glabrata* and *S. mansoni*. After these antigens had been identified, we studied the mechanism governing their diversity. All the results obtained are presented and discussed in the following paragraphs.

2 The Interaction Between *B. glabrata* and *S. mansoni* as a Model to Study the Mechanisms of Antigenic Polymorphism and Diversification

To address these questions, it is necessary to work with an invertebrate–parasite model for which the coevolutionary dynamics is accessible. This is the case in certain host–parasite associations in which only some particular host and parasite phenotypes are compatible, a phenomenon called compatibility polymorphism. It occurs in the interaction between the larval stages (*miracidia*) of the metazoan parasite *Schistosoma mansoni*, the agent of human intestinal schistosomiasis (Gryseels et al. 2006) and its invertebrate intermediate host, the gastropod mollusk *Biomphalaria glabrata* (Basch 1975; Théron et al. 2014). In this interaction, snail–schistosome compatibility was investigated in the field using the whole genetic diversity of wild parasites and snails from the same geographic locality [Guadeloupe, (Théron et al. 2008)]. Numerous naturally infected definitive hosts were used as sources of diversified miracidia, and varying doses of miracidia were used for snail exposures. The main result of this study was that infection rate increased with increasing doses of miracidia and that all experimentally exposed wild snails become infected when 10–20 miracidia/snails are used. We then hypothesized that the probability of infection increases with the dose of miracidia simply because a larger fraction of the phenotypic diversity present in the parasite population is included. In other words, all *B. glabrata* are potentially susceptible to *S. mansoni* and will develop a patent infection if enough phenotypically diverse miracidia are provided. Reciprocally, we speculated that all *S. mansoni* miracidia are potentially infective, if they are exposed to the right individual snail with a matching phenotype. This hypothesis is difficult to verify because the same miracidium cannot be exposed to multiple snails. Consequently, the low natural prevalence of snails with

patent schistosome infection that is usually observed in transmission foci (Anderson and May 1979; Sire et al. 1999) is certainly due to the low probability that a schistosome phenotype encounters its corresponding compatible host phenotype rather than the existence of high level of resistance within host populations. Another result corroborating this hypothesis concerns the different fates of individual parasites that all penetrate the same snail. From histological sections of snails exposed to several miracidia, it was observed that an infected snail may display both developing and encapsulated, i.e., nondeveloping, primary sporocysts in the head–foot tissues (Théron et al. 1997).

Taken together, these results suggest that the infective vs. noninfective phenotype of a parasite could depend on the phenotype of the particular host that it enters. Conversely, susceptibility appears not to be a general characteristic of the host. The susceptible vs. nonsusceptible phenotype of a host is expressed as a function of the parasite phenotype it can harbor. To validate this matching phenotype model (MPM) we proposed (Théron and Coustau 2005), it is necessary to identify the underlying molecular mechanisms. Considering the ideas developed in the introduction of the present chapter, MPM could be based on a system of self/nonself recognition (Mitta, Adema et al. 2012). Some molecules such as the FREPs of *B. glabrata* could be a perfect candidate for the snail molecular determinants in the MPM. One of the chapters of the present book is dedicated to these molecules. Briefly, FREPs are lectin-like hemolymph polypeptides that can precipitate soluble antigens derived from trematodes (Adema et al. 1997); their high level of polymorphism and their somatic diversification were thoroughly studied (Zhang et al. 2004). But as for other invertebrate diversified immune recognition receptors recently characterized, the corresponding antigens of FREPs needed to be identified. The fact that the two species in interaction are two metazoa (and Lophotrochozoa) species could also be an interesting feature. Indeed in such a system, we could expect (1) the necessity to recognize more than classical PAMPs and (2) the emergence of a new antigen-receptor system more sophisticated than the systems involving invertebrate hosts and virus, bacteria, or fungi.

3 Identification of Polymorphic Putative Antigens from *S. mansoni*: The *S. mansoni* Polymorphic Mucins (*Sm*PoMucs)

In order to identify putative *S. mansoni* antigens that could be the parasite counterpart in the MPM, we undertook a comparative molecular approach using two *S. mansoni* laboratory strains: one, totally compatible (C strain) and the other totally incompatible (IC strain) for the same *B. glabrata* strain (Brazilian strain). What is interesting in this experimental system is that newly penetrated parasites from the IC strain are contacted by host hemocytes within 1–2 h post-infection and entirely encapsulated by 8–12 h post-infection. In contrast, newly penetrated miracidia of

the C strain were not encapsulated and primary sporocyst (Sp1) developed normally. These observations suggested that constitutive antigenic differences exist between the two strains, which were then investigated using a comparative proteomic approach (Roger et al. 2008a, b, c). The differentially expressed proteins between strains belonged to two main functional groups: scavengers of reactive oxygen species (ROS) and mucin-like proteins.

The ROS scavengers identified are involved in different oxidative stress scavenging pathways. Among them, a superoxide dismutase (*Sm*SOD) was overrepresented in the C strain. As *B. glabrata* reactive oxygen species (ROS) are the main effector involved in *S. mansoni* sporocysts killing (Hahn et al. 2000), we hypothesized that *Sm*SOD peroxidative function of this enzyme could transform H_2O_2 in $^{\bullet}OH$ that was shown to be less toxic for the sporocysts (Hahn et al. 2001). But among the proteins identified, we considered the newly discovered *S. mansoni* mucins as the most promising candidates because mucins are known to play key roles in the host–parasite interplay. For this reason, we decided to study these molecules more thoroughly (Roger et al. 2008a). We investigated their precursor structures and we analyzed their expression patterns. We showed that these proteins display a mucin-like structure with an N-terminal domain containing a variable number of tandem repeats and a conserved C-terminal domain (Roger et al. 2008a). The proteins are (1) highly glycosylated, (2) only expressed by larval schistosome stages that interact with the snail intermediate host, (3) produced and located in the apical gland of miracidia and sporocysts, and finally (4) highly polymorphic (Roger et al. 2008a). Consequently, we called these molecules *S. mansoni* polymorphic mucins (*Sm*PoMucs). Figure 1 presents the polymorphism of *Sm*PoMucs in the C and IC strains. First, *Sm*PoMuc polymorphism was analyzed at the protein level by 2D gels using a combination of a pH 3–10 nonlinear gradient and a narrow pH 3–6 range to expand the region containing these proteins (Fig. 1a). This approach revealed numerous spots corresponding to *Sm*PoMucs whose position and numbers are different between strains. At the transcript level, *Sm*PoMucs display a large number of molecular variants sharing the same precursor structure (Fig. 1b). They are composed of a signal peptide (22 amino acids in length) followed by a variable number of tandem repeat (VNTR) domain containing repeats of 9 amino acids ($n = 1$ to $n < 55$) and a conserved C-terminal domain. Three different types of repeats were identified, r1, r1′, and r2, that were expressed in both *S. mansoni* strains (Fig. 1b). The variants were classified into three groups considering their similarities in the C-terminal domain. For both strains, groups 1 and 2 share common characteristics: they are always associated with r2 tandem repeats, and the number of repeats is highly variable (Fig. 1b). Major differences between the strains emerge in the third group of molecular variants that are preferentially associated with r1 and r1′ repeats. The same variability in repeat number as the first two groups was observed (Fig. 1b) for the C strain. This is also true for the IC strain, but IC strain harbored an additional subgroup representing about half of the variants containing combinations of the two types of repeats r1 (or r1′) and r2 (Fig. 1b). Finally, we analyzed the expression of the *Sm*PoMucs at the level of

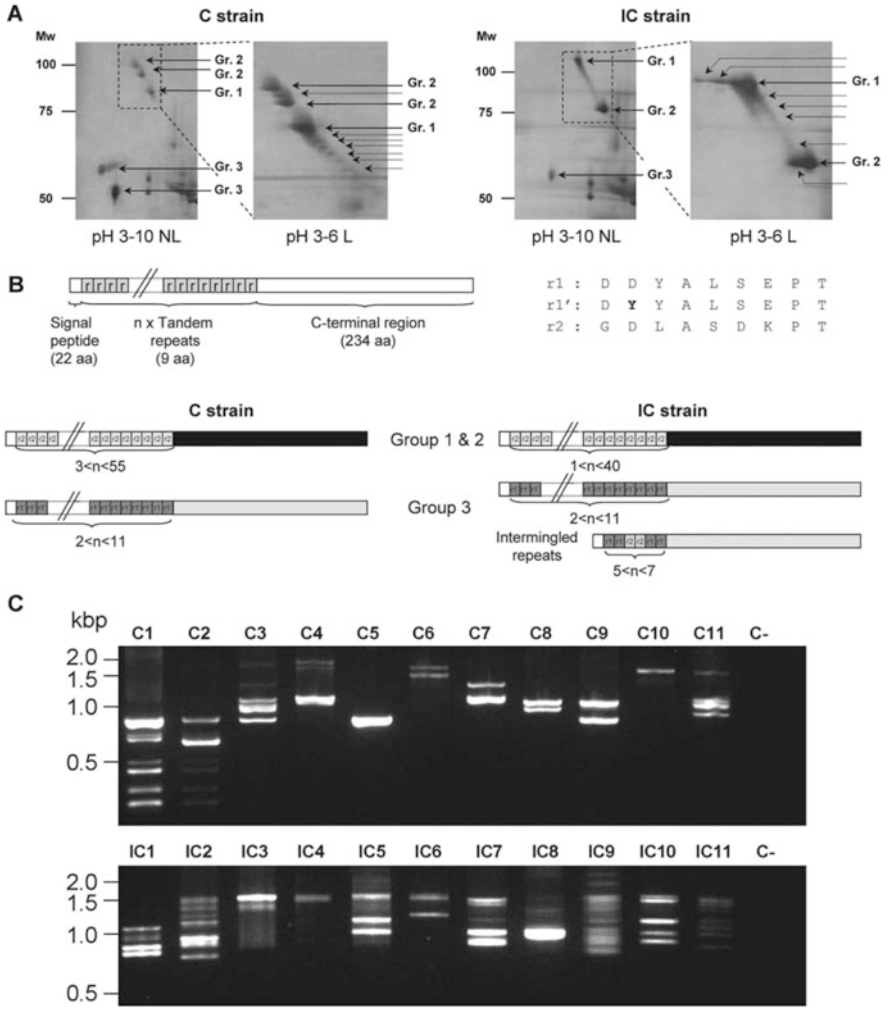


Fig. 1 *Sm*PoMuc polymorphism at the protein and transcript levels. Positional differences between *Sm*PoMucs from compatible (C) and incompatible (IC) strains on silver-stained 2D gels shown with a pH 3–10 nonlinear (NL) gradient or a pH 3–6 linear (L) gradient (**a**). Positions of spots corresponding to *Sm*PoMuc are indicated by arrows. Supplementary spots found in the present study using the pH 3–6 linear gradient are indicated by *dotted arrows*. (**b**) shows the precursor structure and polymorphism of *Sm*PoMuc. Three kinds of repeats were identified in *Sm*PoMuc cDNAs (r1, r1′, and r2). (**c**) Agarose gel separation of RT-PCR amplicons obtained from 11 individual sporocysts (1–11) of both strains (compatible, C, and incompatible, IC). Amplification was performed using consensus primers amplifying the complete coding sequence of all *Sm*PoMuc. C-: negative control of amplification [modified from Roger et al. (2008b)]

individual larvae by nested RT-PCR. It revealed a high degree of polymorphism between individuals within and between strains, all individuals displaying a different *Sm*PoMuc profile (Fig. 1c).

3.1 *The* Sm*PoMuc Genes and the Mechanisms Leading to Their High Level of Diversification*

A detailed analysis of their high level of intra- and interstrain variations showed that *Sm*PoMuc diversification is driven by a complex cascade of mechanisms (Roger et al. 2008b). *Sm*PoMucs are encoded by a multigene family of around 10 members. The typical structure of a *Sm*PoMuc gene is presented in Fig. 2.

Sequencing of these 10 genes in the C and IC strains showed that they fall into four paralogous groups that correspond to the three groups identified at the mRNA level and an additional group whose expression was never detected at the mRNA level. Because of this last feature, the genes of this fourth group were considered as pseudogenes. Analysis of the sequences of *Sm*PoMuc genes revealed that these genes are highly polymorphic and that they undergo frequent recombination events generating new alleles. We found evidence for frequent recombinations between genes of the fourth group and other members of the multigene family. This suggests that pseudogenes can provide an additional pool of genetic variability for the generation of new variants through recombination, gene conversion, or exon shuffling. This type of variation-generating mechanism was observed for the *Trypanosoma brucei* variant membrane surface glycoprotein (Roth et al. 1989), *Anaplasma marginale* membrane surface proteins (Brayton et al. 2002), and MHC (Doxiadis et al. 2006). The numerous insertion/deletion events identified in *Sm*PoMucs, the solo LTR identified in some genes, the truncated genes interrupted by retrotransposition events (contigs identified in the genome assembly), and short tandem repeats flanking some deleted sequences illustrate these frequent reshaping events in *Sm*PoMuc genes. These structural characteristics suggest that retrotransposons could play a central role as mediators of recombination between *Sm*PoMuc genes. The fact that some genes were never expressed and that truncated forms of *Sm*PoMuc were found in the *S. mansoni* genome assembly database suggests that a large proportion of the genes belonging to this multigene family are nonfunctional, exactly as would be expected for multigene families submitted to the birth-and-death model of evolution (Nei and Rooney 2005).

Fig. 2 Schematic representation of a complete *Sm*PoMuc gene. The complete *Sm*PoMuc genes are composed of 15 exons. Exon 2 is included in a genomic repeat that can be repeated several times (a maximum of 20 repeats). These genomic repeats of approximately 1 kb are separated by imperfect polypurine tracts (PPT). Position of a ribozyme between exon 9 and 10 is indicated by an *asterisk*. *Triangles* and *chevrons* indicate complementary sequence positions (12 and 13 nucleotides, respectively) identified in introns of the genomic repeats containing exon 2 [modified from Roger et al. (2008b)]

We showed that *Sm*PoMuc genes occupy four loci in the genome of *S. mansoni*, three on the third chromosome, and one on the fourth chromosome (Roger et al. 2008b). Two of these locations were analyzed in detail using bacterial artificial chromosomes (BACs) and a combination of PCR, sequencing, and Southern blot analysis. These analyses showed that these two locations correspond to *Sm*PoMuc group 2 (gr.2) and group 3 (gr.3), respectively. These genes are organized in clusters in these two distinct genomic regions. The *Sm*PoMuc gr.2 cluster is composed of at least two genes, one complete containing all exons (1–15, Fig. 2) and approximately 20 repeats of exon 2 and a truncated gene with no exon 2 repeats. The *Sm*PoMuc gr.3 cluster is composed of at least six tandemly organized genes containing 1–15 exon 2 repeats. Furthermore, our results revealed that *Sm*PoMuc 2 and 3 clusters are associated with a specific exon 2, the first containing only r2 exons and the second only r1 exons. Nevertheless, intermingled r1 and r2 repeats were found in transcripts of *Sm*PoMuc gr.3 variants of C and IC strains. This suggests that ectopic recombination could occur to generate this genomic level of polymorphism. Particular structural elements found in gene structure could explain these frequent recombination events. Genomic repeats containing exon 2 and flanking intronic sequences are highly conserved within and between all genes of the *Sm*PoMuc family. The level of identity (>93 % at the nucleotide level) shows that these genomic repeats do not evolve independently of each other. The molecular process that leads to homogenization of DNA sequences of a given repetitive family is called concerted evolution and occurs, for example, in ribosomal genes and in certain protein-coding multigene families, such as those encoding histones or ubiquitins (see Liao (1999), for review). This phenomenon was also previously described for several genes (1) encoding proteins containing tandemly repeated domains and (2) displaying genomic repeats like those found *Sm*PoMuc genes.

This is the case for the repetitive part of the single-copy gene encoding the Coccidioides spherule outer wall glycoprotein (SOWgp). This protein contributes to the virulence of *Coccidioides* spp. both by functioning as an adhesion molecule and by modulating the host's immune response (Hung et al. 2000, 2002). In this SOWgp gene, genomic repeats corresponding to repeated exons and associated introns are also nearly identical and evolve by concerted evolution. This type of evolutionary process can be driven by directional and/or stochastic processes. The two mechanisms that are the principal explanations of concerted evolution in nuclear DNA are gene conversion and unequal crossing-over (Nei and Rooney 2005). For unequal crossing-over, increases and decreases in repeat number lead to turnover among repeats and, in principle, stochastic fixation of a single repeat type. According to this model, unequal crossing-over commonly occurs in central regions of the array where repeats can mispair, and unique sequences flanking the repetitive array inhibit exchanges in the edge repeats, i.e., the repeats located at the termini. The involvement of unequal crossing-over was shown for the gene encoding SOWgp (Johannesson et al. 2005), and support for this view comes from both the higher conservation of repeats in the center of the repetitive array and the polymorphism in repeat number in SOWgp. The phenomenon of unequal crossing-over probably occurs in members of the *Sm*PoMuc family because our Southern blot

results indicate a difference in repeat number between individuals, differences that occur when unequal crossing-over takes place. Nevertheless, unequal crossing-over cannot explain homogenization of genomic repeats containing exon 2 between genes situated on different loci and chromosomes. In addition, we did not find a gradient of conservation between central and edge repeats. We therefore favor the alternative hypothesis of gene conversion (nonreciprocal transfer of information).

The molecular mechanism of gene conversion in multigene families is not well understood; nevertheless, several findings indicate that cis-acting sequences can influence this phenomenon. One example is provided by the two early chorionic gene families, ErA and ErB, of the silk moth *Bombyx mori* (Hibner et al. 1991) which are in close proximity on the same chromosome. The genes of the ErA family exhibit 96 % sequence identity, whereas those in the ErB have only 63 % sequence identity. Sequence analysis suggested that microsatellite-like simple repeats present in the ErA family, but not in the ErB family, accounts for the difference in homogenization, because simple sequence repeats can be the sites for initiation of gene conversion (Hibner et al. 1991). Another well-documented example is given by the microsatellite sequences in the human RNU2 locus that were proposed to play a role in concerted evolution (Liao and Weiner 1995). These observations support the hypothesis that the microsatellites that separate the exon 2 genomic repeats in the *Sm*PoMuc genes are involved in the gene conversion mechanism leading to the concerted evolution of these repeats. Several hypotheses for the mechanism of gene conversion induced by microsatellites have been evoked. These microsatellites composed of purines are polypurine tracts that can adopt a triple-helix conformation called H-DNA (Bacolla and Wells 2004). A major role for this kind of conformation (non-B DNA conformation) in chromosomal rearrangements was proposed. Hotspots of rearrangements occurred invariably at nucleotides abutting or within motifs capable of adopting non-B conformations leading to single- or double-strand breaks (Bacolla et al. 2004; Wang and Vasquez 2006). The DNA break repair mechanism could involve invasive DNA replication leading to gene conversion as shown in yeast (Paques and Haber 1999; Szostak et al. 1983).

These different mechanisms could occur in *Sm*PoMuc genes and explain the similarities observed between the genomic repeats in all members of the multigene family. Nevertheless, this gene conversion phenomenon is restricted to the intronic sequence of these genomic repeats as we have shown that exons are different between clusters. We therefore conclude that conservation of exon differences is due to selective pressures. Interestingly, the same molecular architecture and the same type of evolutionary processes were described for the abovementioned modular spider silk protein genes. In these genes, repeated structures in the proteins correspond to genomic tandem repeats composed of exons and introns. Once more, in this case, genomic repeats are subject to concerted evolution, and intron sequences are more homogenized than are the exons that evolve under purifying selection (Hayashi and Lewis 2000). In *Sm*PoMuc genes, the combination of concerted evolution acting on all repeats and exon difference conservation between clusters (by purifying selection) allows combinatory events that we observed at the genomic level (r1 and r2 exons in the same gene) and in cDNA (r1 and r2 repeats in

the same variant) in both strains. The reason is probably ectopic recombination and exon exchange between clusters.

In addition to the structure of the genes that favor generation of polymorphism on a genomic level, several processes generating further levels of polymorphism occur during the expression process. First, genes are transcribed in an individual-specific manner (Roger et al. 2008b). Some individuals express several genes and/or alleles for one group of *Sm*PoMuc and others for two groups that may be different between individuals. This observation raises the question of differential transcriptional regulation of the genes belonging to this multigene family. We demonstrated recently that epigenetic mechanisms are involved in this differential regulation (Perrin et al. 2012). Epigenetic control of transcription can be based on DNA methylation and covalent modifications of histones and other DNA-associated proteins. We showed that DNA methylation is absent from the promoter regions of *Sm*PoMucs and focused therefore on histone modifications. Using a combination of chromatin immunoprecipitation (ChIP) and DNA sequencing, we demonstrated that in the absence of genetic variation, chromatin structure differences are strongly correlated with transcription differences between the C and IC strains (Perrin et al. 2012). In addition, treatment of larvae with an epimutagene, the histone deacetylase inhibitor trichostatin A (TSA), influences the repertoire of *Sm*PoMuc transcripts (Cosseau et al. 2010).

Second, we found evidence for various posttranscriptional regulation events (Roger et al. 2008b). Exhaustive analysis of *Sm*PoMuc cDNAs reveals numerous alternative splicing and aberrant splicing events in the coding region in the 234 residue C-terminal region of the precursor (Fig. 1). Alternative splicing events do not change the ORF and lead to shorter proteins. Aberrant splicing also appears frequently and produces a nonsense codon immediately downstream of the splice sites, leading to truncated proteins.

Third, we suspect that trans-splicing events could occur during the expression level to generate large-size *Sm*PoMuc transcripts. The analysis by Southern blot of genomic DNA from both strains and BACs revealed a maximum of 20 genomic repeats containing the second exon (encoding the repeats in the proteins). Nevertheless, 50 % of the *Sm*PoMuc cDNA variants possess more than 20 repetitive units (25–100 repeats). We decided to investigate the putative involvement of trans-splicing and exon repetition mechanisms. Trans-splicing was reported for *S. mansoni* (Cheng et al. 2006; Davis et al. 1995; Rajkovic et al. 1990), but the absence of a spliced leader sequence in *Sm*PoMuc transcripts suggests that this mechanism cannot be involved in the observed phenomenon. Exon repetition was first identified for the rat carnitine octanoyltransferase gene for which two copies of exon 2 were positioned adjacent to one another in some mRNAs, while the genomic sequence contained only a single copy (Caudevilla et al. 1998). This mechanism has been further studied (Frantz et al. 1999; Rigatti et al. 2004) and the intervention of complementary intron sequences has been hypothesized (Dixon et al. 2007). Since we detected two complementary sequences of 13 and 12 nucleotides (Fig. 2), respectively, in intronic sequences flanking exon 2, this hypothesis can be proposed for *Sm*PoMuc genes. Finally, the presence of hammerhead ribozymes in all

*Sm*PoMuc genes is intriguing. *S. mansoni* hammerhead ribozymes were extensively studied and shown to catalyze cleavage (De la Pena et al. 2003) and ligation (Canny et al. 2007) of transcripts in vitro. Their in vivo function is unknown, but we can hypothesize that *S. mansoni* hammerhead ribozymes could play a role in *Sm*PoMuc transcript processing and posttranslational mucin regulation (Martick et al. 2008).

All the data we present here for the *Sm*PoMuc multigene family show that gene structure, genomic organization, recombination events, and different regulation mechanisms during their expression allow the generation of a remarkably high degree of polymorphism from a limited set of genes. This characteristic is unique for this model compared to the expression of polymorphic molecular variants in other parasites. Indeed, in all previously described cases, the molecular variants are synthesized from a large set of genes belonging to a multigene family. An example is the case of *Trypanosoma cruzi* surface mucins (see Buscaglia et al. (2006) for review) that contribute to parasite protection and to the establishment of a persistent infection. The multigene family encoding these proteins comprises 850 genes covering 1 % of the parasite genome. Other relevant gene families include the vsg or the var family responsible for antigenic variation of *Trypanosoma brucei* [see Navarro et al. (2007), Taylor and Rudenko (2006)] or *P. falciparum* (Kyes et al. 2007; Ralph and Scherf 2005). *T. brucei* has 1000 vsg genes and pseudogenes, and the genome project of *P. falciparum* has identified 59 intact var genes.

As putative antigens and molecular determinants of the compatibility polymorphism, we decided to investigate the impact of all these diversification levels we evidenced on the glycosylation status of *Sm*PoMucs. A combination of in silico analysis of the predicted glycosylation patterns of *Sm*PoMuc sequence variants and of chemical deglycosylation experiments to determine overall *Sm*PoMuc glycosylation levels indicated that sequence polymorphism is directly linked to glycosylation status (Roger et al. 2008b). Indeed, we found quantitative and qualitative differences between carbohydrate patterns of *Sm*PoMucs between strains. *Sm*PoMucs of the IC strain were more heavily glycosylated than *Sm*PoMucs recorded from *S. mansoni* C strain (Roger et al. 2008b). The superimposition of these different levels of variability provides an extraordinary level of polymorphism for *Sm*PoMucs that is the result of a "controlled chaos."

4 *Sm*PoMucs and FREPs: Two Interacting Molecular Repertoires

The structure, expression pattern, glycosylation status, and inter- and intra-strain polymorphism made the *Sm*PoMucs promising candidates in the MPM we propose. Another of our studies strengthened this hypothesis (Mone et al. 2010). Here, a combination of interactome and co-immunoprecipitation (CoIP) experiments was developed.

The interactome experiment aimed on identifying all plasma proteins from the snail that interact with the parasite. This approach revealed several proteins putatively implicated in recognition and presumably in immunity. First, several host lectins and parasite glycoproteins were identified. As expected, among the lectins, FREPs were identified as well as another *B. glabrata* lectin. This latter molecule displays similarities with a secreted galactose-binding lectin characterized in another gastropod, *Helix pomatia* (Perez-Sanchez et al. 2006). Considering the parasite molecular determinants that could be recognized by these lectins, several glycosylated proteins have been identified: *Sm*PoMucs and two other glycoproteins, the 23 kDa integral membrane protein (Sm23) (or tetraspanin) and the glycoprotein K5. Taken together, these results suggest that the recognition process between *S. mansoni* and *B. glabrata* could be multifactorial involving different immune recognition receptors from the host and different carbohydrate components and/or glycoproteins from the parasite. This interactome approach also revealed an interesting new effector from the snail found in the precipitates: a putative cytolytic protein related to a β-pore-forming toxin family whose amino acid sequence displays similarities to aerolysins. In a recent work, we report the molecular cloning and functional characterization of this *B. glabrata* aerolysin (Biomphalysin) which is highly toxic toward *Schistosoma mansoni* sporocysts (Galinier et al. 2013).

The complementary CoIP experiments we developed (Mone et al. 2010) were dedicated to the identification of the suspected interaction between FREPs and *Sm*PoMucs. Antibodies raised against *Sm*PoMucs were used and we confirmed that FREPs and *Sm*PoMucs are together in a molecular complex. We also showed that this immune complex contains a third partner, a thioester-containing protein (TEP) from *B. glabrata* (*Bg*TEP). The presence of TEP in the complex is exciting as some molecules of this family were recently shown to play key roles in other invertebrate/pathogen interactions, especially in insects. Indeed, TEP1 was shown to play a crucial role in the phagocytosis of bacteria and killing of *Plasmodium* parasites in the mosquito *Anopheles gambiae*. TEP1 from the mosquito is secreted by hemocytes and cleaved in hemolymph into an active form [called mature TEP1 or TEP1-cut (Fraiture et al. 2009)]. The C-terminal part of TEP1 binds to bacteria or ookinete surfaces through a thioester bound. The involvement of this complement-like molecule in the antiparasitic defense of mosquitoes was discussed (Volohonsky et al. 2010). Precursor and phylogenetic analysis of *Bg*TEP suggests that it shares the features of invertebrate TEPs that are known to be involved in antiparasitic defense and microbe phagocytosis (Blandin and Levashina 2004, 2007; Blandin et al. 2008; Stroschein-Stevenson et al. 2006). In addition, our LC-MS/MS experiments led to the identification of peptides that are all located in the C-terminal part of *Bg*TEP suggesting a thioester-dependent association with the two other partners of the complex. Therefore, the *Bg*TEP found in the complex is activated and could play a role in opsonization processes as described for the members of this family. This hypothesis is clearly supported by the alpha2-macroglobulin receptor-binding domain (region 1343–1427) found in the C-terminal part of *Bg*TEP precursor (Mone et al. 2010). This domain is known to be involved in the interaction with macrophage- and phagocyte-specific receptors (van Lookeren Campagne

et al. 2007), and it could participate in hemocyte recruitment and capsule formation around the parasite in snail tissues when the interaction is incompatible.

5 Conclusion

In conclusion, we showed that a specific set of highly variable FREPs from *B. glabrata* form complexes with similarly highly polymorphic and individually variable mucins (the *Sm*PoMucs) from its trematode parasite *S. mansoni*. Among the different molecules identified by the combination of approaches reviewed here, it appears that FREPs and *Sm*PoMucs display the appropriate level of polymorphism to explain the compatibility polymorphism in *B. glabrata*/*S. mansoni* natural populations. The conceptual scheme of matching phenotypes that was deduced from population level studies supports the view that two repertoires of highly polymorphic molecules from the host and the parasite are deployed against each other during the early stages of the process of infection, and this interaction is a key component dictating the success or failure of the infection. Several recent studies validate partially this molecular scenario. The knockdown of FREPs by RNA interference was shown to reduce by 20–30 % the resistance of *B. glabrata* to trematode infections (Hanington et al. 2011, 2012). Experiments to knockdown *Sm*PoMucs are currently developed. In addition, we showed that *Sm*PoMuc expression is controlled epigenetically through histone modifications (Perrin et al. 2012) and that the diversity of *Sm*PoMuc expression can be increased by influencing the chromatin status of the *S. mansoni* sporocyst by trichostatin A (TSA) treatment (Cosseau et al. 2010). For this reason, we tested recently the effect of TSA on compatibility, and we showed that the treatment increases (1) the diversity of *Sm*PoMuc expression and (2) the compatibility of some parasite strains toward a specific snail strain (Fneich et al., 2015). These results clearly support the view that FREPs (and *Sm*PoMucs) could be key factors of the compatibility polymorphism, but all the results presented in this chapter taken together suggest that other molecular partners are probably involved. Indeed, we and others identified several other molecular determinants implicated in recognition (glycoprotein and lectins) in effector mechanisms (ROS, Biomphalysin, etc.) that could participate to the compatibility puzzle between *B. glabrata* and *S. mansoni*.

6 Perspectives

In order to validate our current scenario, the full composition of the immune complexes and the nature of the interacting domains of the host and parasite proteins remain to be fully characterized. This characterization is crucial to understand which combinations of molecular determinants define the compatibility or incompatibility status of the interaction. This will be particularly difficult with respect to the identification of the glycan part of *Sm*PoMucs interacting with

other proteins of the complex. Nevertheless, two complementary approaches can be envisioned. The first involves co-immunoprecipitation experiments with full-length or truncated recombinant FREPs incubated with sporocyst extracts (containing *Sm*PoMucs) and *B. glabrata* plasma to define precisely the FREP domains involved in the interaction. For the second approach, the *Sm*PoMuc allelic variants available in compatible and incompatible combinations provide an excellent comparative framework to dissect the nature and composition of complex formation. Both approaches could allow identifying FREP/*Sm*PoMuc combinations leading to compatibility/incompatibility.

In addition and as previously mentioned, it is important to note that *Sm*PoMuc and FREPs are probably only one piece of the puzzle, and other molecules may well be involved in the compatibility polymorphism phenomenon. As the mechanism underlying compatibility polymorphism between *S. mansoni* and *B. glabrata* is multifactorial, we develop currently a complementary genetic approach by X-QTL mapping in collaboration with T. Anderson (San Antonio, USA). We hope that this approach will help us to identify all the factors and their respective weight in the MPM we propose.

References

Adema CM, Hertel LA, Miller RD, Loker ES (1997) A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. Proc Natl Acad Sci USA 94(16):8691–8696

Anderson RM, May RM (1979) Prevalence of schistosome infections within molluscan populations: observed patterns and theoretical predictions. Parasitology 79(1):63–94

Bacolla A, Wells RD (2004) Non-B DNA conformations, genomic rearrangements, and human disease. J Biol Chem 279(46):47411–47414

Bacolla A, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeysinghe SS, O'Connell CD, Cooper DN, Wells RD (2004) Breakpoints of gross deletions coincide with non-B DNA conformations. Proc Natl Acad Sci USA 101(39):14162–14167

Basch PF (1975) An interpretation of snail-trematode infection rates: specificity based on concordance of compatible phenotypes. Int J Parasitol 5(4):449–452

Blandin S, Levashina EA (2004) Thioester-containing proteins and insect immunity. Mol Immunol 40(12):903–908

Blandin SA, Levashina EA (2007) Phagocytosis in mosquito immune responses. Immunol Rev 219:8–16

Blandin SA, Marois E, Levashina EA (2008) Antimalarial responses in *Anopheles gambiae*: from a complement-like protein to a complement-like pathway. Cell Host Microbe 3(6):364–374

Brayton KA, Palmer GH, Lundgren A, Yi J, Barbet AF (2002) Antigenic variation of *Anaplasma marginale* msp2 occurs by combinatorial gene conversion. Mol Microbiol 43(5):1151–1159

Buckley KM, Rast JP (2012) Dynamic evolution of toll-like receptor multigene families in echinoderms. Front Immunol 3:136

Buscaglia CA, Campo VA, Frasch AC, Di Noia JM (2006) *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. Nat Rev Microbiol 4(3):229–236

Cannon JP, Haire RN, Schnitker N, Mueller MG, Litman GW (2004) Individual protochordates have unique immune-type receptor repertoires. Curr Biol 14(12):R465–466

Canny MD, Jucker FM, Pardi A (2007) Efficient ligation of the *Schistosoma* hammerhead ribozyme. Biochemistry 46(12):3826–3834

Caudevilla C, Serra D, Miliar A, Codony C, Asins G, Bach M, Hegardt FG (1998) Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. Proc Natl Acad Sci USA 95(21):12185–12190

Cheng G, Cohen L, Ndegwa D, Davis RE (2006) The flatworm spliced leader 3'-terminal AUG as a translation initiator methionine. J Biol Chem 281(2):733–743

Combes C (2000) Selective pressure in host-parasite systems. J Soc Biol 194(1):19–23

Cosseau C, Azzi A, Rognon A, Boissier J, Gourbiere S, Roger E, Mitta G, Grunau C (2010) Epigenetic and phenotypic variability in populations of *Schistosoma mansoni* – a possible kick-off for adaptive host/parasite evolution. Oikos 119(4):669–678

Davis RE, Hardwick C, Tavernier P, Hodgson S, Singh H (1995) RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite *Schistosoma mansoni*. J Biol Chem 270(37):21813–21819

De la Pena M, Gago S, Flores R (2003) Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. EMBO J 22(20):5561–5570

Di Noia JM, Neuberger MS (2007) Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem 76:1–22

Dixon RJ, Eperon IC, Samani NJ (2007) Complementary intron sequence motifs associated with human exon repetition: a role for intragenic, inter-transcript interactions in gene expression. Bioinformatics 23(2):150–155

Dong Y, Taylor HE, Dimopoulos G (2006) AgDscam, a hypervariable immunoglobulin domain-containing receptor of the Anopheles gambiae innate immune system. PLoS Biol 4(7), e229

Dong Y, Cirimotich CM, Pike A, Chandra R, Dimopoulos G (2012) *Anopheles* NF-kappaB-regulated splicing factors direct pathogen-specific repertoires of the hypervariable pattern recognition receptor AgDscam. Cell Host Microbe 12(4):521–530

Doxiadis GG, van der Wiel MK, Brok HP, de Groot NG, Otting N, t Hart BA, van Rood JJ, Bontrop RE (2006) Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a New World primate species. Proc Natl Acad Sci USA 103(15):5864–5868

Ebert D (2008) Host-parasite coevolution: Insights from the *Daphnia*-parasite model system. Curr Opin Microbiol 11:290–301

Finlay BB, McFadden G (2006) Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. Cell 124(4):767–782

Fraiture M, Baxter RH, Steinert S, Chelliah Y, Frolet C, Quispe-Tintaya W, Hoffmann JA, Blandin SA, Levashina EA (2009) Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of *Plasmodium*. Cell Host Microbe 5(3):273–284

Frantz SA, Thiara AS, Lodwick D, Ng LL, Eperon IC, Samani NJ (1999) Exon repetition in mRNA. Proc Natl Acad Sci USA 96(10):5400–5405

Fneich S, Théron A, Cosseau C, Buard J, Duval D, Arancibia N, Boissier J, Mitta G, Grunau C (2015) Epigenetic origin of adaptive phenotypic variants in the blood-fluke *Schistosoma mansoni* (in press)

Galinier R, Portela J, Mone Y, Allienne JF, Henri H, Delbecq S, Mitta G, Gourbal B, Duval D (2013) Biomphalysin, a new beta pore-forming toxin involved in *Biomphalaria glabrata* immune defense against Schistosoma mansoni. PLoS Pathog 9(3), e1003216

Ghosh J, Lun CM, Majeske AJ, Sacchi S, Schrankel CS, Smith LC (2011) Invertebrate immune diversity. Dev Comp Immunol 35(9):959–974

Gryseels B, Polman K, Clerinx J, Kestens L (2006) Human schistosomiasis. Lancet 368(9541): 1106–1118

Gupta S, Maiden MC, Feavers IM, Nee S, May RM, Anderson RM (1996) The maintenance of strain structure in populations of recombining infectious agents. Nat Med 2(4):437–442

Hahn UK, Bender RC, Bayne CJ (2000) Production of reactive oxygen species by hemocytes of *Biomphalaria glabrata*: carbohydrate-specific stimulation. Dev Comp Immunol 24(6–7): 531–541

Hahn UK, Bender RC, Bayne CJ (2001) Killing of *Schistosoma mansoni* sporocysts by hemocytes from resistant *Biomphalaria glabrata*: role of reactive oxygen species. J Parasitol 87(2): 292–299

Hanington PC, Forys MA, Dragoo JW, Zhang SM, Adema CM, Loker ES (2011) Role for a somatically diversified lectin in resistance of an invertebrate to parasite infection. Proc Natl Acad Sci USA 107(49):21087–21092

Hanington PC, Forys MA, Loker ES (2012) A somatically diversified defense factor, FREP3, is a determinant of snail resistance to schistosome infection. PLoS Negl Trop Dis 6(3), e1591

Hayashi CY, Lewis RV (2000) Molecular architecture and evolution of a modular spider silk protein gene. Science 287(5457):1477–1479

Hibner BL, Burke WD, Eickbush TH (1991) Sequence identity in an early chorion multigene family is the result of localized gene conversion. Genetics 128(3):595–606

Hung CY, Ampel NM, Christian L, Seshan KR, Cole GT (2000) A major cell surface antigen of *Coccidioides immitis* which elicits both humoral and cellular immune responses. Infect Immun 68(2):584–593

Hung CY, Yu JJ, Seshan KR, Reichard U, Cole GT (2002) A parasitic phase-specific adhesin of *Coccidioides immitis* contributes to the virulence of this respiratory Fungal pathogen. Infect Immun 70(7):3443–3456

Johannesson H, Townsend JP, Hung CY, Cole GT, Taylor JW (2005) Concerted evolution in the repeats of an immunomodulating cell surface protein, SOWgp, of the human pathogenic fungi *Coccidioides immitis* and *C. posadasii*. Genetics 171(1):109–117

Kyes SA, Kraemer SM, Smith JD (2007) Antigenic variation in *Plasmodium falciparum*: gene organization and regulation of the var multigene family. Eukaryot Cell 6(9):1511–1520

Liao D (1999) Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64(1):24–30

Liao D, Weiner AM (1995) Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the RNU2 locus) does not prevent rapid diversification of the (CT)n. (GA)n microsatellite embedded within the U2 repeat unit. Genomics 30(3):583–593

Martick M, Horan LH, Noller HF, Scott WG (2008) A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. Nature 454(7206):899–902

Medzhitov R, Janeway CA Jr (1997) Innate immunity: the virtues of a nonclonal system of recognition. Cell 91(3):295–298

Mitta G, Adema CM, Gourbal B, Loker ES, Theron A (2012) Compatibility polymorphism in snail/schistosome interactions: from field to theory to molecular mechanisms. Dev Comp Immunol 37(1):1–8. doi:10.1016/j.dci.2011.09.002, Epub 2011 Sept 16

Mone Y, Gourbal B, Duval D, Du Pasquier L, Kieffer-Jaquinod S, Mitta G (2010) A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. PLoS Negl Trop Dis 4(9)

Navarro M, Penate X, Landeira D (2007) Nuclear architecture underlying gene expression in *Trypanosoma brucei*. Trends Microbiol 15(6):263–270

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152

Pancer Z (2000) Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. Proc Natl Acad Sci USA 97(24):13156–13161

Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD (2004) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 430(6996):174–180

Paques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev 63(2):349–404

Perez-Sanchez R, Ramajo-Hernandez A, Ramajo-Martin V, Oleaga A (2006) Proteomic analysis of the tegument and excretory-secretory products of adult *Schistosoma bovis* worms. Proteomics 6(Suppl 1):S226–236

Perrin C, Lepesant JM, Roger E, Duval D, Fneich S, Thuillier V, Alliene JF, Mitta G, Grunau C, Cosseau C (2012) *Schistosoma mansoni* mucin gene (SmPoMuc) expression: epigenetic control to shape adaptation to a new host. PLoS Pathog 9(8), e1003571

Pianka E (1970) R-selection and K-selection. Am Nat 104(940):592–597

Pradeu T, Cooper EL (2012) The danger theory: 20 years later. Front Immunol 3:287

Rajkovic A, Davis RE, Simonsen JN, Rottman FM (1990) A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. Proc Natl Acad Sci USA 87(22): 8879–8883

Ralph SA, Scherf A (2005) The epigenetic control of antigenic variation in Plasmodium falciparum. Curr Opin Microbiol 8(4):434–440

Rigatti R, Jia JH, Samani NJ, Eperon IC (2004) Exon repetition: a major pathway for processing mRNA of some genes is allele-specific. Nucleic Acids Res 32(2):441–446

Roger E, Gourbal B, Grunau C, Pierce RJ, Galinier R, Mitta G (2008a) Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. Mol Biochem Parasitol 157(2):217–227

Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, Galinier R, Emans R, Cesari IM, Cosseau C, Mitta G (2008b) Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). PLoS Negl Trop Dis 2(11), e330

Roger E, Mitta G, Mone Y, Bouchut A, Rognon A, Grunau C, Boissier J, Théron A, Gourbal BE (2008c) Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: new candidates identified by a global comparative proteomics approach. Mol Biochem Parasitol 157(2):205–216

Roth C, Bringaud F, Layden RE, Baltz T, Eisen H (1989) Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. Proc Natl Acad Sci USA 86(23):9375–9379

Sire C, Durand P, Pointier JP, Théron A (1999) Genetic diversity and recruitment pattern of *Schistosoma mansoni* in a *Biomphalaria glabrata* snail population: a field study using random-amplified polymorphic DNA markers. J Parasitol 85(3):436–441

Stroschein-Stevenson SL, Foley E, O'Farrell PH, Johnson AD (2006) Identification of *Drosophila* gene products required for phagocytosis of Candida albicans. PLoS Biol 4(1), e4

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33(1):25–35

Taylor JE, Rudenko G (2006) Switching trypanosome coats: what's in the wardrobe? Trends Genet 22(11):614–620

Théron A, Coustau C (2005) Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? J Helminthol 79(3):187–191

Théron A, Pages JR, Rognon A (1997) *Schistosoma mansoni:* distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. Exp Parasitol 85(1):1–9

Théron A, Coustau C, Rognon A, Gourbiere S, Blouin MS (2008) Effects of laboratory culture on compatibility between snails and schistosomes. Parasitology 135(10):1179–1188

Théron A, Rognon A, Gourbal B, Mitta G (2014) Multi-parasite host susceptibility and multi-host parasite infectivity: a new approach of the *Biomphalaria glabrata*/*Schistosoma mansoni* compatibility polymorphism. Infect Genet Evol 26:80–88

Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302(5909):575–581

van Lookeren Campagne M, Wiesmann C, Brown EJ (2007) Macrophage complement receptors and pathogen clearance. Cell Microbiol 9(9):2095–2102

Van Valen L (1974) Molecular evolution as predicted by natural selection. J Mol Evol 3(2):89–101

Volohonsky G, Steinert S, Levashina EA (2010) Focusing on complement in the antiparasitic defense of mosquitoes. Trends Parasitol 26(1):1–3

Wang G, Vasquez KM (2006) Non-B DNA structure-induced genetic instability. Mutat Res 598(1–2):103–119

Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI, Schmucker D (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 309(5742):1874–1878

Zhang SM, Adema CM, Kepler TB, Loker ES (2004) Diversification of Ig superfamily genes in an invertebrate. Science 305(5681):251–254

Part II
Hosts: Anticipatory and Adapting Mechanisms

Fibrinogen-Related Proteins (FREPs) in Mollusks

Coen M. Adema

Abstract Anti-parasite responses of the snail *Biomphalaria glabrata* involve antigen-reactive plasma lectins termed fibrinogen-related proteins (FREPs) comprising a C-terminal fibrinogen (FBG) domain and one or two upstream immunoglobulin domains. FREPs are highly polymorphic; they derive from several gene families with multiple loci and alleles that are diversified by exon loss, alternative splicing, and random somatic mutation (gene conversion and point mutations). Individual *B. glabrata* snails have dynamically distinct FREP sequence repertoires. The immune relevance of *B. glabrata* FREPs is indicated by FREP binding to polymorphic antigens of (snail-specific) digenean parasites and altered resistance of *B. glabrata* to digeneans following RNAi knockdown of FREPs. The compatibility polymorphism hypothesis proposes that FREP mutation increases the range of germline-encoded immune recognition in *B. glabrata* to counter antigenically-varied parasites. Somatic mutation may result from sequence exchange among tandemly arranged FREP genes in the genome, and analysis of sequence variants also suggests involvement of cytidine deaminase-like activity or epigenetic regulation. Without current indications of selection or retention of effective sequence variants toward immunological memory, FREP diversification is thought to afford *B. glabrata* immunity that is anticipatory but not adaptive. More remains to be learned about this system; other mollusks elaborate diversified lectins consisting of single FBG domains, and bona fide FREPs were reported from additional gastropod species, but these may not be diversified. Future comparative immunological studies and gene discovery driven by next-generation sequencing will further clarify taxonomic distribution of FREP diversification and the underlying mutator mechanisms as a component of immune function in mollusks.

C.M. Adema (✉)
Biology Department, Center for Evolutionary and Theoretical Immunology, University of New Mexico, Albuquerque, NM 87131, USA
e-mail: coenadem@unm.edu

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_5

1 Lophotrochozoan Immunity

Historically, an emphasis on the study of species with medical or veterinary importance initially yielded a biased view of animal immunity, as revealed from jawed vertebrates, mostly mammals. These organisms provided only limited representation of the deuterostome clade of the animal kingdom, to the exception of invertebrate deuterostomes. Moreover, for some 600 million years, deuterostomes have evolved separately from the protostomes, a clade of invertebrate animals that is subdivided between ecdysozoa (including arthropods and nematodes) and the lophotrochozoa (e.g. annelids and mollusks). Even basal to deuterostomes and protostomes are the prebilateria such as sponges and cnidarians (Erwin et al. 2011). Despite sharing some biological properties due to a common origin, these animal lineages have all evolved independently and differently. Not surprisingly, initial comparative immunological studies of invertebrates revealed basic innate-type immunity such as it also occurs in vertebrates but none of the components that characterize the anticipatory and adaptive immunity of jawed vertebrates such as antibodies, specific antigen receptors generated by rearranging genes (Marchalonis and Schluter 1990). This led to the view that invertebrate immunity was relatively unsophisticated. Invertebrates were thought to be restricted to innate-type responses where germline-encoded pattern recognition factors (PRRs) mediated broad recognition of pathogen-associated molecular patterns (PAMPs), with lectins (proteins that bind specific carbohydrates ligands) interacting with e.g. LPS to recognize Gram (-) bacteria (Medzhitov and Janeway 1997). Limited immune capacities, however, seemed incongruent with the evolutionary success of invertebrate animals. Rather than searching for homologs of vertebrate immune components, comparative immunologists—increasingly profiting from access to molecular techniques—began to explore invertebrates for alternative, other effectors of immune function (Klein 1989). Study of immunity of ecdysozoan protostomes benefited considerably from availability of *Drosophila melanogaster* as model organism, providing novel insights into comparative immunity (Hetru and Hoffmann 2009). No such well-established model system was available for study of lophotrochozoan immunity. Nevertheless, a long-standing research tradition into immunological determinants that rule the snail-mediated transmission of flatworm parasites such as the human pathogen *Schistosoma mansoni* (Digenea, Platyhelminthes) provided the aquatic fresh water snail *B. glabrata* (Gastropoda, Mollusca; see Bayne 2009) as an important organism for study.

2 Snail Lectins Regulate Immune Activities

2.1 Snail Immunity Regulation by Soluble Factors

Digenea (trematodes) are parasitic flatworms of vertebrates. While the biology of these parasites is highly diverse (over 25,000 species estimated), the life cycles of digeneans share a common feature in the use of snails (gastropod mollusks) as obligate intermediate hosts for transmission to a vertebrate final host (Esch et al. 2002). Among several digeneans that can be transmitted by the freshwater snail *Biomphalaria glabrata* are *Echinostoma paraensei* (and related parasites of rodents) and *Schistosoma mansoni*, a significant human pathogen that causes intestinal schistosomiasis, a debilitating infectious disease (Bruun and Aagaard-Hansen 2008). Especially the latter parasite–snail host interaction has been studied extensively toward better understanding and control of transmission of human disease. Through field and laboratory research it was recognized that *B. glabrata* was not always a permissive host that supports successful development and transmission of infecting schistosome parasites. In fact, only a low percentage of field collected snails harbors productive schistosome infections. Lie and coworkers observed that a small proportion of lab-maintained *B. glabrata* snails was naturally resistant at first exposure to a particular digenean species, and *B. glabrata* can develop transient levels of acquired resistance following previous exposure to incompatible parasites with modest cross protection to related digenean species (Lie et al. 1975a, b). Genetic factors determine host–parasite suitability; several laboratory strains of *B. glabrata* have been generated that are resistant or susceptible to one or even several schistosome isolates or yet other digenean parasites (Richards et al. 1992; Langand and Morand 1998). Compatible parasites did not attract encapsulation reactions and proceeded to develop unharmed within the snail, whereas incompatible schistosomes were observed to be rapidly encapsulated and destroyed by hemocytes (Théron and Coustau 2005). Snail hemocytes are phagocytic cells circulating in the hemolymph that can be activated to exert potent cellular cytotoxicity through a respiratory burst and lytic proteins and enzymes (Adema and Loker 1997). Laboratory investigation of the determinants that rule this immunological incompatibility replicated these observations in vitro: hemocytes isolated from susceptible *B. glabrata* did not encapsulate schistosomes, yet hemocytes from resistant *B. glabrata* encapsulated and destroyed parasites (Bayne et al. 1980a). Dramatically, when incubated with the cell-free blood fluid of a resistant *B. glabrata*, hemocytes from a susceptible *B. glabrata* did engage and eliminated a normally compatible schistosome. Thus, hemocytes from these *B. glabrata* strains did not differ a priori in their cytotoxic capabilities, but their response required activation by some factor(s) for recognition of schistosome parasites as non-self that was absent from the blood of susceptible snails yet present in resistant *B. glabrata* (Bayne et al. 1980b).

2.2 *Characterization of* Biomphalaria glabrata *Lectins*

The initial experimental identification of soluble factors that could function as snail innate-type non-self recognition factors relied on incubation of various foreign cells (bacteria and mammalian erythrocytes) in blood fluid of snails. Agglutination of such probes, in concentration-dependent and a monosaccharide-inhibitable fashion, indicated the presence of lectins, non-antibody, nonenzymatic multivalent proteins that specifically recognize and reversibly bind to carbohydrates leading to cross-linking of cells bearing such surface epitopes (see Horak and van der Knaap 1997). Lectins function as innate immunity-type pathogen recognition receptors (PRR) for specific recognition of pathogen-associated molecular patterns (PAMPs) consisting of repetitive carbohydrate moieties that characterize the cell surface of groups of pathogens, such as bacteria, fungi, and metazoan parasites (Medzhitov and Janeway 1997). Lectin binding may depend on divalent metal ions as cofactors. These properties allow for sophisticated regulation of ligand binding by snails lectins and activation of immune activities (Richards and Renwrantz 1991). A range of lectin activities was recorded from *B. glabrata* (Horak and van der Knaap 1997), initially without molecular characterization of the lectin molecules responsible. Monroy et al. (1992) applied six different monosaccharides for column affinity purification of lectins from *B. glabrata* hemolymph. Purified lectins resolved on SDS-PAGE as protein bands of 80–120 kDa and ~200 kDa; occasionally other proteins bands were observed also. These proteins were most abundant in snails at 8-day post-exposure to *E. paraensei* infection versus untreated snails, suggesting that lectin expression was upregulated expression following pathogen stimulation. The indication from a consistently unfocused, diffuse aspect of the gel bands that the purified *B. glabrata* lectins were diverse in composition was confirmed by two-dimensional gel electrophoresis. These groups of lectins consisted of multiple components, differing in composition between the blood of *B. glabrata* harboring different digenean parasites, either *S. mansoni* or *E. paraensei* (Monroy and Loker 1993). The lectin-like properties of the purified snail proteins were further evident from the monosaccharide-inhibitable binding to the surface of foreign cells including bacteria, erythrocytes, and *E. paraensei* parasites (Hertel et al. 1994). Moreover, experimental coincubation of these snail lectins, as contained in cell-free plasma from *E. paraensei*-infected *B. glabrata*, with soluble secreted-excreted products (SEP) obtained from culture medium that held intramolluskan stages of the parasite *E. paraensei*, yielded a precipitate. The cross-linking between lectins and parasite SEP was concentration dependent and reversible; the precipitate dissolved when more SEP was added. Low SEP dose yielded precipitates containing snail lectins of 80–120 kDa and 200 kDa, The precipitates that formed at high dose of SEP contained another group of lectins that resolved on SDS-PAGE gels as a prominent diffuse band centered at 65 kDa, suggestive of diverse composition. The difference in affinities was used to purify this newly observed category of lectins by stepwise addition of SEP to snail plasma and removing the precipitate at each step to deplete the lectins of higher molecular weight; the precipitates formed contained only

65 kDa lectins. Analysis by gel electrophoresis showed that the 65 kDa lectins form multimers, organizing in hexamers of 400 kDa under nondenaturing conditions that combine to a native form of 1600 kDa (Adema et al. 1997a).

3 Characterization of FREPs

3.1 Identification of Lectins as FREPs

The previous function and biochemical studies clearly connected lectin activities to particular *B. glabrata* plasma proteins that increased in abundance after infection and bound relevant antigens in lectin-like fashion. Importantly, purified 65 kDa proteins were now available for further molecular characterization of this category of lectins as immune factors of *B. glabrata*. Amino acid sequences of internal peptides that were obtained from gel-purified 65 kDa proteins were used to instruct PCR primer design. PCR amplicons obtained from genomic DNA of *B. glabrata* encoded a partial sequence of 354 bp with similarities to previously sequenced fibrinogen (FBG)-like sequences. Accordingly, the 65KdA lectin from *B. glabrata* was designated a fibrinogen-related protein or FREP. The screening of a *B. glabrata* cDNA library yielded several clones that hybridized the 354 bp FREP PCR amplicon. Insert sequencing revealed that *B. glabrata* FREPs comprise a C-terminal FBG domain with Immunoglobulin-superfamily (IgSF)-like domains at the N-terminus (Adema et al. 1997b). This particular combination of domains defines bona fide FREPs at the distinction of many other, unrelated FBG domaincontaining sequences that have been reported since, and that were also designated as FREP even though they have a different domain organization. For instance, the 53 FREPs reported the arthropod *Anopheles gambiae* do not contain IgSF sequence (Wang et al. 2005).

3.2 Levels of Diversity B. glabrata *FREPs*

Screening of cDNA inserts yielded three similar yet different FREP sequences from *B. glabrata* that were named BgMFREP2-4, additional to the initial PCR fragment now designated FREP1. Each sequence clearly encoded canonical or conserved residues for the IgSF or FBG domains, yet differed in sequence content and in the length of the so-called interceding region (ICR) that connects the IgSF and FBG domains. The characterization of the structure of FREP genes from a genomic DNA insert library revealed that two different types of FREPs exist in *B. glabrata*. FREPs contain either one single or two tandemly arranged IgSF domains upstream of the 3′ FBG domain (Fig. 1). Full-length sequences were completed for four FREP genes and these show that BgMFREPs 2 and 4 have a single IgSF domain and FREP3 and

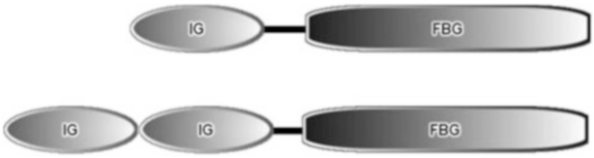


Fig. 1 Structure of two categories of *B. glabrata* FREPs. FREPs are soluble proteins circulating in snail blood, which are increased in abundance after infection and that bind parasite antigens in a lectin-like fashion. Molecular analysis revealed that FREPs consist of N-terminal immunoglobulin (IG) domains and a C-terminal fibrinogen (FBG) domain. FREPs 2, 4, and 14 each have single IG domain (*top*); the structure of other FREPs like FREP 3 and 7 includes two IG domains in a tandem arrangement (Adema et al. 1997b; Léonard et al. 2001; Zhang et al. 2001)

7 are examples of genes with two tandemly arranged IgSF domains (Léonard et al. 2001; Zhang et al. 2001) The genomic analysis identified several additional yet different FREP genes, but these were characterized only partially although complete cDNA sequences are available for several of these FREPs (Zhang and Loker 2003, 2004; Zhang et al. 2008b). It became clear that the genome of *B. glabrata* contains several FREP gene families. At one point, Loker and coworkers recognized up to 14 FREP gene families that were delineated by the arbitrary criterion of less than 85 % nucleotide identity between gene families, but recent next-generation sequencing approaches indicate that the number may be considerably greater (Dheilly et al. 2015). Particular FREP gene families comprise between 1 and 9 loci in the *B. glabrata* genome, as estimated from Southern blotting experiments (Zhang and Loker 2004). Recovery of truncated cDNA sequences for several FREPs indicated that exon loss in FREP genes or alternative splicing may increase the diversity of FREP sequences of *B. glabrata* yet further (Zhang and Loker 2003). These phenomena contribute to provide *B. glabrata* with a high diversity of FREP genes and transcripts, but continued study revealed yet another aspect of FREP diversification (Sect. 3.3).

3.3 Somatic Diversification of FREP Genes

During assembly of FREP genes from sequences derived from different inserts of genomic libraries, it was noted that otherwise identical genomic FREP sequence fragments occasionally differed randomly at single nucleotide positions (Léonard et al. 2001; Zhang et al. 2001). Further investigation of this observation focused on the FREP3 gene because it was most frequently affected in this manner. To explore diversity of this FREP, high fidelity PCR was applied with specific primers to amplify a region of exon 2 that encodes the upstream IgSF domain of FREP3. Amplicons were generated from genomic DNA of two individual *B. glabrata* (snail A and B) and subcloned to sequence individual inserts. Remarkably, this yielded a surprisingly high level of sequence diversity that far exceeded the estimated number of FREP genes in the genome of *B. glabrata* (Zhang et al. 2004). All

amplicons clearly derived from FREP3 (greater than 85 % nucleotide identity). Yet 183 clones from snail A yielded 45 different sequences; 37 different sequences were present among 173 clones from snail B, encoding for a total of 36 and 31 sequence variants at amino acid level, respectively. Only a single DNA sequence and two predicted amino acid sequences were identical between the two snails. Several sequence variants, including the one shared by snail A and B, were recovered multiple times. By contrast most variant sequences were recovered only from single clones. Some of the variants differed over a length of sequence, but most sequences were different from others due to randomly distributed, single nucleotide point mutations (SNPs); no frame shifts or stop codons were observed. Computational analysis indicated that all variant sequences combined could result from nine original (source) sequences through mutation involving point mutation and gene conversion. The computationally identified source sequences matched with the particular variants that had been recovered from multiple clones. These sequences were interpreted to represent (alleles of) FREP genes encoded in the germline of *B. glabrata*. For example, snail A contained five source sequences and these were adequate to serve as origin for all sequence variants in that snail. Source sequences were hereditary, as they were also transferred from parent to offspring. Additional analyses showed that other regions of the FREP3 gene, like the downstream FBG domain, were also affected by somatic mutation. Comparative analysis indicated that cDNA from the same exon 2 of FREP3 did not gain additional diversity and this led to the conclusion that FREP diversification occurs at the somatic level, through mutation of FREP gene sequences (Zhang et al. 2004). The mutation mechanism effects changes in complete intact FREPs genes in a manner unrelated to the gene rearrangements that drive somatic mutation of vertebrate IgSF genes. At the time of these discoveries, somatic diversification of innate immunity effector molecule in a snail was a novel concept that challenged the notion that invertebrates relied on limited numbers of germline-encoded receptors like lectins for broad recognition of groups of pathogens as proposed by Medzhitov and Janeway (1997). However, similar observations were made involving representative species of other groups of animal phylogeny that employed yet other alternative mechanisms (e.g., increased gene copy number or alternative splicing) to generate diversified immune factors, different from FREPs. It is now generally accepted that diversification of immune factors also occurs outside the jawed vertebrates (Ghosh et al. 2011; Boehm et al. 2012). In hindsight, it is remarkable that the presence of diversified innate immune factors was considered unlikely in invertebrate animals whereas it has been long known that a single-celled organism like the protozoan pathogen *Trypanosoma brucei* employs complex mechanisms for somatic mutations to effect antigenic variation (e.g., Borst et al. 1996).

The extent of sequence diversity that may be attained among different *B. glabrata* through somatic mutation of FREP3 is impressive; individuals possess different complements of alleles that are present in snail populations, while gene conversions can generate novel sequences additional to those encoded in the germline (Zhang et al. 2004). Furthermore, each region of these genes, whether IgSF, interceding region, or FBG domain, can incur point mutations (Zhang

et al. 2004; Hanington et al. 2010). Sequences from other FREP gene families may also be diversified. This has been confirmed by targeted analysis for FREP12, which belongs to the category of FREPs that have two IgSF domains (Dheilly et al. 2015), as well as FREP2 which has a single IgSF domain (Moné et al. 2010). Yet additional alleles were uncovered of previously recognized FREP gene families, and new gene families, providing greater diversity of known genomic FREP sequences when a novel strain of *B. glabrata* was queried using Illumina next-generation sequencing (Dheilly et al. 2015). Keeping in mind that expressed FREP proteins may organize in multimers under native conditions, potentially combining diverse FREP sequences (Adema et al. 1997a; Zhang et al. 2008b), it seems highly unlikely that different individual *B. glabrata* snails will possess the same repertoire of FREP-mediated binding capabilities at any time.

4 FREPs and Immunity

4.1 Functional Study of FREPs in B. glabrata

Several experimental approaches were taken to study the function of FREPs to gain further evidence in support of the putative involvement of FREPs in immunity that was already suggested by the binding in lectin-like fashion of FREP to non-self, and the increase in FREP abundance following infection of *B. glabrata*. Previously determined nucleotide sequences were applied toward recombinant expression of FREPs for the purpose of generating specific antisera against FREPs 2, 3, and 4. The use of the resulting immune-reagents in immunoblotting experiments indicated differential binding preferences of FREPs in the plasma of *E. paraensei*-infected snails (4-day-old infections): FREP2 and 4 bound primarily to *E. paraensei* parasites, whereas bacteria and fungi were bound by FREP3. Thus, different FREP categories exhibited some specialization in their binding toward particular pathogens (Zhang et al. 2008b).

4.2 FREPs Associate with Immunity

Both 70-mer oligonucleotide- and cDNA-based microarray platforms have been available to study gene expression in *B. glabrata* (Adema et al. 2010; Lockyer et al. 2008, 2012). Study of the response of *B. glabrata* at different times post-exposure to pathogens including bacteria and digenean parasites, using mRNA from either whole body tissues or hemocytes, has consistently shown differential expression of FREP3 and 4. In particular, increased expression of FREP3 was associated with effective defense responses from *B. glabrata* that were resistant to the pathogens. Specifically, FREP3 was uniquely upregulated in three different

types of pathogen resistance in *B. glabrata* (Lockyer et al. 2008, 2012; Hanington et al. 2010, 2012). These include (a) strain-specific resistance to *S. mansoni* infection (tested at 2 h and 12 h post-exposure), (b) the age-dependent resistance to *E. paraensei* parasites that develops in adult *B. glabrata*, and finally (c) so-called acquired resistance to *E. paraensei*. In the latter case, experimental exposure to irradiation-attenuated *E. paraensei* that fail to develop and die after penetration into the snail yields an interval of acquired resistance to newly encountered *E. paraensei* for which it is normally susceptible (Lie and Heyneman 1979; Hanington et al. 2010). FREP3 expression increased in response to the first encounter and remained elevated, such that it may prime the snail for a rapid and vigorous defense response that eliminates viable, non-attenuated *E. paraensei* at a renewed encounter. Conversely, downregulated expression of FREP3 (among other gene transcripts) due to an initial infection with *E. paraensei* associated with the advent of susceptibility to *S. mansoni* in otherwise resistant *B. glabrata* snails (Hanington et al. 2012). These combined observations indicate that FREP3 is a strong candidate as an immune factor that contributes to an effective defense response in *B. glabrata*.

4.3 FREPs Contribute to Immunity

RNA-interference (RNAi) techniques have been adapted for functional transcriptomics in *B. glabrata*, specifically with study of FREP function in mind. It proved feasible to employ injection of long double-stranded RNA sequences to target and knockdown transcripts from specific FREP gene families without significant impact on other FREP gene families (Jiang et al. 2006). More recently, RNAi targeted at FREPs has also been effected in *B. glabrata* using gene-specific siRNAs, delivered by injection (Hanington et al. 2010, 2012). Successful RNAi knockdown of FREP3 (confirmed at mRNA and protein level) altered the phenotype of normally resistant *B. glabrata* by increasing susceptibility to digenean infection; 30 % of exposed snails became infected with either *E. paraensei* or *S. mansoni*. This established that FREP3 is a co-determinant in deciding the resistance of a snail to infection by digenean parasites (Hanington et al. 2010, 2012).

4.4 Compatibility Polymorphism

How does sequence diversification contribute to the immune function of FREPs? Consider that in a susceptible *B. glabrata* snail, not all invading digenean parasites survive equally. A viable, healthy parasite may be observed next to the remnants of another individual parasite that was encapsulated and eliminated. Immunocompatibility is decided on an individual basis for every snail–parasite encounter (Théron and Coustau 2005). In the case of *S. mansoni*, this differential outcome is thought to result from capability of the parasite to effect antigenic variation.

Briefly, intramolluskan stages of *S. mansoni* employ gene conversions, alternative splicing, and posttranslational modification to generate a randomly diversified set of antigens consisting of polymorphic mucins (SmPoMucs). It was postulated that this prevents effective recognition of non-self and avoid a snail defense response (Roger et al. 2008; Moné et al. 2010). It is attractive to consider that such a polymorphic virulence system is countered by a polymorphic host defense system as provided by diversified FREPs in *B .glabrata*. The compatibility polymorphism hypothesis (Mitta et al. 2012) proposes that if the overall diversity of the FREP repertoire, resulting from alleles of germline genes and the variant sequences thereof, in an individual snail is adequate to recognize the unique antigenic diversity of a particular *S. mansoni*, FREP binding can signal for the activation of snail defenses resulting in elimination of that parasite by hemocyte encapsulation. The parasite escapes elimination if its antigens are not adequately bound by FREPs. Moné et al. (2010) have confirmed that SmPoMucs and FREPs bind to form precipitates that also include a snail Thioester containing protein (TEP); this is of interest because TEPs are thought to signal activation of defense responses in invertebrates. In this model, FREPs are important immune recognition factors that are diversified through somatic mutation to broaden the range of antigens that they can bind to in order to counter antigenically variable pathogens.

5 Dealing with Diversification

5.1 A Model for FREP Diversification in B. glabrata

For further study of the role of diversified FREPs in immune function of *B. glabrata*, it is important to resolve the underlying mechanisms that generate sequence diversity. An initial question relates to the fact that *B. glabrata* snails always display FREP diversity, both when infected and in the absence of (observed) pathogens (Zhang et al. 2004). Moreover, the majority of expressed FREP sequences are true to the source (or germline) sequences, individual variant sequences are observed at low frequency among FREP genes. Significantly, FREPs are expressed by hemocytes, the circulating defense cells of *B. glabrata* (Adema et al. 1997b; Hanington et al. 2010). A model for FREP diversification can be based on consideration of the biological characteristics of these cells. Hemocytes of *B. glabrata* do not multiply by cell division; rather these cells proliferate in the amebocyte producing organ (APO) and other areas of specialized connective tissue where they differentiate from progenitor cells (Lie et al. 1975a, b). It is assumed that random somatic mutations may occur during cell differentiation in a way that modifies germline FREP genes in some but not all hemocytes (Stout et al. 2009). The outcome of this is that the majority of cells in the hemocyte population express germline-encoded (source) FREP gene sequences; variant sequences are only expressed by individual cells that incurred a unique mutation in a FREP gene.

This pattern is observed experimentally; multiple subsets of 20–40 hemocytes from individual *B. glabrata* contained distinct repertoires of modified FREP sequences (Hanington et al. 2010). Unique FREP variants are expressed in the snail only for the limited life span of the hemocyte from which they originate. Pathogen stimulation may increase hemocyte proliferation in *B. glabrata* (Lie et al. 1976) and thereby cause plasma FREP proteins to increase both in abundance and in diversity. Persistence of the increased repertoire of diverse FREP sequences may prime the defenses of *B. glabrata* to respond effectively to new pathogens, at least until the resident hemocyte population is replaced by a new generation of FREP-producing hemocytes (Hanington et al. 2012; Adema and Loker 2015). As a consequence of this model, individual hemocytes are functionally different; moreover, a continuous replacement of hemocytes alters the immunological identity of an individual *B. glabrata* over time.

5.2 *Mechanisms for Somatic Mutation of FREPs*

It must then be considered how somatic mutation of FREP genes, consisting of gene conversions and point mutations (Zhang et al. 2004; Hanington et al. 2010), is achieved at the genomic and at the biochemical/molecular level. To be effective, obviously, somatic mutations may not generate lethal mutations in housekeeping genes or damage the germline. A BAC library produced with genomic DNA from *B. glabrata* (Adema et al. 2006) was employed to probe the genomic organization of FREP genes. The insert of a clone that was reactive with a FREP-specific probe was sequenced and assembled (Hanington et al. 2010) to reveal the clustering of 4 FREP genes within a genomic region of 70,000 nt. (Fig. 2). Additional to FREP14 (which has one IgSF domain) and three members of the FREP3 gene family, partial FREP gene-like fragments were encountered.

This configuration, clustering of highly similar genes (and sequence fragments), is frequently associated with somatic mutation and gene conversion (Hanington et al. 2010; Chen et al. 2007).

Inspection of the patterns of diversity caused by point mutations in FREP sequences, and the lack thereof in control genes (Zhang et al. 2004; Hanington et al. 2010), provides clues for the nature of the mechanisms of mutation. The specificity toward FREP target sequences and the rare occurrence of frame shifts or premature stop codons in variant FREP sequences may indicate that somatic mutation is effected by cytidine deaminase (APOBECs)-like enzymes, present in bacteria and eukaryotes, that recognize target DNA sequence motifs of which they mutate only a single nucleotide (Bransteitter et al. 2009; Larijani et al. 2007). This mechanism targets particular sequences and avoids indels and thus frameshifts, and somatic mutation of only particular sequence motifs may reduce the likelihood of creating premature stop codons (Kühn et al. 1993; Rogozin et al. 2007). *Biomphalaria glabrata* possesses a cytidine deaminase-encoding sequence that was found to be upregulated during immune responses (Genbank accession

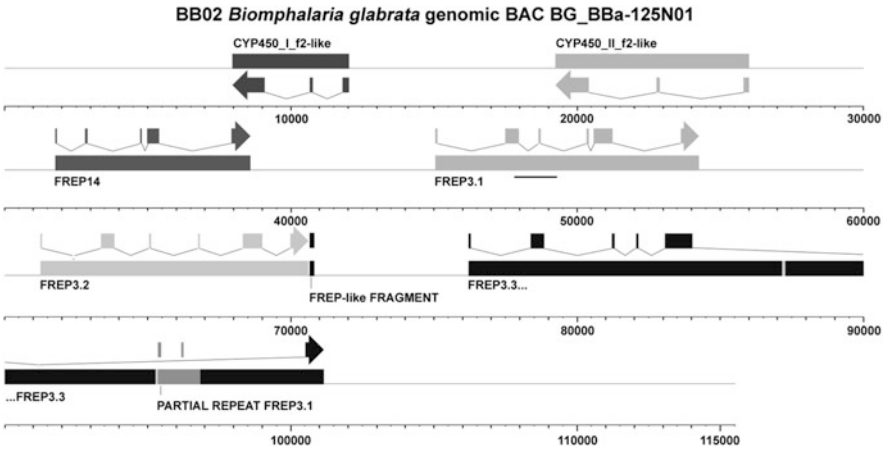


Fig. 2 A cluster of FREP genes in the genome of *B. glabrata*. The assembly of a BAC clone (Genbank accession JN382156) showed four FREP genes in addition to two cytochrome P450 genes. *Boxes* indicate the full-length coding region, with exons indicated aside. *Arrowheads* indicate the directionality of the genes. The *underlined region* of FREP3.1 is duplicated in an intron of FREP3.3 (~95,000–97,000 nt). A fragment of FREP-like coding sequence occurs just downstream from FREP3.2. *Scale bar* indicates nucleotide positions

DQ117977, Bouchut et al. 2006). The occurrence of only single point mutations in the FREP sequences also suggests epigenetic control as another alternative candidate mechanism for somatic mutation. Epigenetic mechanisms target CpG dinucleotides for DNA methylation and this may lead to spontaneous mutation of the cytosine. *Biomphalaria glabrata* has the essential elements of the machinery needed for DNA methylation (Geyer et al. 2011; Fneich et al. 2013). DNA methylation of exonic sequences occurs in invertebrates, including mollusks (Gavery and Roberts 2010), and it is proposed to mediate epigenetic regulation of genes that manage environmental stress and immune genes. While these considerations provide leads for research, the mechanism(s) for targeted somatic diversification of FREP genes remain to be resolved.

6 FREPs and Fibrinogen-Containing Lectins in Other Mollusks

The knowledge of function and somatic diversification of FREPs as molluskan immune factors derives almost exclusively from *B. glabrata*. In large part, this is because bona fide FREPs, which by definition combine IgSF and FBG domains, were exclusively known only from *B. glabrata* for a long interval of time. It seemed unlikely, however, that diversified immune genes like FREPs occur only in this single planorbid gastropod. A biochemical study, not designed to generate sequence

data, makes it plausible that FREPs also function in the immunity of *Helisoma trivolvis*, a species of the sister genus (*Helisoma)* to the genus *Biomphalaria* (Adema et al. 1999). This comparative study showed side by side that digenean pathogens evoked complex patterns of plasma proteins from *H. trivolvis*, with dynamics akin to those of the FREPs from *B. glabrata*. Searches employing Southern blotting and PCR with conserved primers failed to provide conclusive evidence for FREPs from other genera of gastropods. New insights arose with the more recent availability of genome sequences. While computational analyses remained negative in the case of the basal gastropod *Lottia gigantea*, two bona fide FREP genes (each containing two IgSF domains and one FBG domain) were described from the sea hare *Aplysia californica* (Opisthobranchia, Gastropoda). Interestingly, experimental work did not provide indications that these FREPs were diversified (Gorbushin et al. 2010). Considerable sequence differences among FREPs from *A. californica* and *B. glabrata* provide no opportunity for successful design of general FREP-specific PCR primers to explore the presence of FREP genes across the phylogenetic range between *Aplysia* and *Biomphalaria*. The anticipated benefit from next-generation sequencing for extending the scope of comparative immunology (Dheilly et al. 2014) was evident from a recent report of a bona fide FREP (with two IgSF domains) from a marine gastropod *Littorina littorea*, although no further details are available at this time regarding possible diversification of this sequence (Gorbushin and Borisova 2015). With the anticipation that FREPs will be recovered from additional species of mollusks in the near future, current indications suggest that FREPs are distributed only among the Heterobranchia (Euthyneura), a subgroup of more derived gastropods (Gorbushin et al. 2010, also see Kocot et al. 2011). This does not mean that polymorphic systems involving FBG domains containing proteins do not exist in other mollusks. The molecular characterization of a sialic acid-binding lectin from the slug *Limax flavus* (Gastropoda) that agglutinates bacteria yielded three closely related sequences consisting of only an FBG domain (Kurachi et al. 1998). A great diversity of FBG-containing sequences has been reported from bivalves, another class of the Molluska that includes mussels and oysters. Many of these are single FBG domain lectins (Gorbushin and Iakovleva 2011; Yang et al. 2014; Zhang et al. 2008a). These types of fibrinogen-containing lectins may also function as polymorphic systems to counter pathogens. Romero et al. (2011) showed that collectin-like lectins from the mussel *Mytilus galloprovincialis*, designated as MuFREPs, were increased in abundance after infection and possessed opsonic activities; moreover, three individual *M. galloprovincialis* each had a distinct repertoire of MuFREP sequences. More recently, it was reported that the genome of the oyster, *Crassostrea gigas*, contains over 190 genes that include an FBG domain and at least some of these exhibit high levels of polymorphism, including some due to diversification by point mutation (Zhang et al. 2012; Huang et al. 2015). Future studies are required to further document the involvement of such polymorphic systems in bivalve immunity. It is interesting to speculate that perhaps the evolution of molluskan immunity cannot be traced by the identities of particular polymorphic immune effector genes, but rather by the mechanism(s) that

effect sequence diversification, like the cytidine deaminase (AID-CDA) that affects variable lymphocyte receptors (VLRs) in agnathans and IgSF in gnathostomes (Boehm et al. 2012; Tran et al. 2010).

7 Do Variant FREPs Aid *B. glabrata* Immunity?

The polymorphic nature of FREPs and the immune relevance of FREPs have been demonstrated in multiple ways (see Sects. 3.3 and 4). It remains unclear to what extent FREP diversity benefits immune function of *B. glabrata*. Consideration of the following issues may guide future research toward better understanding of snail immunity.

It is remarkable to consider that a complex system for somatic mutation is in place to generate polymorphic FREP sequences, yet that the individual variant sequences occur only at very low frequency among a majority of unmodified

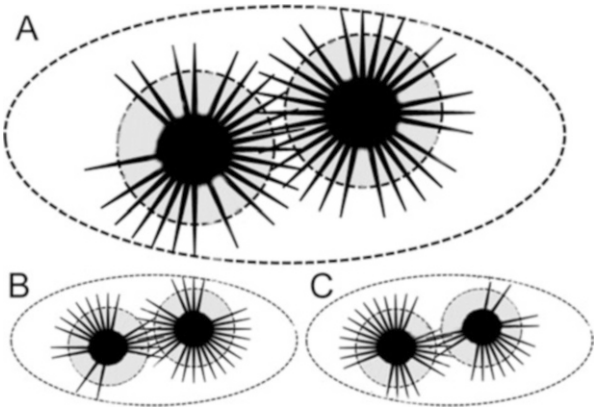


Fig. 3 A hypothetical model for how sequence diversification may contribute to immune function of FREP genes. (**a**) The area in the large oval represents the total range of antigenic diversity that can be achieved by the digenean *S. mansoni* as a species; the antigenic identity of a single parasite is located as a point somewhere within this range. An individual *Biomphalaria glabrata* snail is immunologically resistant to parasites with antigenic signatures that fall within the non-self recognition capabilities of two germline-encoded FREP genes (*gray circles*). Random mutations provide FREP variants with new recognition capabilities that extend (*black spikes*) the range of immune resistance to include parasites with antigenic identities not covered by the germline. In this manner, FREP diversification serves in anticipation of polymorphic pathogens that may be encountered. The impact of random mutations that reduce the recognition capabilities of a single FREP variant (*black circle*) is negated by the continued expression in large numbers of unaltered germline FREP sequences. (**b**) Unique variant FREP sequences are expressed by single individual hemocytes. The continuous replacement of these defense cells leads to changes in the immunological identity (immune recognition afforded by germline plus diversified FREPs) over time [compare (**a**) and (**b**)]. (**c**) The random nature of somatic mutation produces individually unique FREP repertoires. Accordingly, individual *B. glabrata* with the same germline-encoded FREP genes have distinct immune profiles that differ over time [compare (**a**), (**b**) and (**c**)]

germline-encoded (source) sequences. Additionally, the mutational process is random and may affect all domains of a FREP sequence (Zhang et al. 2004; Hanington et al. 2010; Dheilly et al. 2015). Thus, it is conceivable that sequence diversification generates both FREP variants with novel specificity and FREP variants that are less efficient than the germline-encoded sequences. Finally, FREPs are expressed in hemocytes, what then are the consequences of continuous replacement of these nondividing cells with newly proliferated hemocytes for the immune function of FREPs? Figure 3 and the associated legend provide a theoretical framework to consider these issues. In summary, FREP diversification may well extend the non-self recognition capabilities of *B. glabrata* to a greater range than provided by the germline-encoded FREP genes, in a manner that anticipates future encounters with polymorphic pathogens while providing *B. glabrata* snails with distinct immunological identity that varies over time. These considerations provide significant impetus to resolve the mechanism(s) responsible for somatic mutation of FREP genes. In theory at least, such mechanism(s) could be targeted experimentally to inhibit or reduce FREP diversification and investigate how effective the immunity of *B. glabrata* is when it lacks polymorphic FREPs and must rely only on the germline-encoded FREP genes.

8 Conclusion

Persistent comparative immunology research in invertebrates, with *B. glabrata* as example here, continues to expand our understanding of animal immunology. The diversification of FREPs likely augments the innate-type defense capabilities beyond that of germline-encoded genes and bestows a unique immunological identity upon individual snails that reduces vulnerability to pathogens. Characterization of the underlying mechanism(s) will give a better understanding of how this polymorphic system is regulated and make it possible to explore the contribution of FREPs to immunity of gastropods that encounter pathogens in natural environments. FREPs will inform on evolution of immunity in gastropod mollusks, but may not provide a uniform model for all mollusks. Already, the polymorphic systems recorded from bivalves involve genes that do contain fibrinogen sequences, but are otherwise distinct and clearly have a different evolutionary history. Insights regarding yet other polymorphic defense factors in mollusks, especially in other classes of mollusks (including cephalopods, chitons, and others), may arise with future explorations that are enabled by next-generation sequencing approaches. At this time however, available data restrict us to interpret the implications of polymorphic FREPs for the immune function of *B. glabrata*. Importantly, somatic diversification gives rise to continuously changing, randomly individualized FREP repertoires in *B. glabrata* snails that provide increased recognition capabilities but there is no apparent long-term retention of effective FREP variants. Accordingly, *B. glabrata* immunity is not adaptive (no immune memory) but it is

anticipatory, in that it effects diversification of FREPs to counter pathogens that actively employ antigenic variation.

Acknowledgments CMA acknowledges support from CETI and NIH grant number P20GM103452 from the National Institute of General Medical Sciences (NIGMS).

References

Adema CM, Loker ES (1997) Specificity and immunobiology of larval digenean-snail associations. In: Fried B, Graczyk TK (eds) Advances in trematode biology. CRC Press, Boca Raton, FL, p 230

Adema CM, Loker ES (2015) Digenean-gastropod host associations reveal foundations of specific immunity in snails. Dev Comp Immunol 48:275–283. doi:10.1016/j.dci.2014.06.014

Adema CM, Hertel LA, Loker ES (1997a) Infection with *Echinostoma paraensei* (Digenea) induces parasite-reactive polypeptides in the hemolymph of the gastropod host Biomphalaria glabrata. In: Beckage N (ed) Parasite effects on host physiology and behavior. Chapman Press, New York, p 77

Adema CM, Hertel LA, Miller RD et al (1997b) A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. Proc Natl Acad Sci USA 1997(94):8691–8696

Adema CM, Hertel LA, Loker ES (1999) Evidence from two planorbid snails of a complex and dedicated response to digenean (echinostome) infection. Parasitology 119:395–404

Adema CM, Luo M-Z, Hanelt B et al (2006) A BAC library for *Biomphalaria glabrata*, intermediate snail host of *Schistosoma mansoni*. Mem Inst Oswaldo Cruz 101:167–177

Adema CM, Hanington PC, Lun CM et al (2010) Differential transcriptomic responses of *Biomphalaria glabrata* (Gastropoda, Mollusca) to bacteria and metazoan parasites, *Schistosoma mansoni* and *Echinostoma paraensei* (Digenea, Platyhelminthes). Mol Immunol 47:849–860. doi:10.1016/j.molimm.2009.10.019

Bayne CJ (2009) Successful parasitism of vector snail *Biomphalaria glabrata* by the human blood fluke (trematode) *Schistosoma mansoni*: a 2009 assessment. Mol Biochem Parasitol 165:8–18. doi:10.1016/j.molbiopara.2009.01.005

Bayne CJ, Buckley PM, DeWan PC (1980a) Macrophagelike hemocytes of resistant *Biomphalaria glabrata* are cytotoxic for sporocysts of *Schistosoma mansoni* in vitro. J Parasitol 66:413–419

Bayne CJ, Buckley PM, DeWan PC (1980b) *Schistosoma mansoni*: cytotoxicity of hemocytes from susceptible snail hosts for sporocysts in plasma from resistant *Biomphalaria glabrata*. Exp Parasitol 50:409–416

Boehm T, McCurley N, Sutoh Y et al (2012) VLR-based adaptive immunity. Annu Rev Immunol 30:203–220. doi:10.1146/annurev-immunol-020711-075038

Borst P, Rudenko G, Taylor MC et al (1996) Antigenic variation in trypanosomes. Arch Med Res 27:379–388

Bouchut A, Sautiere PE, Coustau C et al (2006) Compatibility in the *Biomphalaria glabrata*/*Echinostoma caproni* model: potential involvement of proteins from hemocytes revealed by a proteomic approach. Acta Trop 98:234–246

Bransteitter R, Prochnow C, Chen XS (2009) The current structural and functional understanding of APOBEC deaminases. Cell Mol Life Sci 66:3137–3147, PMID: 19547914

Bruun B, Aagaard-Hansen J (2008) The social context of Schistosomiasis and its control an introduction and annotated bibliography. World Health Organization, Geneva

Chen JM, Cooper DN, Chuzhanova N et al (2007) Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8:762–775

Dheilly NM, Adema CM, Raftos DA et al (2014) No more non-model species: the promise of next generation sequencing for comparative immunology. Dev Comp Immunol 45:56–66. doi:10.1016/j.dci.2014.01.022

Dheilly NM, Duval D, Mouahid G et al (2015) A family of variable immunoglobulin and lectin domain containing molecules in the snail *Biomphalaria glabrata*. Dev Comp Immunol 48:234–243. doi:10.1016/j.dci.2014.10.009

Erwin DH, Laflamme M, Tweedt SM et al (2011) The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science 334:1091–1097. doi:10.1126/science.1206375

Esch GW, Barger MA, Fellis KJ (2002) The transmission of digenetic trematodes: style, elegance, complexity. Integr Comp Biol 42:304–312. doi:10.1093/icb/42.2.304

Fneich S, Dheilly N, Adema CM et al (2013) 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*. Parasit Vectors 6:167. doi:10.1186/1756-3305-6-167

Gavery MR, Roberts SB (2010) DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). BMC Genomics 11:483. doi:10.1186/1471-2164-11-483

Geyer KK, Rodríguez López CM, Chalmers IW et al (2011) Cytosine methylation regulates oviposition in the pathogenic blood fluke Schistosoma mansoni. Nat Commun 9(2):424. doi:10.1038/ncomms1433

Ghosh J, Lun CM, Majeske AJ et al (2011) Invertebrate immune diversity. Dev Comp Immunol 35:959–74. doi:10.1016/j.dci.2010.12.009

Gorbushin AM, Borisova EA (2015) Lectin-like molecules in transcriptome of *Littorina littorea* hemocytes. Dev Comp Immunol 48:210–20. doi:10.1016/j.dci.2014.10.007

Gorbushin AM, Iakovleva NV (2011) A new gene family of single fibrinogen domain lectins in *Mytilus*. Fish Shellfish Immunol 30(1):434–438. doi:10.1016/j.fsi.2010.10.002

Gorbushin AM, Panchin YV, Iakovleva NV (2010) In search of the origin of FREPs: characterization of *Aplysia californica* fibrinogen-related proteins. Dev Comp Immunol 34:465–473. doi:10.1016/j.dci.2009.12.007

Hanington PC, Forys MA, Dragoo JW et al (2010) Role for a somatically diversified lectin in resistance of an invertebrate to parasite infection. Proc Natl Acad Sci USA 107:21087–21092. doi:10.1073/pnas.1011242107

Hanington PC, Forys MA, Loker ES (2012) A somatically diversified defense factor, FREP3, is a determinant of snail resistance to schistosome infection. PLoS Negl Trop Dis 6, e1591. doi:10.1371/journal.pntd.0001591

Hertel LA, Stricker SA, Monroy FP et al (1994) *Biomphalaria glabrata* hemolymph lectins: binding to bacteria, mammalian erythrocytes, and to sporocysts and rediae of *Echinostoma paraensei*. J Invertebr Pathol 64:52–61

Hetru C, Hoffmann JA (2009) NF-kappaB in the immune response of *Drosophila*. Cold Spring Harb Perspect Biol 1:a000232. doi:10.1101/cshperspect.a000232

Horak P, van der Knaap WPW (1997) Lectins in snail – trematode immune interactions: a review. Folia Parasitol (Praha) 44:161–172

Huang B, Zhang L, Li L et al (2015) Highly diverse fibrinogen-related proteins in the Pacific oyster *Crassostrea gigas*. Fish Shellfish Immunol 43(2):485–490. doi:10.1016/j.fsi.2015.01.021, pii: S1050-4648(15)00035-2

Jiang Y, Loker ES, Zhang SM (2006) *In vivo* and *in vitro* knockdown of FREP2 gene expression in the snail *Biomphalaria glabrata* using RNA interference. Dev Comp Immunol 30:855–866

Klein J (1989) Are invertebrates capable of anticipatory immune responses? Scand J Immunol 29:499–505. doi:10.1111/j.1365-3083.1989.tb01152.x

Kocot KM, Cannon JT, Todt C et al (2011) Phylogenomics reveals deep molluscan relationships. Nature 477:452–456. doi:10.1038/nature10382

Kühn K, Bertling WM, Emmrich F (1993) Cloning of a functional cDNA for human cytidine deaminase (CDD) and its use as a marker of monocyte/macrophage differentiation. Biochem Biophys Res Commun 15:1–7

Kurachi S, Song Z, Takagaki M et al (1998) Sialic-acid-binding lectin from the slug Limax flavus–cloning, expression of the polypeptide, and tissue localization. Eur J Biochem 254:217–222

Langand J, Morand S (1998) Heritable non-susceptibility in an allopatric host-parasite system: *Biomphalaria glabrata* (Mollusca)-*Echinostoma caproni* (platyhelminth Digenea). J Parasitol 84:739–742

Larijani M, Petrov AP, Kolenchenko O et al (2007) AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence independent manner. Mol Cell Biol 27:20–30

Léonard PM, Adema CM, Zhang SM et al (2001) Structure of two FREP genes that combine IgSF and fibrinogen domains, with comments on diversity of the FREP gene family in the snail *Biomphalaria glabrata*. Gene 269:155–165

Lie KJ, Heyneman D (1979) Acquired resistance to echinostomes in four *Biomphalaria glabrata* strains. Int J Parasitol 9:533–537

Lie KJ, Heyneman D, Lim HK (1975a) Studies on resistance in snails: specific resistance induced by irradiated miracidia of *Echinostoma lindoense* in *Biomphalaria glabrata* snails. Int J Parasitol 5:627–631

Lie KJ, Heyneman D, Yau D (1975b) The origin of amoebocytes in *Biomphalaria glabrata*. J Parasitol 61:574–576

Lie JK, Heyneman D, Jeong KH (1976) Studies on resistance in snails. 4. Induction of ventricular capsules and changes in the amebocyte-producing organ during sensitization of *Biomphalaria glabrata* snails. J Parasitol 62:286–291

Lockyer AE, Spinks J, Kane RA et al (2008) *Biomphalaria glabrata* transcriptome: cDNA microarray profiling identifies resistant- and susceptible-specific gene expression in haemocytes from snail strains exposed to *Schistosoma mansoni*. BMC Genomics 9:634. doi:10.1186/1471-2164-9-634

Lockyer AE, Emery AM, Kane RA et al (2012) Early differential gene expression in haemocytes from resistant and susceptible *Biomphalaria glabrata* strains in response to *Schistosoma mansoni*. PLoS One 7(12), e51102

Marchalonis JJ, Schluter SF (1990) Origins of Immunoglobulins and Immune Recognition Molecules. BioScience 40:758–768

Medzhitov R, Janeway CA (1997) Innate immunity: the virtues of a non-clonal system of recognition. Cell 91:295–298

Mitta G, Adema CM, Gourbal B et al (2012) Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. Dev Comp Immunol 37:1–8. doi:10.1016/j.dci.2011.09.002

Moné Y, Gourbal B, Duval D et al (2010) A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. PLoS Negl Trop Dis 4. pii: e813. doi:10.1371/journal.pntd.0000813.

Monroy FP, Loker ES (1993) Production of heterogeneous carbohydrate-binding proteins by the host snail *Biomphalaria glabrata* following exposure to *Echinostoma paraensei* and *Schistosoma mansoni*. J Parasitol 79(3):416–423

Monroy F, Hertel LA, Loker ES (1992) Carbohydrate-binding plasma proteins from the gastropod *Biomphalaria glabrata*: strain specificity and the effects of trematode infection. Dev Comp Immunol 16:355–366

Richards EH, Renwrantz LR (1991) Two lectins on the surface of *Helix pomatia* haemocytes: a Ca2 + -dependent, GalNac-specific lectin and a Ca2 + -independent, mannose 6-phosphate-specific lectin which recognises activated homologous opsonins. J Comp Physiol B 161:43–54

Richards CS, Knight M, Lewis FA (1992) Genetics of *Biomphalaria glabrata* and its effect on the outcome of *Schistosoma mansoni* infection. Parasitol Today 8:171–174

Roger E, Grunau C, Pierce RJ et al (2008) Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). PLoS Negl Trop Dis 2(11), e330. doi:10.1371/journal.pntd.0000330

Rogozin IB, Iyer LM, Liang L et al (2007) Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AIDAPOBEC family cytosine deaminase. Nat Immunol 8:647–656

Romero A, Dios S, Poisa-Beiro L et al (2011) Individual sequence variability and functional activities of fibrinogen-related proteins (FREPs) in the Mediterranean mussel (*Mytilus galloprovincialis*) suggest ancient and complex immune recognition models in invertebrates. Dev Comp Immunol 35:334–344. doi:10.1016/j.dci.2010.10.007

Stout BA, Adema CM, Zhang S-M et al (2009) The biology of FREPs: diversified lectins with fibrinogen-related domains from the freshwater snail *Biomphalaria glabrata*. In: Vasta G, Ahmed H (eds) Animal lectins: a functional view. CRC Press and Taylor & Francis, Boca Raton, FL, p 475

Théron A, Coustau C (2005) Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? J Helminthol 79:187–191

Tran TH, Nakata M, Suzuki K et al (2010) B cell-specific and stimulation-responsive enhancers depress AICDA by overcoming the effects of silencers. Nat Immunol 11:148–154

Wang X, Zhao Q, Christensen BM (2005) Identification and characterization of the fibrinogen-like domain of fibrinogen-related proteins in the mosquito, *Anopheles gambiae*, and the fruitfly, *Drosophila melanogaster*, genomes. BMC Genomics 6:114

Yang C, Wang L, Zhang H et al (2014) A new fibrinogen-related protein from *Argopecten irradians* (AiFREP-2) with broad recognition spectrum and bacteria agglutination activity. Fish Shellfish Immunol 38:221–229. doi:10.1016/j.fsi.2014.03.025

Zhang SM, Loker ES (2003) The FREP gene family in the snail *Biomphalaria glabrata*: additional members, and evidence consistent with alternative splicing and FREP retrosequences. Dev Comp Immunol 27:175–187

Zhang SM, Loker ES (2004) Representation of an immune responsive gene family encoding fibrinogen-related proteins in the freshwater mollusc *Biomphalaria glabrata*, an intermediate host for *Schistosoma mansoni*. Gene 341:255–266

Zhang SM, Léonard PM, Adema CM et al (2001) Parasite-responsive IgSF members in the snail *Biomphalaria glabrata*: characterization of novel genes with tandemly arranged IgSF domains and a fibrinogen domain. Immunogenetics 53:684–694

Zhang SM, Adema CM, Kepler TB et al (2004) Diversification of Ig superfamily genes in an invertebrate. Science 305:251–254

Zhang H, Wang L, Song L et al (2008a) A fibrinogen-related protein from bay scallop *Argopecten irradians* involved in innate immunity as pattern recognition receptor. Fish Shellfish Immunol 26:56–64. doi:10.1016/j.fsi.2008.07.019

Zhang SM, Nian H, Zeng Y et al (2008b) Fibrinogen-bearing protein genes in the snail *Biomphalaria glabrata*: characterization of two novel genes and expression studies during ontogenesis and trematode infection. Dev Comp Immunol 32:1119–1130. doi:10.1016/j.dci.2008.03.001

Zhang LL, Li L, Zhang GF (2012) Sequence variability of fibrinogen-related proteins (FREPs) in *Crassostrea gigas*. Chin Sci Bull 57:3312–3319. doi:10.1007/s11434-012-5155-6

Somatic and Germline Diversification of a Putative Immunoreceptor within One Phylum: Dscam in Arthropods

Daniela Brites and Louis Du Pasquier

Abstract Arthropod Dscam, the homologue of the human Down Syndrome cell adhesion molecule, is a receptor used by the nervous and immune systems. Unlike in vertebrates, evolutionary pressure has selected and maintained a vast Dscam diversity of isoforms, known to specifying neuronal identity during the nervous system differentiation. This chapter examines the different modes of Dscam diversification in the context of arthropods' evolution and that of their immune system, where its role is controversial.

In the single Dscam gene of insects and crustaceans, mutually exclusive alternative splicing affects three clusters of duplicated exons encoding the variable parts of the receptor. The Dscam gene produces over 10,000 isoforms. In the more basal arthropods such as centipedes, Dscam diversity results from a combination of many germline genes (over 80) with, in about half of those, the possibility of alternative splicing affecting only one exon cluster. In the even more basal arthropods, such as chelicerates, no splicing possibility is detected, but there exist dozens of germline Dscam genes. Compared to controlling the expression of multiple germline genes, the somatic mutually alternative splicing within a single gene may offer a simplified way of expressing a large Dscam repertoire. Expressed by hemocytes, Dscam is considered a phagocytic receptor but is also encountered in solution. More information is necessary about its binding to pathogens, its role in phagocytosis, its possible role in specifying hemocyte identity, its kinetics of expression, and the regulation of its RNA splicing to understand how its diversity is linked to immunity.

D. Brites (✉)
Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland
e-mail: d.brites@unibas.ch

L. Du Pasquier
Institute of Zoology and Evolutionary Biology, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland
e-mail: louis.dupasquier@unibas.ch

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_6

1 Introduction: A Question of Mind-Sets

The past 45 years have seen waves of theories and fashions concerning our understanding of the evolution of immune systems of metazoa. Opinions have varied from invertebrates having no "immunology" [that is, "no antibody no immunology" (Jerne 1970)] to having adaptive specific responses similar to those of vertebrates (Hildemann et al. 1979). Often conclusions were reached prematurely, simply excluding one or the other type of immunity. This chaotic evolution of our theories was often due to predetermined ideas, mind-sets fostered by the great discoveries in the vertebrate immune system: the greater the discovery, the more firmly entrenched the mind-set!

> The great advance of one generation may retard progress in the next. This is because each advance induces a mindset in the scientists that slants the interpretation of data and new speculations. (Silverstein 1989).

Let us take one example, within the immune system. Two great advances made between the 1950s and the mid 1970s were:

1. The discovery that graft rejection was an immunological specific reaction with memory (Billingham et al. 1954).
2. The discovery of the somatic generation of antibody diversity in the adaptive immune system of gnathostome vertebrates (Hozumi and Tonega 1976).

Both findings created mind-sets.

> Thus, we saw above how the brilliant progress in antibody research early in the century led to later delays and false leads in the exploration of cellular immune reactions. (. . .) We must take care always to question our preconceptions, rather than taking the easier path of designing experiments to confirm them. (Silverstein 1989)

These mind-sets slanted the approach toward understanding invertebrate immunity. It was logical to find out whether specific allograft rejection with memory and the "antibody model" (*sensu lato*) applied to all metazoa. But many scientists quickly veneered the invertebrates with models influenced by the vertebrate mind-sets, after overinterpreting experiments that were not performed rigorously. This lasted until molecular biology analyses failed to confirm the soft results that suggested the existence of specific adaptive response in invertebrates. Yet, almost silently, since 1972 invertebrate immunity was undergoing a revolution initiated by Hans Boman. Working with the giant silk moth *Hyalophora cecropia*, Boman set the stage for the study of insect immunology, independently of the "Vertebrate mind-set." Soon *Drosophila* was introduced among invertebrate immunity models (Boman et al. 1972). In 1993, almost simultaneously, two papers demonstrated the involvement of the NF-Kappa B cascade in the triggering of the production of antimicrobial peptides in *Drosophila*. This research has led to numerous seminal discoveries including the discovery of Toll—and Toll-like receptors—as key regulators of the immune response, first in cultured *Drosophila* cells (Rosetto et al. 1995), then in *Drosophila* in vivo (Lemaitre et al. 1996), and finally in

mammalian cells (Medzhitov et al. 1997). In mammals these discoveries marked the beginning of the modern innate immunity era.

But...History repeating itself, a new mind-set was created and adaptive immunity examples in invertebrates left the scene. Everybody wanted to see an "all innate" configuration of the immune systems, which left no room for somatic adaptations. Yet, considering the selective pressures and constraints that apply to any immune system, adaptive responses with somatic generation of receptor diversity confer some advantages, and neglecting these, as many "innate immunity mind-set" brains did, was premature. In fact the majority of creatures in the tree of life use a specific adaptive immune system with memory, as long as we include in our survey Bacteria and Archaea of which a large fraction uses CRISPR-mediated specific immunity (Horvath and Barrangou 2010). In addition, a set of results obtained originally by evolutionary biologists in the early 2000s suggested that specific responses with memory could indeed be observed in several invertebrates. Unfortunately terminology problems and lack of experimental rigor cast doubt on many examples of the so-called "evidence for specific memory" in invertebrates. Yet some observations remain that cannot be explained by known nonspecific innate immunity mechanisms (Schmid-Hempel 2005; Hauton and Smith 2007; Sadd and Schmid-Hempel 2006).

So how are the immunity cards distributed among metazoa? What have been the evolutionary pathways? How have selection pressures to individualize responses been met? How did somatic adaptation come about? The picture is not going to be black and white as an earlier review has already suggested (Loker et al. 2004). First, do we have models to answer some of these questions? At least one group of animals, the arthropods, offers a category of receptors that can be used to address one issue pertinent to the above questions, that of diversification and somatic adaptations. Arthropods express a vast diversity of a putatively immunologically relevant receptor, Dscam, the homologue of the human Down syndrome cell adhesion molecule. This diversity can be obtained by two strategies: either by expression of many germline genes, or somatically, that is to say during ontogeny, by an alternative splicing mechanism within a single gene (Brites et al. 2013). This chapter discusses these diversities, but in doing this we will also reconsider the role played by this receptor, because we think many interpretations concerning the role of Dscam in immunity have been tainted by mind-sets!

2 Dscam: A Brief Introduction

Dscam of arthropods (Schmucker and Chen 2009) is the homologue of the Down syndrome cell adhesion molecule of vertebrates [DSCAM (Yamakawa et al. 1998; Schmucker and Chen 2009)]. It is an integral membrane protein with external features of a cell adhesion molecule. In its extracellular part, it consists of an assembly of immunoglobulin superfamily (IgSF) and fibronectin type III (FnIII) domains, followed by a hydrophobic transmembrane segment and by a long

cytoplasmic tail rich in motifs involved in signaling (Fig. 1a). This structure is common among cell adhesion molecules of metazoans (Fig. 1a). The crystal structure of the first seven Ig domains of *Drosophila* Dscam-hv has been elucidated and the shape of the extracellular part of molecule determined. The first 4 Ig domains form a sort of horseshoe-like structure (Meijers et al. 2007) also encountered in several molecules with distal Ig domains like roundabout, hemolin, neurexin, axonin, and abscam (Brites et al. 2013) (Fig. 1b). The whole extracellular part forms an S-shaped structure with Ig7 lining up with the axis of the horseshoe, resulting in the configuration of various surfaces providing possibilities of interaction with ligands (Sawaya et al. 2008) (Fig. 1b). Three ligands or categories of ligands are known to bind Dscam in different regions of the molecule:

1. Dscam isotypes in homologous interactions within the nervous system (Hattori et al. 2008).
2. Heterologous diverse ligands from parasite surfaces, pathogen or not (virus, bacteria, perhaps protozoa), hence a putative role in immunity (Sawaya et al. 2008)
3. Netrin, a diffusible protein provoking chemotactic clues for axon guidance (Ly et al. 2008; Lai Wing Sun et al. 2011).

Dscam gene architecture is well conserved throughout metazoans (Fig. 1a), but compared to Dscam in vertebrates and in mollusks, the peculiarity of Dscam in arthropods is its diversity of isoforms. The most interesting aspect is that this diversity is not generated in all arthropods in the same way. One aspect common to all arthropods is that there are several Dscam paralogous genes in one genome. However, the number of paralogues can vary enormously from dozens of genes in chelicerates and centipedes to an average of four in pancrustacean genomes (Brites et al. 2013). In centipedes some Dscam genes are modestly diversified by mutually exclusive alternative splicing affecting one exon of the receptor (Brites et al. 2013). But in pancrustaceans, where the number of Dscam paralogues is small, duplications have generated sets of alternative exons in several sections of the gene (Graveley et al. 2004). These are used in a mutually exclusive alternative splicing process. Following random usage of these multiple exons, the Dscam gene generates somatically a vast repertoire. All splicing events included, the number of isoforms can reach >150,000 in one *Drosophila melanogaster* fly. The combinatorial association of multiple alternative germline-encoded duplicated exons multiplies the number of isoforms considerably (Schmucker et al. 2000) compared to centipedes. The term "combinatorial association" is reminiscent of the combinatorial usage of DNA gene segments during differentiation of the lymphocyte antigen receptors. However, there is a big difference: in the case of Dscam the diversity of isoforms is generated only at the RNA level, whereas for antibody or TCR the transcripts are issued from a gene segment modified at the DNA level and perpetuated in a lymphocyte clone.

The mechanisms by which diversity is created are as follows: In its full-fledged form, observed in pancrustaceans, three kinds of splicing affect the genesis of the extracellular part of the Dscam receptor, in its Ig2, Ig3, and Ig7 domains. The three

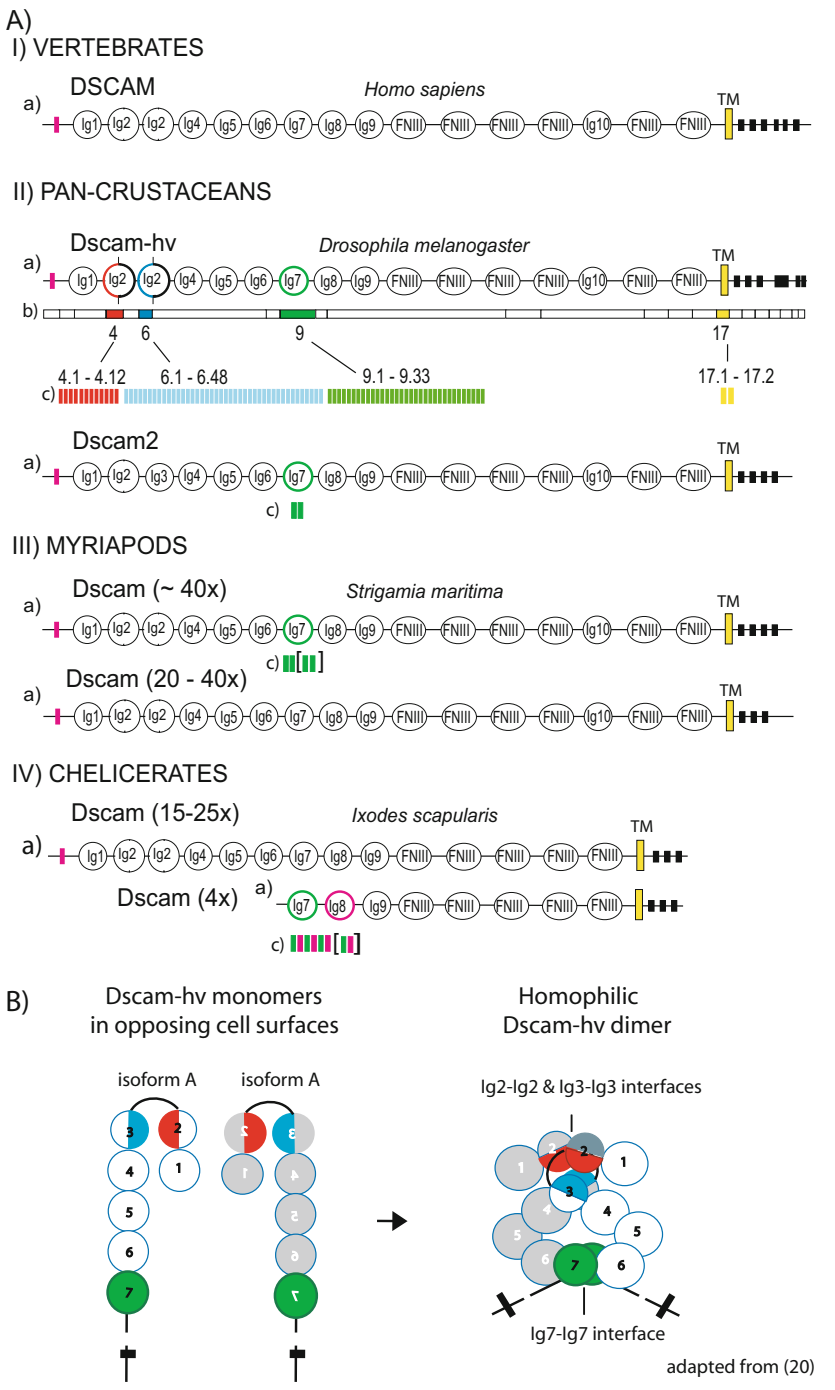


Fig. 1 Dscam in vertebrates and arthropods. (**a**) Examples of Dscam molecules and genes. I) Vertebrates' DSCAM (*Homo sapiens*); II) Pancrustaceans' Dscam (*Drosophila melanogaster* Dscam-hv and Dscam2; this species has two other homologues with no alternative exons); III)

exons (that from now on will be called by their *Drosophila* nomenclature exons 4, 6, and 9, respectively) that potentially undergo splicing, code for the specific parts of the molecule affecting the shape of Dscam epitopes that interact with their ligands. They encode half domains for the first two IgSF domains Ig2 and Ig3 and encode the full domain of Ig7. In pancrustaceans those three exons are duplicated many times (Fig. 1a). Their specific mutually exclusive splicing depends on structural features of the exons and the introns, as a well as on complex enzymatic processes that are not necessarily identical for all the exons. The mechanisms responsible for Ig2 (exon 4) and 3 (exon 6) are relatively well known (especially for the *Drosophila* exon 4) but not the ones responsible for Ig7 (exon 9) splicing (Kreahling and Graveley 2005; Park et al. 2004).

Several lines of experiments have shown that the role of Dscam is primordial for the wiring of the nervous system in several species (Kise and Schmucker 2013), the diversity in insects being an important factor to specify the identity of neurons via differential expression and homologous Dscam recognition. It has also been suggested and partially demonstrated that Dscam could be important for the immune system (Watson et al. 2005; Armitage et al. 2014a; Ng et al. 2014). The suggestion came from the demonstration of a possible role in phagocytosis, from patterns of expression of Dscam in immunocytes, and then from the discovery of soluble forms that could act as antibodies. However, in the case of Dscam the biological significance for diversity is not clear and the problem comes from the mind-sets mentioned above. Many scientists want to compare (read "equate"!) Dscam diversity to antibody diversity, but it may not be that simple a parallel.

Globally, several strategies seem to concur at providing individuals with a great diversity of Dscam receptors, and if we cannot yet envisage with precision the role of this diversity, we can first try to understand how it came about.

Fig. 1 (continued) Myriapods' Dscam (*Strigamia maritima* contains around 40 other Dscam paralogues that have two to four duplicated exons coding for Ig7 and 20–40 paralogues that do not contain duplications); IV) Chelicerates' Dscam (*Ixodes scapularis* contains 15–25 Dscam paralogues plus incomplete paralogues including genes with multiple exons coding for Ig7 and Ig8. a) Protein domains; b) mRNA structure; c) arrays of exons (*bars*) coding for the N-terminal parts of Ig2 and Ig3 and the complete Ig7 domains. The transmembrane domain in *D. melanogaster* is encoded by two alternative exons. The cytoplasmic tails do not represent the true domain structure of *S. maritima* and *I. scapularis* Dscams. (**b**) Model based on Dscam1–8 crystal structure for the conformation of the first seven Ig domains of Dscam in monomers (*right*) and after the formation of dimers (*left*) (Hattori et al. 2008). In monomers, the first four Ig domains form a compact horseshoe structure, whereas the remaining Ig domains have a flexible structure. Upon homophilic binding between identical isoforms (here, isoform A) mediated by the variable regions of Ig2, Ig3, and Ig7 (in *color*), the dimer acquires an S shape

3 The Origin of Dscam's Alternative Splicing Pathways

As stated above, looking at the recent literature one is under the impression that millions of pancrustacean species use a huge Dscam diversity, somatically acquired, to fulfill two purposes: one being the wiring of the nervous system and the second being immunity. However, in some basal arthropods, the centipede *Strigamia maritima*, the splicing of Dscam is simpler than in *Drosophila*. Only one type of alternative splicing exists, the one affecting Ig7, but the number of whole Dscam genes is much higher in the germline (dozens) than in pancrustaceans (maximum six) (Brites et al. 2013).

In the most basal group of all arthropods, the chelicerates, it is possible that mutually exclusive alternative splicing does not exist and that only germline diversity is present, as in the tick *Ixodes scapularis* (Brites et al. 2013). We came to this hypothesis because no structure suggesting a canonical alternative splicing pattern for exon 9 has been observed in the tick *I. scapularis*. Although duplicates of the exons encoding Ig7 and Ig8 have been seen in some of the *Ixodes* Dscam homologues, none so far has been demonstrated to be part of a *bona fide* Dscam gene homologue (Brites et al. 2013). But at least these duplicates show that building cassettes of duplicated exons is possible in the Dscam gene family within chelicerates. Similar cassettes could eventually have been incorporated and used in a full-fledged Dscam-hv system in other species. In addition chelicerates lack several of the genes encoding splicing factors that play a role in splicing Dscam elements (Table 1).

Most recent molecular phylogenies strongly suggest that chelicerates are the most basal group within extant arthropods (Rota-Stabelli et al. 2011; Rota-Stabelli and Telford 2008; Bourlat et al. 2008) and that myriapods, to which centipedes belong, crustaceans, and insects share a common ancestor which is not shared by chelicerates (Rota-Stabelli et al. 2011). Furthermore, myriapods are a basal group to both crustaceans and insects, and the latter group has probably diverged from crustaceans (Cook et al. 2001, 2005; Regier et al. 2005).

The Dscam features described above suggest a stepwise acquisition of the Dscam splicing strategy during arthropod evolution. The question is in which order did the specific features change during evolution: exon structure first or enzyme machinery?

4 The Dscam Loci, from Chelicerates to Insects

The genealogy of the Dscam gene family from vertebrates and invertebrates confirmed that the diversifying form of the gene evolved from an unvarying form before the split of insects and crustaceans (Brites et al. 2013). Can one then find genes related to the ancestors of Dscam-hv among the Dscam genes of chelicerates and centipedes? How can one recognize them? The necessary structural features for

Table 1 Genes involved in alternative splicing in *Drosophila melanogaster* (Park et al. 2004) and the similarities (in percentage) of homologues of those genes in other arthropod species

Genes involved in AS	Percent of similarity in homologous genes		
Drosophila melanogaster	*Ixodes scapularis*	*Strigamia maritima*	*Daphnia pulex*
CG10279 (594)	64.4 % (435)	68 % (433)	63 % (474)
CG 1658 doa (580)	72 % (346)	66.5 % (402)	59.4 % (414)
CC1.3 CG11266 (594)	27.8 % (394)	65 % (355)	54 % (574)
CG12085 Hfp (545)	74 % (215)	58 % (264)	59 % (464)
CG11360 (492)	n. h.	39 % (492)	34.5 % (492)
CG4602 (512)	40 % (266)	52 % (261)	52 % (242)
CG13425 (496)	30 % (144)	42 % (211)	36 % (405)
CG31762 Aret (810)	79 % (93)	65 % (283)	66 % (181)
CG8241 (1242)	81.2 % (960)	81 % (643)	73 % (1242)
CG7437 mub (359)	n.h.	57 % (333)	64 % (359)
CG16941 Prp21 (784)	n.h	53 % (766)	44 % (737)
CG8144 ps (572)	46 % (417)	47 % (564)	63 % (89)
CG7879 (985)	n.h	31 % (568)	50 % (362)
CG8019 (775)	71.9 % (797)	64 % (611)	68 % (510)
CG9998 (416)	67 % (378)	69 % (379)	57 % (461)
CG14641 (418)	64 % (337)	76 % (223)	69 % (299)
CG5931 (2142)	75 % (2135)	73 % (2151)	68 % (2142)
CG3582 (264)	83 % (179)	82 % (179)	74 % (158)
CG10210 (1197)	52 % (368)	42 % (1176)	56 % (635)
CG7269 (424)	36 % (365)	80 % (389)	64 % (399)

BLAST searches were done using *Drosophila* sequences as probes. We tested sequences from *Strigamia maritima*, *Ixodes scapularis*, and *Daphnia pulex* transcripts and scaffolds. In parenthesis is indicated the length in amino acids of the protein in *Drosophila melanogaster* and for the remaining species the length in amino acids over which similarity has been measured. This was done because in several cases the original *Drosophila* gene consists of different domains, only one of which shows prominent homology with the other arthropods' molecules (e.g., *Aret*). Abbreviations: n.h.—no homologue

a full-fledged Dscam alternative splicing are the existence of cassettes of duplicated independent Ig7 exons and of duplicated Ig2 and Ig3 half domain exons.

4.1 Ig7 (Exon 9)

Exon 9 encoding Ig7, in the chelicerates *I. scapularis*, has duplicated, but so far the duplicates have been found in noncanonical Dscam genes and often together with Ig8 coding exon duplicates. Full-length Dscam-related genes differing from the pancrustacean architecture in the membrane proximal part of the molecule (often no Ig10 and one or two less FnIII domains) are present in large numbers (dozens) (Fig. 1a).

In the centipede *Strigamia*, exon 9 has been duplicated in several of the Dscam paralogues, not in very large numbers (up to 4, i.e., fewer than in pancrustaceans) (Fig. 1a) and the alternative exons are used in mutually exclusive alternative splicing (Brites et al. 2013). But the number of germline non-variable genes remains high (several dozens, i.e., the same order of magnitude as in *Ixodes*) (Fig. 1a). Unfortunately (Kreahling and Graveley 2005), nothing is known about the specific mechanism governing the splicing of the exon encoding Ig7.

4.2 Ig2 Ig3 (Epitopes I and II) (Exons 4 and 6)

In addition to Ig7 variability, diversity of pancrustacean Dscam is noticeable in important epitopes made of the interface between the Ig2 and 3 domains, in the regions encoded by the variable alternative half domain exons 4 and 6. These represent stretches of DNA of 165 bp for Ig2 (exon 4) and 117 bp for Ig3 (exon 6), respectively, in *Drosophila*. They contribute to the special region of the distal part of the receptor proposed to display the epitopes responsible for binding: epitope I on one face for homologous interactions and epitope II on the other face for binding to heterologous ligands, including pathogens. Let us see whether any basal arthropod possesses Dscam homologues with at least these features, with or without duplication. Indeed among the many *Ixodes* Dscam-related genes some (Is1, 4, 6, 13, 14, 17 for example) possess exons 4 and 6 with the above mentioned characteristics (example in Fig. 2).

Ixodes scapularis Is14, (Brites et al 2013)
Ig2
VVEQYYEVQVYDEFTIAGNTAVLRCHVPSFVKEDVVVVSWEHKLAQKTEVITTGGRMSVFPSGELHVRRVQPSDASADFRCRTWHR
LTGETKLSSYGRLVVT
Ig3
DLKVNVPPRITNVRSTVVARDGDTVELPCAAQGYPPPKYLWERLPTSDSLSSRRSVLAGSSRFEPSDGSLIIRKVEPEDAGKYLCL
VSNGVGEERATVTLDVQ

Strigamia maritima Sm35, (Brites et al 2013)
Ig2
VLQPYDVYVYDVYVIKGNTAVFRCHVPSFLVDYVKVTSWVRDSAFVIQSTFADVTSYHFSLFYQQDGKYIVMPTGELYVRDVAAND
AMTTFRCQTQHRLTGEVKMSATAGRLF
Ig3
VTEPQGKVQPRVTDSKTSIKANQHDTVVLPCIAQGHPVPAPKWFTKVANGHLLPVYVGDRIHQPNGALVIRDAEVADTGTYVCVIS
NNASSERIETSVAITAPLT

Drosophila melanogaster Dscam-hv
Ig2
VVSQHYEEDIHKAFVIRGNSAILKCDIPSFVADFVNVISWHSDEKENFYPGTEYDGKYLVLPSGELHIREVGPEDGYKSYQCRTKH
RLTGETRLSATKGRLVITE
Ig3
PVGSVSPQLSGNGNQEHITLTRVPKMGSVTLMCPAQAYPVPFFRWYKFIEGTTRKQAVVLNDRVKQVSGTLIIKDAVVEDSGKYLC
VVNNSVGGESVETVLTVT

Fig. 2 Exon and intron boundaries in Dscam genes. In italic exon 4 in Ig2 and 6 in Ig3. The two genes from *I. scapularis* and *S. maritima* are organized in the same way as that of *D. melanogaster* Dscam-hv. An eventual variation due to duplication of those exons would therefore contribute to similar diversified receptor architectures in its horseshoe distal region

Similarly, in the centipede *Strigamia*, several of the multiple Dscam genes harbor a similar organization of exons 4 and 6 but, again, not duplicated.

In addition, many more genes have exon 4 in the proper configuration, but not exon 6.

4.3 Motifs in the Cytoplasmic Tail Associated with Dscam-hv. A Dscam-hv Lineage in Centipedes?

The above features are necessary for permitting an alternative splicing pattern typical of Dscam-hv. Yet many genes that have these features do not use them in this way. Are there then other tags of Dscam-hv? A look at the cytoplasmic tail in Dscam-hv genes and in the various Dscam-related genes of centipedes reveals a sequence motif of high predictive value in identifying Dscam-hv-related genes. We found it systematically associated with Dscam-hv of pancrustaceans and in those *Strigamia* Dscam genes that have been proven to use alternative splicing for exon 9 like *Sm35* (Brites et al. 2013), or that have multiple Ig7 exons plus an exon 4–6 structure with the appropriate abovementioned features (e.g., *Sm546*) (Fig. 3). This motif is a variant of a motif encountered in the roundabout family of receptors expressed in the nervous system, the CC1-like motif (Fig. 3) (Prasad et al. 2007). It is present in all Dscam-hv so far analyzed, but it is missing in *Ixodes* homologues (Brites et al. 2013). This variant motif is present only in arthropods. In *Ixodes Is10* is the closest relative showing the conservation of the PYA motif but not of the flanking residues (Fig. 3). The *Is10* gene does not have a separate exon 6 that would put it in the Dscam-hv lineage.

Not all the *Strigamia* genes that have duplicated exons 9 show this specific sequence. This means that we are able to split a sample of 21 *Strigamia* Dscam genes into two groups, one small group with 2 genes: *Sm35* and *Sm546* that do use Ig7 exon alternative splicing but with only a few alternative Ig7 alternative exons, and the other group with 18 genes, more heterogeneous and without Dscam-hv features. In this "other" group some of the *Strigamia* genes have two exons 9, like the pancrustacean paralogues *Drosophila* Dscam2 and *Apis* Abscam that are specifically expressed in the nervous system and that do use alternative splicing of exon 9 only (Millard et al. 2007). In those the Dscam-hv-specific motif in the cytoplasmic segment is missing. So arthropods, as early as in the centipedes, seem to have selected at least two major types of Dscam genes differing in their signaling moiety and using mutually alternative splicing to generate different isotypes. However, phylogenies do not suggest an obvious link between Dscam2 or Abscam and the *Strigamia* or *Ixodes* Dscam genes. In conclusion the gene architecture suitable for Dscam-hv alternative splicing can be present in ancient arthropod lineages like the centipedes but with a limited number of duplications. It has been proven to work at least in one example: the *Strigamia Dscam* gene *Sm35*; therefore,

Homo sapiens Ro	TTYSRPGQ**PTPYATT**QLIQSNLSNN	124
Dugesia Dscam	NNDDEDEMLV**PYAT**YESLSKPDSST	105
Aplysia Dscam	SFRSDEGNIN**PYAT**YNEIKPTFIPE	139
Strong. Dscam	EPRRHRGLAD**PYAT**FDYHDGSIYPS	126
Is6	LEGRLDYY**PTPYATT**RVTDIDERKL	68
Is23	ECSTSAFFPA**PYATT**HLGTRGPEKR	72
Is10	PRGDPLYFPS**PYATT**HISVYSGDND	69
Is15	PSKDQIYYPS**PYA**LGGREPVLHRQG	69
Sm52294	GSHVDSDEL**TPYAT**ARLADFQEHRR	61
Sm321807	QNSLRRGDVA**PYAT**GHLSDHYQAEE	95
Sm34	TIPRRGADPS**PYAT**SHLTDCHHPEH	94
Sm35	LVKGSSDEI**TPYATT**QLPNFHYGEM	66
Sm53.1	YTQTSLEDVC**PYAT**YRIPESSNKAQ	98
Sm605.1	TREGVHDDAC**PYAT**FQLSENKQNSN	102
Sm91	KRVAPRGEIQ**PYAT**YQLPECCTDAF	91
Sm546	VALGSQEDLA**PY**SAYQLPNYHYGEL	79
***Dro*.Dscam-hv**	RHPGMEDEIC**PYAT**FHLLGFREEMD	162
*Dro.*Dscam2	EGNEYIEDIC**PYAT**FQLNKQTYSES	108
*Dro.*DscamL3	GNESEMYEIS**PYAT**FSVNGGRTGAP	92
***Daph*.Dscam-hv**	LYAGMDDEIC**PYAT**FHLLGFREEMD	151
*Daph.*Dscam2	LSDYAPDQVS**PYA**VFPSLTSSGGKS	104
*Daph.*DscamL3	DNPNQLGDIT**PYAT**FTLKPINGMDT	123
***Pacifast* Dscam-hv**	LRSGGDDEIC**PYAT**FHLLGFREEMD	165
*Dro.*DscamL4	KIPETSEDIS**PYAT**FQLSEAGGNMS	96
Apis Abscam	KIPETAEDIS**PYAT**FQLSEGGGGSL	77

Fig. 3 Occurrence of the CC1-related motif GMDDEICPYATFHLLGFREEMD in Dscam and related genes. In bold the core sequence; on the *left* the species and the Dscam gene name. The numbers on the *right* refer to the amino acid position within the transmembrane domain. Abbreviations: *H. sapiens* Ro: Human roundabout; Duguesia Dscam: *Duguesia japonica* Dscam; Aplysia Dscam: *Aplysia californica* Dscam; *Strong.* Dscam: *Strongylocentrotus purpuratus* Dscam; *Sm: Strigamia maritima* Dscams; Is: *Ixodes scapularis* Dscams; *Pacifast.*: *Pacifastacus leniusculus*; Dro: *Drosophila melanogaster* Dscams; Daph: *Daphnia pulex* Dscams; Apis Abscam: *Apis mellifera* Dscam2

at least some of the splicing factors' machinery known in pancrustaceans must be present and operate in *Strigamia*.

5 The Splicing Factors

We assumed that mutually exclusive alternative splicing of Dscam is only possible thanks to certain splicing factors. Therefore, chelicerates might differ from all the other taxa by lacking more homologues of these factors than the species known to use mutually exclusive alternative splicing. To verify this hypothesis, we surveyed the chelicerates *Ixodes scapularis*, the centipede *Strigamia maritima*, and the crustacean *Daphnia pulex* for the presence of the homologues of known *Drosophila* splicing factors, the involvement of which in Dscam splicing has been demonstrated by RNAi testing (Park et al. 2004). In the present stage of our knowledge, we

could study only the case of the exon 4 and 17 splicing (second transmembrane exon, Fig. 1a).

Thirty-six genes that in *Drosophila* affected specifically exon 4 (Ig2) or exon 4 plus exon 17 splicing were selected and homologies looked for in the genomes of the above species. Strong similarities (>60 % amino acid level) for the majority of them could be detected. In most cases the gene homology extended over the full length of the coding segment (Table 1). In addition other proteins involved in splicing regulation like hp36 and b52 have homologues in crustaceans that seem to function like in *Drosophila* (Chiang et al. 2013).

However, a few genes had no homologues or yielded only short homology stretches. This was the case for *CG11360*, *CG7437 (mub)*, *CG16941 (Prp21)*, and *CG7879* in chelicerates (Table 1). *Prp21* and *mub* in particular were shown to be essential for Dscam Ig2 encoding exons' splicing. Chelicerate is precisely the taxon where no perfect equivalent of alternative exons within a true Dscam gene was found.

In summary, at the present stage of our knowledge of chelicerates, the lack of duplicated exons in the expected place and the absence of some splicing factor homologues suggest that these most basal arthropods, here represented by *Ixodes* might not use mutually exclusive alternative splicing at all to diversify their many germline-encoded Dscam homologues. The pressure for a diversity of Dscam had, until centipedes evolved, only selected germline amplification as a tool to generate multiple Dscam isoforms.

Centipedes' Dscam genes have both germline diversity and the possibility of some limited somatic variation with alternative splicing of Ig7 coding exons. The number of isotypes generated is large but much smaller than in the Dscam-hv of pancrustaceans (Fig. 4) and the nature of the differences (no variation other than the

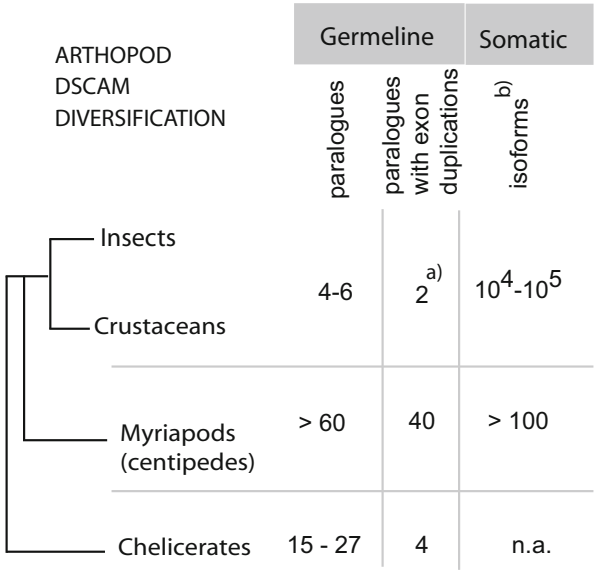


Fig. 4 Putative numbers of Dscam isoforms in different arthropod species caused by either gene and/or exon duplication. a) Dscam-hv and Dscam2; the latter has two Ig7 alternative coding exons. b) Potential number of Dscam isoforms per individual. n.a.: no expression data available

germline one in the horseshoe structure) might have a different meaning in the two groups of animals.

So far the most elaborated form of mutually exclusive alternative splicing exists only in pancrustaceans with minor quantitative variations from one organism to the other (Graveley et al. 2004).

The evolving complexity of the splicing pattern seems to follow the history of arthropods, and the acquisition of the pancrustacean strategy must have been the result of a stepwise process starting with splice variation in Ig7 exons. But what is all this diversity good for? What is the advantage, if any, of generating it somatically?

6 The Resulting Diversities

6.1 How Do Diversities of the Relevant Regions Compare Across Arthropods?

Figure 5 depicts the amino acid sequence diversity of half of Ig2 (exons 4) and Ig3 (exons 6) domains and the complete Ig7 (exons 9) domain from representatives of different arthropod groups. Diversity is expressed as pairwise amino acid distances among paralogue domains for different arthropod species (e.g., a pairwise distance of almost zero indicates that the amino acid composition of two sequences was almost identical). The diversity of Ig2, Ig3, and Ig7 generated by germline gene duplications as in *I. scapularis* seems to be generally similar to the diversity generated in Dscam-hv of pancrustaceans (Fig. 5a). In contrast, *S. maritima* has lower diversity for the two half Ig2 and Ig3 and complete Ig7 domains (Fig. 5a). This could be the result of more recent duplication events (in evolutionary time) and shorter divergence times. However, if Ig5 and FNIII-1 (taken as control domains) are compared, amino acid distances between *Strigamia* and pancrustaceans paralogues are similar to *I. scapularis* (Fig. 5b). Gene conversion is a very common phenomenon among paralogous genes (Nei and Rooney 2005), and it has been suggested to happen among different Dscam paralogues in *S. maritima* (Brites et al. 2013). It could therefore be one reason why some regions of the *S. maritima* Dscam genes are more similar to each other than others. Interestingly, in *Daphnia* and *Drosophila* Dscam-hv, gene conversion does not seem to happen significantly between the exon duplicates (Brites et al. 2011), suggesting selection against amino acid composition homogenization of Dscam-hv variable regions. In other words, selection acts to preserve Dscam diversity obtained via duplication and sequence divergence in pancrustaceans. In centipedes, Dscam diversity also seems advantageous given the high number of putative functional Dscam genes in *S. maritima*. However, in centipedes selection against genetic mechanisms which homogenize the Dscam paralogues does not seem as strong as in pancrustaceans. This indicates that Dscam diversity might not be as essential for *S. maritima* as it is for

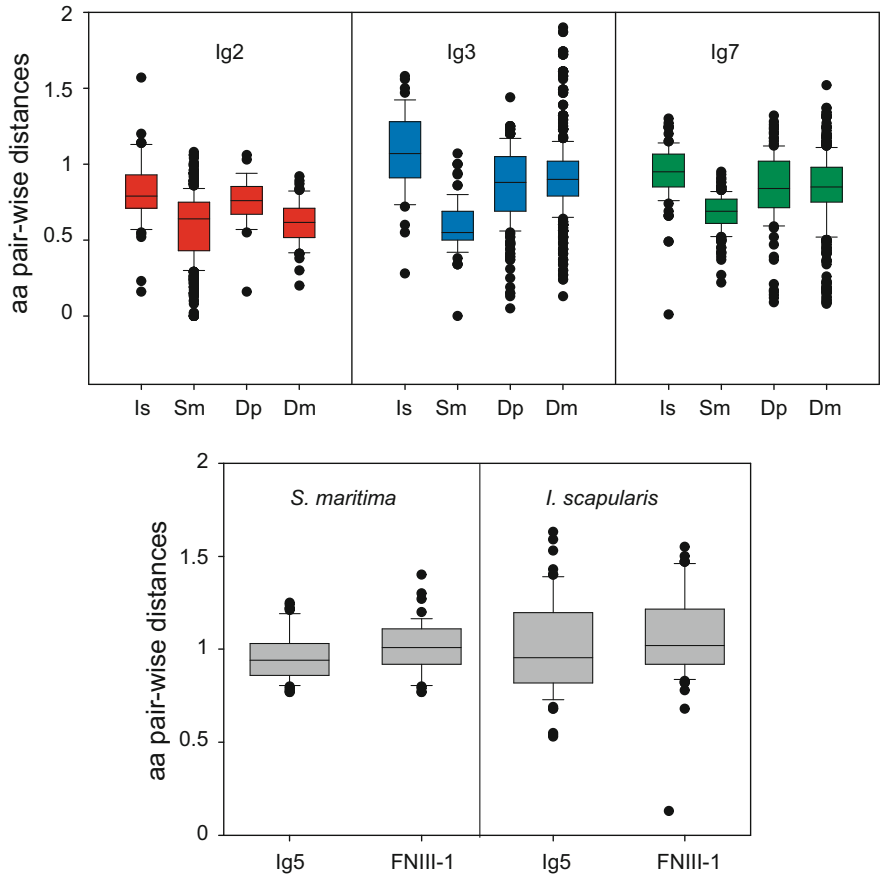


Fig. 5 Amino acid diversity of Dscam Ig2, 3, and 7 domains in species that use or do not use alternative splicing. (**a**) Box plots depicting pairwise amino acid differences among paraloguous Dscam genes (*S. maritima* and *I. scapularis*) or exons (*D. melanogaster* and *D. pulex*) encoding half (exons 4 and 6) of Ig2 and Ig3, respectively, or the complete ig7 (exon 11). For *S. maritima* and *I. scapularis*, only Dscam paralogues in which Ig2 and/or Ig3 are coded by two exons were included; Pairwise amino acid distances were estimated using an amino acid model with Poisson correction as implemented in MEGA (Tamura et al. 2007). Sites with gaps were excluded from pairwise comparisons. The following number of paralogue domains were used: *S. maritima*—exons 4 ($n = 37$), exons 6 ($n = 15$), exons 9 ($n = 20$); *I. scapularis*—exons 4 ($n = 11$), exons 6 ($n = 10$), exons 9 ($n = 15$); *D. pulex*—exons 4 ($n = 8$), exons 6 ($n = 24$), exons 9 ($n = 16$); *D. melanogaster*—exons 4 ($n = 12$), exons 6 ($n = 48$), exons 9 ($n = 33$). (**b**) *Box plots* depicting pairwise amino acid differences among Dscam Ig5 and FNIII-1 domains from the selected paralogues depicted in (**a**). The following numbers of paralogues were used: *S. maritima* exons 4 ($n = 11$), FNIII-1 ($n = 10$); *I. scapularis* Ig5 ($n = 13$); FNIII-1 ($n = 12$). Abbreviations: Is: *Ixodes scapularis*; Sm: *Strigamia maritima*, Dp: *Daphnia pulex*; Dm: *Drosophila melanogaster*

pancrustaceans, suggesting that the function of that diversity differs between both groups of animals. The myriapod Dscams might represent an evolutionary intermediary step preceding the evolution of Dscam-hv system in the ancestor of all pancrustaceans. Interestingly, once a Dscam-hv has evolved in that ancestor, most other Dscam paralogues were lost, suggesting that they were functionally not relevant anymore or maybe their regulation was even too costly.

6.2 *Does Dscam Diversity Expression Vary in the Tissues of the Different Species?*

One thing remains mysterious given the randomness of exon usage during the alternative splicing: the splicing pattern differs in different tissues whether in crustaceans or insects. The representation of the diverse exons is not the same in the nervous system and in cells of the immune system. This is particularly clear in *Drosophila* where the exon usage has been studied in detail for three tissues: nervous system, fat bodies and hemocytes (Watson et al. 2005), and S2 cells of the hematopoietic lineage (Neves et al. 2004). Even within the fruit fly nervous system many different cell types express different Dscam splicing repertoires. *Daphnia* seems to follow a similar pattern of expression with big differences of repertoires between hemocytes and nervous system (Brites et al. 2008). However, the difference between hematopoietic tissue and nervous tissue might turn out to be only apparent. As suggested recently by decomposition analysis of the brain dataset, it is very likely that in some distinct types of neurons, the Dscam isoform repertoire might be similarly small as that observed in S2 cells, so the expression pattern of those cells might not be unique. In addition within pancrustaceans, in shrimps, the expression is far from being restricted to hemocytes and nervous system; message can be detected in hepatopancreas, brain, stomach, testis, abdominal nerve, gill, hemocytes, and midgut, but repertoires have not been analyzed in detail (Chou et al. 2009). Often diversity is measured in samples too small to give a good impression of the whole repertoire.

Transcripts from Dscam have been recovered from the centipede *Strigamia* hemocytes, but repertoire comparisons have not been made with other tissues even though expression was observed in the brain and other tissues.

At the single cell level in vivo, the average number of different Dscam molecules expressed per cell is not known. In *Drosophila* S2 cultured cells, it is two dozen isotypes (Neves et al. 2004), but one does not have a good idea of the density nor of the total number of receptors on the cell surface. This could prove a major factor in speculating on the usefulness of a specific Dscam isoform as a receptor. What is the threshold level?

Independent exon usage seems to be the rule during the splicing. The expression pattern in the uniform S2 population of cells has been studied and the results demonstrated that the alternative splicing between different exon clusters is

independent, a fact also implied when Neves et al. (2004) did not find specific exon 4 and 6 alternatives associated with either of the two exon 9s (Neves et al. 2004; Sun et al. 2013). Nothing is known about the expression of *Ixodes* and *Strigamia* Dscam in terms of tissue specificity or repertoire nature.

7 The Meaning of Dscam Diversity for the Immune System

Many papers in arthropods attribute to Dscam a role in immunity. It has been proposed to be a phagocytic receptor as an integral membrane protein on hemocyte surface. It has also been suggested that it worked as a soluble effector. The soluble form could be due to two very different processes: passive shedding from the membrane or active secretion in the hemolymph. Dscam is supposed to bind pathogens or autologous ligands (following apoptosis) and the diversity of its repertoire is often compared to that of the antibodies produced by adaptive immune systems of vertebrates (although it is still orders of magnitude lower). Yet in the Dscam-hv case, diversity is due to a process at the RNA level and in the antibody case, to rearrangement, gene conversion, and somatic mutation at the DNA level (Armitage et al. 2014a; Ng et al. 2014). So the analogy stops pretty fast.

We think that there are some inconsistencies in many of these immunological experiments and their interpretations, and we think also that many theoretical considerations linked to the nature of Dscam diversity are not taken enough into account when interpreting Dscam as a pathogen receptor. This is not to say that diversified Dscam does not play a role in the immune system in those species, but it may be a different role.

7.1 Dscam as an Immuno-Receptor/Effector

Results obtained in phagocytosis experiments in *Drosophila* suggested initially that Dscam-hv could be a phagocytosis receptor. Experimentally, the capacity of hemocytes to engulf bacteria was reduced in cells lacking Dscam on their surface following a specific RNAi treatment or exposure to an anti-Dscam-hv antibody. In addition, engineering of Dscam molecules in vitro by varying exon combinations created forms that had different specificities of binding. This reinforced the idea that Dscam-hv molecules could bind heterologous parasite ligands, arguing therefore strongly for a role in immunity (Watson et al. 2005). What we know of the signaling pathways following Dscam-hv engagement is indeed compatible with a role in phagocytosis. In S2 cells Dscam-hv appears to be at the head of a cascade of interactions implicating Wasp, a known regulator of actin polymerization in the cytoskeleton, the modulation of which is probably crucial both in axonal extension and in mobility, in endocytosis, and therefore in phagocytosis by hemocytes (Worby et al. 2001; Li and Kun-Liang 2004). This phase of Dscam signaling

pathway probably depends on the motifs present in the cytoplasmic tail. And precisely on this subject variation of Dscam at the cytoplasmic tail level was also demonstrated in *Daphnia* (Brites et al. 2008), in *Drosophila*, and in shrimp (Chou et al. 2011), due again to different splicing possibilities of the cytoplasmic tail exons which modify the distribution of tyrosine residues, motifs that may be involved in modulation of endocytosis and therefore in phagocytosis (Yu et al. 2009). Along those lines, isoforms with alternatively acquired endocytosis motif are more abundantly expressed in the hemocytes of crustaceans (Chou et al. 2011). A role as a phagocytic receptor could also be compatible with an important function, not strictly immunological, during development: engulfment of autologous cells that die by apoptosis and the recognition of which might require a fine-tuned mechanism where Dscam could be implicated and its diversity necessary for modulating recognition capacity. Recognition could be of a different nature there; it could be a Dscam–Dscam homologous interaction because Dscam might be expressed on the autologous dead cells. The situation could therefore be similar to that found in the nervous system with Dscam arrays specifying the identity of certain autologous cell populations to be targeted during remodeling. Indeed, in remodeling studies during development, 35 genes known to be involved in phagocytosis are significantly upregulated in hemocytes. Among them, several of these genes encode putative phagocytic receptors including *Dscam* (Regan et al. 2013).

Dscam has also been reported to be upregulated following introduction of some pathogens with some degree of specificity [reviews in Armitage et al. (2014a) and Ng et al. (2014)]. In *Anopheles gambiae* Dscam was reported to be upregulated in response to infection by *Plasmodium*, apparently producing "pathogen challenge"--specific splice form repertoires. Like in *Drosophila* experiments, transient silencing of Dscam compromised the mosquito's resistance to infections with *Plasmodium* but also with bacteria. It was concluded that Dscam mediates phagocytosis of bacteria and most remarkably in a splice form-specific manner. Dscam-hv was therefore considered as an important determinant for resistance to infection in this species (Dong et al. 2006). But no mechanism was proposed to explain how this was made possible. Significant increases in Dscam receptor diversity were observed in parasite-exposed mosquitoes, and a cluster of AgDscam exon 4 variants that become especially common during *Plasmodium* invasion was identified (Smith et al. 2011). However, an increase in diversity is not what might be the most interesting form of response; one would rather see the amplification of a given useful variant with some specificity. But we may simply not understand the mode of action of Dscam.

This is perhaps why in 2014 new avenues have been explored within the *Anopheles* model, and why different mechanisms have been suggested, that place Dscam perhaps more downstream in the response than originally thought, more like an effector perhaps? It has now been claimed (Dong et al. 2012) that "in the *Anopheles* malaria vector, IMD and Toll pathways mediate species-specific defenses against *Plasmodium* and bacteria through the transcriptional regulation of splicing factors Caper and IRSF1 that, in turn, determine the production of pathogen-specific splice variant repertoires of the hypervariable pattern recognition

receptor AgDscam." A stunning observation in these studies was the effect of some Dscam isoform knock-in experiments on the composition of the microflora. Some exon combinations would reduce by orders of magnitude the amount of proliferation of the flora following the blood meal (Dong et al. 2012). Another result pointing also in the direction of a link between Dscam expression and IMD and Toll pathway is the observation in the honeybee of a close relationship between Dscam, Imd, and MyD88 induction in the midgut-ileum tissues infected with *N. ceranae* (Schwarz and Evans 2013).

In conclusion, certain things happen at the Dscam locus level following introduction of parasites. This issue has been reviewed by Ng et al. (2014) and recently by Armitage and Kurtz (Armitage et al. 2014a) who were especially interested in asking what was the evidence for an involvement of Dscam in pancrustacean immune-specific memory and who concluded rightly:

"We think that crucial experiments to address whether it plays a role in specificity upon secondary encounter with a pathogen still remain to be done."

We might extend this criticism even further in the next section.

7.2 Problems with Experiments

Many of the experiments dealing with Dscam as a parasite receptor have not been independently repeated or have not dealt with certain theoretical implications. Also we feel that not enough connections have been made between the knowledge about Dscam properties derived from the nervous system studies and the putative role of Dscam in immunity.

7.2.1 Role in Phagocytosis

There has been no direct confirmation of a role in phagocytosis. Hemocytes from *Drosophila* carrying a Dscam-loss-of-function mutation were still able to recognize and cross-link bacteria on their surface with equal efficiency to their wild-type counterparts. This suggested that Dscam is dispensable for recognition of *Escherichia coli* by embryonic hemocytes (Vlisidou et al. 2009). If Dscam works as an opsonin, then the receptor for it on the surface of the phagocytic cell has not yet been discovered, unless it is Dscam itself.

7.2.2 The Receptor Upregulation Problem

Many will agree that it is possible to observe induction of Dscam expression after some "antigenic" exposure whether at the receptor level or perhaps as soluble effector. But why would a receptor be upregulated? To increase efficiency of stimulation with a more active phagocytosis? Then the receptor has to be specific,

and only the relevant receptor or a small population of cross-reacting ones has to be upregulated. This upregulation is easier to understand if it concerns Dscam as an effector that would be produced in larger amounts like an antibody, but again it makes sense essentially if the Dscam molecule interacts specifically with its ligand. This brings us to an issue that has been neglected in most studies: the concentration of each isoform either on the cell surface in the hemocyte population or in the biological fluids.

7.2.3 The Concentration of Each Dscam Isoform Problem

If one assumes that one variant of Dscam has a better avidity for its ligand, the question arises then whether this advantage can manifest itself and be useful. It is difficult to imagine the use of a single variant diluted in the middle of thousands other forms, whether on a cell surface as receptor or in the hemolymph as a soluble form. As such, there is a need for an amplification step. With alternative splicing that occurs at the RNA level cell proliferation is of no use, except by providing more possible sources of the protein but without specificity. Since apparently a single cell expresses more than one Dscam isotype (a profound difference from any lymphocyte) the best that can be produced is a shotgun spray of unrelated Dscam molecules. Perhaps an increased abundance of receptors favors more phagocytosis. It could help the cell reach a threshold of sensitivity by increasing receptor density and thus cross-linking capacities.

There is a lot of confusion around Dscam analogy with antibodies. Dscam is not produced monoclonally by hemocytes. Hemocytes do not proliferate after antigen injection. Dscam upregulation can therefore only be due to a control of transcription.

Contrary to what has been proposed (Boehm 2007), there is no clonal amplification of the cells producing Dscam. In fact, no proliferation follows pathogen introduction in any invertebrate immune response so far reported, except perhaps in tunicates (Raftos and Cooper 1991). The changes occur and are limited to the RNA level. All Dscam responses that have been observed are upregulations of protein synthesis. Some models have been proposed to explain the stabilization of the production of a given isotype, but in terms of experiments, nothing is available and "it remains a puzzle how a specific recognition of a pathogenic antigen by one Dscam isoform may lead to changes in alternative splicing or splice form production of that same isoform" (Schmucker and Chen 2009).

It would help to know the affinity with which an isoform of Dscam binds to its heterologous ligand. What if the binding to pathogen is a by-product that may occasionally affect phagocytosis and what if the main point of Dscam expression on hemocyte is simply its diversity, just like for the nervous system?

All the results dealing with variability are plagued by the huge diversity of the repertoire, by the small amount of data concerning putatively relevant isoforms and the lack of precise kinetics and controls. As such, one does not have a clear picture of the repertoire in nature, of its evolution with time, or of its stability or instability

after encounter with parasites. We know that these repertoires differ from one tissue to the other. This implies selection after random production of isoforms or an upstream tissue-specific control of expression, two steps about which we know nothing.

7.2.4 Induction, Kinetics, Variations and Biochemical Inconsistencies

There are many differences among and within models with respect to Dscam expression after an encounter with an antigen. The bases are not understood.

Several groups have followed the modulation of Dscam expression after the introduction of a parasite in the host [reviewed in Armitage et al. (2014a)].

In crustaceans several papers do support the idea that Dscam can be upregulated early in response to some bacterial challenges or virus challenges (Armitage et al. 2014a). In the honeybee significant Dscam mRNA production was seen 6, 24, and 72 h, transiently after exposure to bacteria, with fluctuation (Schwarz and Evans 2013). In the crayfish the kinetics are different, with an increase 5 days after virus challenge and continuing for a long period (month) (Watthanasurorot et al. 2011).

Many of these studies rely on anti-Dscam antibodies to monitor expression, which would be valid if all the forms detected on western blots were proven to be Dscam. The discrepancy between measured molecular weight and expected molecular weights can be explained only if many posttranslational modifications take place. In some cases the measured molecular weights are so imprecise that we cannot decide whether Dscam or cross-reactants are being detected.

In *Drosophila*, immunoprecipitation data correlated well with the expected molecular weight of the Dscam receptor (>220 kDa) (Watson et al. 2005). In some instances soluble forms of Dscam have been reported. However, these soluble forms detected with rabbit antisera on western blots, either in Drosophila or in crustaceans, are much shorter than they should be after elimination of the transmembrane segments and cytoplasmic tails, which casts some doubt on the interpretation. In *Drosophila* the size measured was *ca.* 160 kDa size compatible with a fragment cut after Ig10 and excluding therefore the last two FNIII and transmembrane domains and cytoplasmic tail regions and ignoring possible addition due to carbohydrates.

In the crab *Eriocheir* hemocytes, where soluble Dscam were reported (Jin et al. 2013), the MW measured is even closer to 140 kDa which would be only possible following the elimination of a segment of Dscam starting at the Ig10 domain level. In all these cases, if the molecules brought down were actually Dscam modified by posttranslational modifications or chemical breakdown products or cross-reacting molecules should be verified by microsequencing.

7.2.5 Increased Diversity Following Stimulation?

In *Anopheles*, it was reported that infection did not induce an overall increase of Dscam transcripts but rather profound changes of its variable exon representation through alternative splicing (Dong et al. 2006). This was never repeated. Further clouding the issue, there are now repeated reports that Dscam diversity simply increases upon stimulation, or in other words, that there is no specific amplification. This would imply that the whole population of Dscam variants is involved in immunity, but then in a more mysterious way. Whereas the data indicate that Dscam diversity increases with *P. falciparum* exposure, they do not suggest that Dscam diversity rises further in response to increased parasite diversity (Smith et al. 2011). A recent work in *Drosophila* following exposure of the whole organism or of cell lines to *E. coli* suggested that there were changes neither in expression nor in splicing patterns of Dscam-hv (Armitage et al. 2014b).

What would be anyway the meaning of an increased diversity upon stimulation? With more kinds of isoforms produced, each one will be present at a lower specific concentration. Then the issue of its usefulness comes up, unless what matters in Dscam response and role is simply its diversity (whatever it is), like for the nervous system, but in ways not yet elucidated in the immune system. Perhaps the expression of Dscam diversity results in a population of hemocytes all unique in their molecular profile, in the same way as it has been suggested for lymphocytes (Lefkovits 2013). Finding out whether specific forms are amplified or not and whether diversity increases upon stimulation are key issues to be solved and sorted out.

8 Features that Do Not Fit with a Simple Antigen Recognizing Immunoreceptor or Immunoeffector Role

We do not know enough about Dscam expression in the immunological context. Many features that are associated with a function of Dscam in the nervous system do not have a justification in immunity.

8.1 The Alternative Transmembrane Segments

The role of alternative transmembrane segments is well understood within the nervous system context, with their differential expression in subfamilies of neurons. The two transmembrane segments seem to play a role during dendrite elaboration and axonal arborization (Shi et al. 2007), in other words a role quite specific of the nervous system. What is the pattern of expression of the transmembrane domains in hemocytes? No one knows.

8.2 Endodomain Diversity

Diversity of the Dscam endomains is also easily linked to neuronal development but not to immunity. In Drosophila besides 19,008 possible ectodomains and the two alternative transmembrane segments, Dscam may carry cytoplasmic regions containing or lacking exons 19 and 23. The resulting Dscam molecules (e.g., with or without exon 19) are utilized to govern different stage-specific neuronal morphogenetic processes, possibly due to differences in protein targeting.

8.3 Specificity of Expression Patterns

The pattern of expression is not that of an immunoreceptor and the S2 cells' expression pattern once thought to be typical of hemocytes might be encountered in some subpopulation of nervous cells (Sun et al. 2013). In addition, the demonstrated instability of Dscam expression with protein variation during the life of the cell and the difference seen between Dscam mRNA and the proteins produced are also difficult to reconcile with a simple immunoeffector role in *Drosophila* (Yu et al. 2009), simply because no stable and continuous production of a single useful variant seems possible. In fact, expression of a large diversity of Dscam repertoire is more important in the nervous system than the actual sequence of each isoform. Perhaps an analogous situation exists in the immune system.

Could the role of Dscam in immunity be a side product, one that would still profit from a somatic generation of diversity? Yet it is difficult not to think about an immunoeffector role when looking at shrimps and species with soluble forms in greater abundance than in *Drososphila*, and with splice forms that produce soluble forms like the one predicted in *Ixodes scapularis Is21 Dscam* gene, precisely one that yielded an EST only coding for Dscam ectodomains.

8.4 The Genesis of Splice Variants

In neurons Dscam splicing does not occur in a cell-type-specific fashion. Cells sharing the same anatomical location in different individuals can express different exon 4 variants, and the splicing pattern in a given neuron can change over time. Splicing is probabilistic (Miura et al. 2013). If the mechanism is similar in immune cells, it is difficult to reconcile with a role in immunity demanding a specific amplification of a given adequate receptor of effector. One has to imagine a process at the RNA level that will fix the production of one single variant; otherwise no amplification of the useful product can be obtained. The role might not be analogous to that of an antibody.

Many questions remain therefore to be answered and precision provided.

9 Other Possible Roles for Dscam in the Immune System?

9.1 Specifying Hemocyte Identity

Given the many uncertainties encountered when attributing an immunological role to Dscam and in particular to its somatically acquired diversity, one can perhaps look at other functions that would profit from the expression of a large somatic random diversity of receptors by hemocytes. The minimal hypothesis is that the diversities are used in both nervous and immune system for the same purpose: specifying cell identities. Hemocytes migrate, assemble, and circulate in the hemolymph, and they do aggregate under some circumstances (Wood and Jacinto 2007). This behavior may need some cellular individualization and recognition among the cells moving around. Dscam could therefore be used like for the specification of neuronal identities and again the somatic adaptation would simply be a way to save on genetic material (Pasquier 2005).

9.2 A Role for Hemocytes in the Development of the Brain

Another developmental process that may require Dscam expression by hemocytes is the control of the nervous system condensation during development. Condensation is a process whereby a tissue undergoes a coordinated decrease in size and increase in cellular density during development. Two major events coincide with brain condensation during embryogenesis: the local deposition of extracellular matrix by hemocytes, and the onset of central nervous system activity. Preventing hemocyte migration inhibits condensation (Olofsson and Page 2005). Dscam on hemocytes could therefore be useful for these remodeling steps.

In the two above hypotheses, the problem of the relative concentration of each isoform does not exist like in the "antibody" interpretation, because what matters is simply diversity. And the fact that the diversity is generated randomly at the RNA level has no lastingly adverse consequences because one does not need the cell progeny to have the same diversity but just some diversity. One problem though remains with this hypothesis: how does one explain the qualitative differences among Dscam repertoires in the different tissues?

9.3 A Role for the Nervous System Controlling Hematopoiesis

In *Drosophila*, hemocytes rely on the peripheral nervous system to survive, to proliferate, and to migrate. Important interactions take place in special

microenvironments (niches) where contact must take place among cell expressing Dscam (Makhijani and Bruckner 2012).

10 Conclusion and Further Questions: Always Mind-Sets

The role of Dscam in arthropod immunity is far from being fully elucidated. Many hypotheses, not necessarily mutually exclusive, can still be put forward. Yet diversity is at the center of all of them. It is a constant of Dscam biology in arthropods with a convergence of solutions that lead to large or immense repertoires of Dscam molecules. The features suggesting a stepwise acquisition of the Dscam splicing strategy during evolution of arthropods shed a new light on the interpretations concerning the recognition capacities of Dscam, so far only derived from *Drosophila*.

It has been argued (Sawaya et al. 2008) that in *Drosophila* Dscam "each variable Ig domain appears to have very weak homophilic self-binding, which is insufficient alone to elicit functional recognition. It is only when all three variable regions are identical between a Dscam pair that productive engagement can take place." If that is the case one may wonder about the meaning of the other forms of Dscam where "somatic" alternative splicing affects Ig7 only like in centipedes, or like some paralogues in insects [e.g., Dscam2 of *Drosophila* (Lah et al. 2014)]. The expression of Dscam isoforms in *Ixodes* and perhaps even those in *Strigamia* might resemble more the expression of the isoforms of Dscam in human than the multiple isoforms of Dscam-hv in pancrustaceans. Comparing Dscam functions and expression in these two taxons might prove useful. The introduction of mutually exclusive splicing, with combinatorial usage of 3 alternative exons, caused an abrupt change of scale in Dscam repertoire size as the evolution of arthropods reaches the pancrustaceans, and we are not sure whether the selection pressures that were exerted on this Dscam-hv gene are the same as the ones acting on Dscam of centipedes or chelicerates. Too much is missing in our information on expression in these two taxa so that we can formulate a decent hypothesis. Diversity "per se" remains perhaps a reasonable hypothesis to explain the randomness of the generation of the repertoire, its lack of inheritability, and its usages in different tissues where the individualization of cells and various forms of recognition have to take place.

The introduction of alternative splicing raises also the question of the "cost" of Dscam expression and of the "cost" of genetic load. In comparing the species using germline diversity to those using alternative splicing within one gene, one realizes that a major difference may exist at the level of the control of gene expression. Dscam-hv is presumably under the control of one single promoter, whereas each independent Dscam in *Ixodes* or *Strigamia* must have its own promoter, making perhaps simultaneous coexpression more complicated. In terms of cost, it might be "cheaper" to regulate the expression of alternatively spliced isoforms from one gene than to regulate the expression of dozens of independent genes scattered in the

genome, even if some form clusters. Therefore, it is probably not by chance that, so far, none of the pancrustaceans species has kept a large number of Dscam paralogues, six at the maximum.

The evolutionary pressures have led to organisms producing as diverse as possible Dscam repertoires. Several solutions were used, first the duplication and maintenance of many Dscam genes in basal arthropods, and second, combinatorial association of exons used in alternative splicing in pancrustaceans that created a potential diversity of receptors larger than the number of genes encoding all Dscam segments. To this large individual diversity polymorphism adds variation within populations (Brites et al. 2011). What is the significance of this polymorphism when so much diversity is generated somatically already? The analysis of orthologous exons in *Drosophila* and in *Daphnia* revealed an excess of nonsynonymous polymorphisms in the epitopes putatively involved in pathogen binding, over what would be expected if the sequence of those epitopes evolved neutrally. This may be a sign of balancing selection. Indeed, in *D. melanogaster* the same derived nonsynonymous alleles segregate in several populations around the world. Yet, other hallmarks of balancing selection were not found. Hence, we cannot rule out that the excess of nonsynonymous polymorphisms is caused by segregating slightly deleterious alleles, thus potentially indicating reduced selective constraints in the putative pathogen binding epitopes of Dscam. It would be interesting in this context to evaluate the polymorphism of the Dscam gene in the species that use moderately or do not use alternative splicing, that is, *Ixodes* and *Strigamia*.

To conclude: perhaps we are prevented from conceiving the real role of Dscam in arthropod immunity because we might still be too much under the influence of the "somatically acquired repertoire" mind-set, a repertoire that has to be able to recognize the environmental stressors and pathogens. Perhaps it would be better to go back to concepts involving network implicating autologous determinants. But perhaps then we fall under the mind-set of neurobiologists! O Scylla! O Charybdis!

Acknowledgments We thank Prof. I. Lefkovits and for his critical reading of the manuscript.

References

Armitage SA, Peuss R, Kurtz J (2014a) Dscam and pancrustacean immune memory – a review of the evidence. Dev Comp Immunol 48(2):315–23

Armitage SA, Sun W, You X, Kurtz J, Schmucker D, Chen W (2014b) Quantitative profiling of *Drosophila melanogaster* Dscam1 isoforms reveals no changes in splicing after bacterial exposure. PLoS One 9(10), e108660

Billingham RE, Brent L, Medawar PB (1954) Quantitative studies on tissue transplantation immunity. II. The origin, strength and duration of actively and adoptively acquired immunity. Proc R Soc Lond B Biol Sci 143(910):58–80

Boehm T (2007) Two in one: dual function of an invertebrate antigen receptor. Nat Immunol 8(10):1031–1033

Boman HG, Nilsson I, Rasmuson B (1972) Inducible antibacterial defence system in *Drosophila*. Nature 237(5352):232–235

Bourlat SJ, Nielsen C, Economou AD, Telford MJ (2008) Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. Mol Phylogenet Evol 49(1):23–31

Brites D, McTaggart S, Morris K, Anderson J, Thomas K, Colson I, Fabbro T, Little TJ, Ebert D, Du Pasquier L (2008) The Dscam homologue of the crustacean *Daphnia* is diversified by alternative splicing like in insects. Mol Biol Evol 25(7):1429–1439

Brites D, Encinas-Viso F, Ebert D, Du Pasquier L, Haag CR (2011) Population genetics of duplicated alternatively spliced exons of Dscam in *Daphnia* and *Drosophila*. PLoS One 6(12), e27947. doi:10.1371/journal.pone.0027947

Brites D, Brena C, Ebert D, Du Pasquier L (2013) More than one way to produce protein diversity: duplication and limited alternative splicing of an adhesion molecule gene in basal arthropods. Evolution 67(10):2999–3011

Chiang YA, Hung HY, Lee CW, Huang YT, Wang HC (2013) Shrimp Dscam and its cytoplasmic tail splicing activator serine/arginine (SR)-rich protein B52 were both induced after white spot syndrome virus challenge. Fish Shellfish Immunol 34(1):209–219

Chou PH, Chang HS, Chen IT, Lin HY, Chen YM, Yang HL, Wang KCHC (2009) The putative invertebrate adaptive immune protein *Litopenaeus vannamei* Dscam (LvDscam) is the first reported Dscam to lack a transmembrane domain and cytoplasmic tail. Dev Comp Immunol 33(12):1258–1267

Chou PH, Chang HS, Chen IT, Lee CW, Hung HY, Han-Ching Wang KC (2011) *Penaeus monodon* Dscam (PmDscam) has a highly diverse cytoplasmic tail and is the first membrane-bound shrimp Dscam to be reported. Fish Shellfish Immunol 30(4-5):1109–1123

Cook CE, Smith ML, Telford MJ, Bastianello A, Akam M (2001) Hox genes and the phylogeny of the arthropods. Curr Biol 11(10):759–763

Cook CE, Yue Q, Akam M (2005) Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. Proc Biol Sci 272(1569):1295–1304

Dong Y, Taylor HE, Dimopoulos G (2006) AgDdscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. PLoS Biol 4(7), e229. doi:10.1371/journal.pbio.0040229

Dong Y, Cirimotich CM, Pike A, Chandra R, Dimopoulos G (2012) Anopheles NF-kappaB-regulated splicing factors direct pathogen-specific repertoires of the hypervariable pattern recognition receptor AgDscam. Cell Host Microbe 12(4):521–530

Graveley B, Amardeep K, Dorian G, Lawrence ZS, Lee R, c. CJ (2004) The organization and evolution of the Dipteran and Hymenopteran Down syndrome cell adhesion molecule (*Dscam*) genes. RNA 10(10):1499–1506

Hattori D, Millard SS, Wojtowicz WM, Zipursky SL (2008) Dscam-mediated cell recognition regulates neural circuit formation. Annu Rev Cell Dev Biol 24:597–620

Hauton C, Smith VJ (2007) Adaptive immunity in invertebrates: a straw house without a mechanistic foundation. Bioessays 29(11):1138–1146

Hildemann WH, Johnson IS, Jokiel PL (1979) Immunocompetence in the lowest metazoan phylum: transplantation immunity in sponges. Science 204(4391):420–422

Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327(5962):167–170

Hozumi N, Tonegawa S (1976) Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc Natl Acad Sci USA 73(10):3628–3632

Jerne NK (1970) General discussion. In: Chateaureynaud-Duprat P, Du Pasquier L, Valembois P (eds) Colloque sur les réactions immunitaires chez les invertébrés, Bordeaux. Université de Bordeaux I, pp 30–41

Jin XK, Li WW, Wu MH, Guo XN, Li S, Yu AQ, Zhu YT, He L, Wang Q (2013) Immunoglobulin superfamily protein Dscam exhibited molecular diversity by alternative splicing in hemocytes of crustacean, Eriocheir sinensis. Fish Shellfish Immunol 35(3):900–909

Kise Y, Schmucker D (2013) Role of self-avoidance in neuronal wiring. Curr Opin Neurobiol 23(6):983–989

Kreahling JM, Graveley B (2005) The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila Dscam* pre-mRNA. Mol Cell Biol 25(23):10251–10260

Lah GJ, Li JS, Millard SS (2014) Cell-specific alternative splicing of *Drosophila* dscam2 is crucial for proper neuronal wiring. Neuron 83(6):1376–1388

Lai Wing Sun K, Correia JP, Kennedy TE (2011) Netrins: versatile extracellular cues with diverse functions. Development 138(11):2153–2169

Lefkovits I (2013) Alacrity of cells engaged in the immune response. Scand J Immunol 77(1):1–12

Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA (1996) The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. Cell 86(6):973–983

Li W, Kun-Liang G (2004) The Down Syndrome Cell Adhesion Molecule (DSCAM) interacts with and activates Pak. J Biol Chem 279(31):32824–32831

Loker ES, Adema CM, Zhang SM, Kepler TB (2004) Invertebrate immune systems–not homogeneous, not simple, not well understood. Immunol Rev 198:10–24

Ly A, Nikolaev A, Suresh G, Zheng Y, Tessier-Lavigne M, Stein E (2008) DSCAM is a netrin receptor that collaborates with DCC in mediating turning responses to netrin-1. Cell 133(7): 1241–1254.

Makhijani K, Bruckner K (2012) Of blood cells and the nervous system: hematopoiesis in the *Drosophila* larva. Fly (Austin) 6(4):254–260

Medzhitov R, Preston-Hurlburt P, Janeway CA Jr (1997) A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. Nature 388(6640):394–397

Meijers R, Puettmann-Holgado R, Skiniotis G, Liu J-h, Walz T, Wang J-h, Schmucker D (2007) Structural basis of Dscam isoform specificity. Nature 449:487–491

Millard SS, Flanagan JJ, Pappu KS, Wu W, Zipursky SL (2007) Dscam2 mediates axonal tiling in the *Drosophila* visual system. Nature 447(7145):720–U714

Miura SK, Martins A, Zhang KX, Graveley BR, Zipursky SL (2013) Probabilistic splicing of Dscam1 establishes identity at the level of single neurons. Cell 155(5):1166–1177

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152

Neves G, Zucker J, Daly M, Chess A (2004) Stochastic yet biased expression of multiple Dscam splice variants by individual cells. Nat Genet 36:240–246

Ng TH, Chiang YA, Yeh YC, Wang HC (2014) Review of Dscam-mediated immunity in shrimp and other arthropods. Dev Comp Immunol 46(2):129–138

Olofsson B, Page DT (2005) Condensation of the central nervous system in embryonic *Drosophila* is inhibited by blocking hemocyte migration or neural activity. Dev Biol 279(1):233–243

Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR (2004) Identification of alternative splicing regulators by RNA interference in Drosophila. Proc Natl Acad Sci USA 101(45): 15974–15979

Pasquier LD (2005) Insects diversify one molecule to serve two systems. Science 309(5742): 1826–1827

Prasad A, Qamri Z, Wu J, Ganju RK (2007) Pivotal advance: Slit-2/Robo-1 modulates the CXCL12/CXCR4-induced chemotaxis of T cells. J Leukoc Biol 82(3):465–476

Raftos DA, Cooper EL (1991) Proliferation of lymphocyte-like cells from the solitary tunicate, Styela clava, in response to allogeneic stimuli. J Exp Zool 260(3):391–400

Regan JC, Brandao AS, Leitao AB, Mantas Dias AR, Sucena E, Jacinto A, Zaidman-Remy A (2013) Steroid hormone signaling is essential to regulate innate immune cells and fight bacterial infection in Drosophila. PLoS Pathog 9(10), e1003720

Regier JC, Shultz JW, Kambic RE (2005) Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc Biol Sci 272(1561):395–401

Rosetto M, Engstrom Y, Baldari CT, Telford JL, Hultmark D (1995) Signals from the IL-1 receptor homolog, Toll, can activate an immune response in a *Drosophila* hemocyte cell line. Biochem Biophys Res Commun 209(1):111–116

Rota-Stabelli O, Telford MJ (2008) A multi criterion approach for the selection of optimal outgroups in phylogeny: recovering some support for Mandibulata over Myriochelata using mitogenomics. Mol Phylogenet Evol 48(1):103–111

Rota-Stabelli O, Campbell L, Brinkmann H, Edgecombe GD, Longhorn SJ, Peterson KJ, Pisani D, Philippe H, Telford MJ (2011) A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. Proc Biol Sci 278(1703): 298–306

Sadd BM, Schmid-Hempel P (2006) Insect immunity shows specificity in protection upon secondary pathogen exposure. Curr Biol 16(12):1206–1210

Sawaya MR, Wojtowicz WM, Andre I, Qian B, Wu W, Baker D, Eisenberg D, Zipursky SL (2008) A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms. Cell 134(6):1007–1018

Schmid-Hempel P (2005) Natural insect host-parasite systems show immune priming and specificity: puzzles to be solved. Bioessays 27(10):1026–1034

Schmucker D, Chen B (2009) Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. Genes Dev 23(2):147–156

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zypursky S (2000) *Drosophila*. Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101(9):671–684

Schwarz RS, Evans JD (2013) Single and mixed-species trypanosome and microsporidia infections elicit distinct, ephemeral cellular and humoral immune responses in honey bees. Dev Comp Immunol 40(3–4):300–310

Shi L, Yu HH, Yang JS, Lee T (2007) Specific *Drosophila* Dscam juxtamembrane variants control dendritic elaboration and axonal arborization. J Neurosci 27(25):6723–6728

Silverstein AM (1989) A history of immunology, vol XXII. Academic, San Diego, CA

Smith PH, Mwangi JM, Afrane YA, Yan G, Obbard DJ, Ranford-Cartwright LC, Little TJ (2011) Alternative splicing of the *Anopheles gambiae* Dscam gene in diverse Plasmodium falciparum infections. Malar J 10:156

Sun W, You X, Gogol-Doring A, He H, Kise Y, Sohn M, Chen T, Klebes A, Schmucker D, Chen W (2013) Ultra-deep profiling of alternatively spliced *Drosophila* Dscam isoforms by circularization-assisted multi-segment sequencing. EMBO J 32(14):2029–2038

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24(8):1596–1599

Vlisidou I, Dowling AJ, Evans IR, Waterfield N, ffrench-Constant RH, Wood W (2009) Drosophila embryos as model systems for monitoring bacterial infection in real time. PLoS Pathog 5(7), e1000518

Watson LF, Püttmann-Holgado FT, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI, Schmucker D (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 309:1874–1878

Watthanasurorot A, Jiravanichpaisal P, Liu H, Soderhall I, Soderhall K (2011) Bacteria-Induced Dscam Isoforms of the Crustacean, *Pacifastacus leniusculus*. PLoS Pathog 7(6), e1002062

Wood W, Jacinto A (2007) Drosophila melanogaster embryonic haemocytes: masters of multitasking. Nat Rev Mol Cell Biol 8(7):542–551

Worby CA, Simonson-Leff N, Clemens JC, Kruger RP, Muda M, Dixon JE (2001) The sorting nexin, DSH3PX1, connects the axonal guidance receptor, Dscam, to the actin cytoskeleton. J Biol Chem 276(45):41782–41789

Yamakawa K, Huo Y-K, Haendel MA, Hubert R, Chen X-N, Lyons GE, Korenberg JR (1998) DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. Hum Mol Genet 7(2): 227–237

Yu HH, Yang JS, Wang J, Huang Y, Lee T (2009) Endodomain diversity in the *Drosophila* Dscam and its roles in neuronal morphogenesis. J Neurosci 29(6):1904–1914

An Immune Effector System in the Protochordate Gut Sheds Light on Fundamental Aspects of Vertebrate Immunity

Assunta Liberti, Brittany Leigh, Rosaria De Santis, Maria Rosaria Pinto, John P. Cannon, Larry J. Dishaw, and Gary W. Litman

Abstract A variety of germline and somatic immune mechanisms have evolved in vertebrate and invertebrate species to detect a wide array of pathogenic invaders. The gut is a particularly significant site in terms of distinguishing pathogens from potentially beneficial microbes. *Ciona intestinalis*, a filter-feeding marine protochordate that is ancestral to the vertebrate form, possesses variable region-containing chitin-binding proteins (VCBPs), a family of innate immune receptors, which recognize bacteria through an immunoglobulin-type variable region. The manner in which VCBPs mediate immune recognition appears to be related to the development and bacterial colonization of the gut, and it is likely that these molecules are critical elements in achieving overall immune and physiological homeostasis.

A. Liberti • R. De Santis • M.R. Pinto
Department of Animal Physiology and Evolution, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy
e-mail: assusy.liberti@gmail.com; desantis@szn.it; pinto@szn.it

B. Leigh
Department of Marine Sciences, University of South Florida St. Petersburg, St Petersburg, FL, USA
e-mail: bleigh@mail.usf.edu

J.P. Cannon • L.J. Dishaw • G.W. Litman (✉)
Department of Pediatrics, University of South Florida College of Medicine, St Petersburg, FL, USA
e-mail: jcannon@health.usf.edu; ldishaw@health.usf.edu; litmang1@jhmi.edu

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_7

1 The Antiquity and Evolution of Immune Genetic Diversity

Extreme genetic variation in the small genomes of some viruses, bacteria, and other protozoa results in extraordinary functional diversity, of which some is utilized to escape host defense strategies (Bainard and Gregory 2013; Steward et al. 2000). Since all living organisms can generate genetic diversity, which may result in functional novelty, an evolutionary arms race is maintained between host immune receptors and potential pathogens. Similar dynamics influence harmless as well as beneficial microbes (Hooper and Gordon 2001); the host must evolve strategies that either ignore or actively protect some organisms while aggressively responding to those that otherwise could prove fatal. A remarkable number of different mechanisms have evolved at all levels of invertebrate and vertebrate phylogeny; however, in the vertebrates, an adaptive strategy for generating an extraordinary level of receptor variation at the somatic cell level gives rise to highly specific responses, which can mature affinity and, through providing memory, effect heightened responses upon secondary contact. This strategy can afford protection to nonpathogenic or nonthreatening flora and has the potential to both interfere with and minimize collateral damage common to innate responses.

The immunoglobulin superfamily (IgSF) is one of the largest and most functionally diversified gene families. Through domain duplication and shuffling, diversified gene families, consisting of multiple nonhomologous segments, have emerged. These diverse genes serve roles in both innate and adaptive immune responses, as well as in some aspects of developmental patterning. Because of the vast sequence variation thus created in certain members of the IgSF, bona fide orthologs may conserve only short amino acid segments, confounding efforts to detect homologous forms in lower chordate species. Although the ancestral elements responsible for the somatic diversification of adaptive antigen receptors are not understood, characterizing novel Ig-like domains in the more ancient phylogenetic taxa has led to the discovery of alternative forms of the Ig-domain-containing proteins. It is likely that organisms whose ancestry predates vertebrate divergence may provide the most meaningful insight into the evolution of the components that were assembled into modern adaptive immune systems.

2 Immunity in Protochordates: The VCBPs

Protochordates are deuterostome invertebrates belonging to the phylum Chordata, of which two subphyla are recognized: the cephalochordates, of which amphioxus is representative, and the urochordates, of which the sea squirt is representative (Fig. 1). Cephalochordates diverged from a common chordate ancestor and urochordates diverged more recently from a common ancestor with vertebrates (Delsuc et al. 2006). Protochordates maintain chordate anatomical and developmental characteristics such as: an endostyle, a notochord, pharyngeal slits, segmented

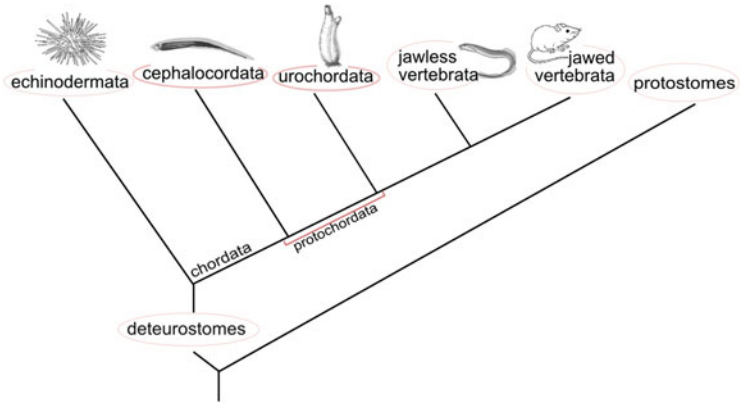


Fig. 1 The evolution of Chordates. Divergence of Chordata, including subphyletic grouping of Vertebrata, Urochordata, and Cephalochordata. The urochordates, which include *Ciona intestinalis*, and the cephalochordates, which include *Branchiostoma floridae*, both are classified as protochordates

muscle blocks, a hollow dorsal nerve cord, and a postanal tail (Berrill 1947) but do not possess adaptive immune systems. Organized lymphoid tissues, which play important roles in various phases of the adaptive immune response, are lacking in protochordates and all other invertebrates. Protochordates afford an opportunity to further characterize a chordate system that relies exclusively on streamlined innate immune networks to achieve and maintain homeostasis in the absence of a highly complex adaptive immune system (e.g., redundant pathways, complex signaling, rearranging receptors, and additional specialized cell types).

2.1 VCBPs in Branchiostoma floridae

In the course of an investigation into the origins of early forms of the immunoglobulin/T cell antigen receptor (Ig/TCR) variable region domain, the variable region-containing chitin-binding protein (VCBP) genes were identified in the cephalochordate amphioxus, *Branchiostoma floridae* (Cannon et al. 2002). The VCBPs include at least five proteins (VCBP1–5) that possess two tandem V-type domains at the N-terminus and a single chitin-binding domain (CBD) at the C-terminus (Fig. 2A). *VCBP1*, *VCBP2*, *VCBP4*, and *VCBP5* map to the same chromosomal locus and *VCBP3* maps to a separate locus.

An extraordinarily large number of haplotypes of VCBPs have been identified in amphioxus (Cannon et al. 2004). Although at the time of their discovery VCBP function was not understood, it was hypothesized that extensive polymorphism may compensate for a lack of somatic variation in receptor diversity. *VCBP2* and *VCBP5* share closer overall organization and sequence identity as compared to *VCBP1*, *VCBP3*, and *VCBP4*, which are related more distantly to each other as well

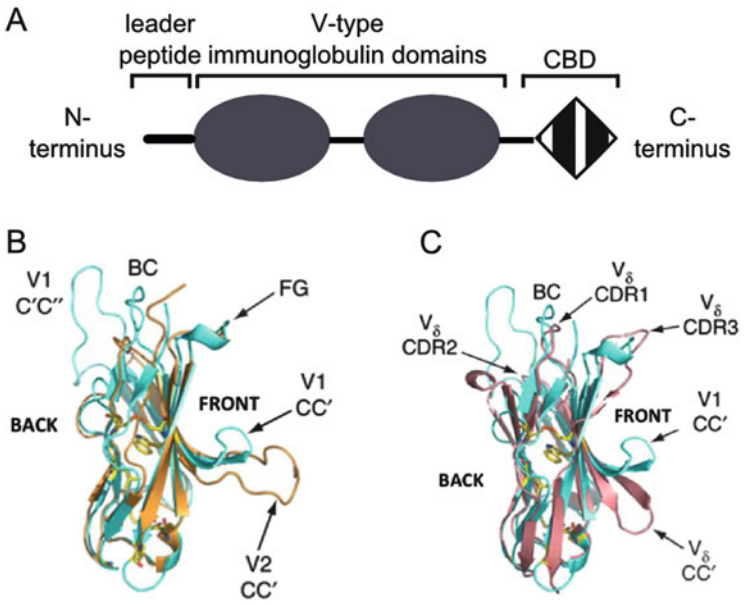


Fig. 2 VCBP protein structure. (**A**) A generic VCBP protein possessing a leader peptide, two tandem N-terminal V-type immunoglobulin domains, and a C-terminal chitin-binding domain (CBD). (**B**) The crystal structure of the VCBP3 domain compared with other antigen receptors, with the V1 domain being the most closely related to TCR Vδ (**C**). VCBP3 V1 superimposed on the V2 domain reveals that the CC′ and BC loops differ. Adapted from Dishaw et al. (2008) and Hernandez Prada et al. (2006)

as to the *VCBP2*/*VCBP5* cluster (Dishaw et al. 2008). The *VCBP2*/*VCBP5* cluster has been characterized extensively; the genes display indel polymorphism in both coding and noncoding regions. In particular, one of the haplotypes (B) exhibits a much higher instance of inverted repeats (IRs) compared to the rest of the genome and that in turn may influence genomic stability. Gene inactivation, diversity, and transcriptional conformation could be affected (Dishaw et al. 2008).

VCBPs are expressed predominantly as full-length versions; shorter-length spliced variants also have been identified. VCBP haplotypes are characterized both by copy number variation and by extensive variation in repeat type and density. Gene duplication, gene conversion, locus differentiation, unequal crossover, and/or other mechanisms of gene exchange may generate exceptional VCBP haplotype and allelic diversity in wild populations (Dishaw et al. 2010).

VCBPs exhibit certain characteristics of what we consider as immune recognition molecules, including: a high degree of germline polymorphism, V-family distribution, tissue-specific expression (largely confined to the gut), and a chimeric immunoglobulin-lectin structure similar to other non-vertebrate receptor types (Cannon et al. 2002; Ghosh et al. 2011). The structures of the V domains in amphioxus are situated head-to-tail (Fig. 2B, C), as opposed to the jawed vertebrate Igs and TCRs, where the two V domains are packed in a head-to-head orientation

(Cannon et al. 2002; Hernandez Prada et al. 2006). Hypervariable complementarity-determining regions (CDRs) in vertebrate Igs are positioned in a manner that cooperatively form an antigen-binding site comprising two V domains. In VCBPs, the orientation of the two V-type domains positions the sequence segments corresponding to CDRs in Ig vertebrate V regions at the opposite ends (Fig. 2B, C). It is important to note that the sequence regions that exhibit the greatest variation in VCBPs do not correspond to the positions of CDRs in vertebrate Ig V regions and likely are not involved in interactions of VCBPs with their ligands (Hernandez Prada et al. 2006). The contiguous joining J-region, which is present in all rearranging antigen receptor genes, is important for antigen receptor dimerization and combinatorial diversity (Gough and Bernard 1981). However, J-region-like segments in VCBPs are intimately involved in V-domain packing (Hernandez Prada et al. 2006). The V domains of VCBPs represent an alternative utilization of the V domain as a platform for immunological function.

In addition to the V-type domains, VCBPs possess C-terminal regions consisting of conserved structural residues (e.g., characteristically spaced cysteines) that are present in CBDs. Chitin, beta-(1–4)-poly-*N*-acetyl D-glucosamine, is the second most abundant polysaccharide found in nature, after cellulose, and is present in every kingdom including bacteria, Archaea, fungi, plants, animals, and protists (Jeuniaux and Vossfoucart 1991). Chitin exerts a number of effects on animal physiology and immunity as a signaling molecule, likely mediated by host enzymes such as chitinases. Host production of chitin via chitin synthase genes is observed throughout all classes of metazoans except mammals, indicating an important and integral ancient role for this biopolymer (Lee 2009). More than 250 CBD-encoding open reading frames, scattered over 31 scaffolds, have been identified in the amphioxus genome. *VCBP1* and *VCBP3* contain exons that encode the CBM_14-type CBD, a domain in the carbohydrate-binding family that is related very closely to those described in peritrophins, which are glycoproteins that are incorporated into the chitin-rich gut environment of insects (Dishaw et al. 2008). Expression of VCBP proteins in amphioxus essentially is specific to the gut of filter-feeding adults that are exposed to a constant influx of potential pathogens.

2.2 *VCBPs in* Ciona intestinalis

Amphioxus at best is a challenging experimental model system for characterizing molecular functions. In order to determine VCBP function, a different protochordate model system, *Ciona intestinalis*, is being developed. *Ciona* is a sessile, filter-feeding protochordate that has served as a significant model in developmental biology (Satoh 2003). The genome of *Ciona* has been sequenced and annotated, and four unlinked VCBPs have been identified: *VCBP-A*, *VCBP-B*, *VCBP-C*, and *VCBP-D* (Azumi et al. 2003; Cannon et al. 2002, 2004; Dishaw et al. 2011).

VCBP-A, *VCBP-C*, and *VCBP-D* are located on chromosomes 5, 4, and 10, respectively (JGI version 2.0). *VCBP-B* was placed on scaffold 41 in the first *Ciona* genome draft (JGI version 1.0), but assembly-related problems and/or allelic ambiguities have precluded its mapping in the second *Ciona* genome draft (JGI version 2.0). The predicted transcript lengths of *VCBP-A*, *VCBP-B*, and *VCBP-C* are ~1 kb. *VCBP-A* and *VCBP-B* are encoded by nine exons of similar length and are related most closely; however, their respective introns and intergenic regions exhibit only limited sequence similarity, indicating that they do not represent allelic variants (Dishaw et al. 2011). *VCBP-C* and *VCBP-D* are encoded in five exons and are more similar to each other than to *VCBP-A* and *VCBP-B*. The 3′ terminal exons of *VCBP-D* encode termination and polyadenylation signals but not a CBD (Dishaw et al. 2011).

Multiple, randomly distributed, single nucleotide polymorphisms (SNPs) have been identified in coding regions, introns, and intergenic regions of *VCBP-A*, *VCBP-B*, and *VCBP-C*. *VCBP-D* is the least polymorphic (Dishaw et al. 2011). Further inspection of polymorphisms indicates that *VCBP-C* and, to a minor extent, *VCBP-B* are represented by limited main haplotypes. Point mutations and frequent indels have been identified in intron and intergenic regions. Frequent indels have been identified in introns of alleles that exhibit moderate polymorphism in their coding regions (Dishaw et al. 2011). Although the VCBP genes in *Ciona* are polymorphic, they lack the high degree of haplotypic variation seen in amphioxus VCBPs (Cannon et al. 2004; Dishaw et al. 2008).

VCBP (A–C) in *Ciona* exhibit the same domain organization as VCBPs in amphioxus consisting of a leader peptide, two N-terminal V-type Ig domains, and a single C-terminal CBD. A gene corresponding to VCBP-D, which contains two complete V-type domains but lacks a CBD, has not been identified in amphioxus. However, an alternatively spliced form of VCBP3, in which the CBD is absent, is found in this species (Dishaw et al. 2008). Comparisons of VCBPs at the peptide level indicate ~72 % relatedness at the predicted peptide level between VCBP-A and VCBP-B and ~29 % identity when VCBP-C is compared with VCBP-A and VCBP-B. VCBP-D shares ~47 % identity with VCBP-C and ~26 % with VCBP-A and VCBP-B, taking into account its shorter length (Dishaw et al. 2011).

2.2.1 VCBP Expression in Immunocompetent Tissues from Larval Stage to the Adult

At an early stage in the life cycle of *Ciona*, the embryo develops into a swimming tadpole larva, which attaches to a substrate within the course of several hours. After settlement, the larva undergoes metamorphosis, during which the body axes are reorganized, the tail retracts, and consequently the notochord and the tubular nervous system are lost (Satoh 2003). VCBP genes in *Ciona* are expressed in defined territories from the larval stage through metamorphosis to adulthood (Fig. 3). Specifically, *VCBP-A* and *VCBP-C* are expressed the earliest, paralleling the development of the digestive tract. The patterns of expression essentially trace

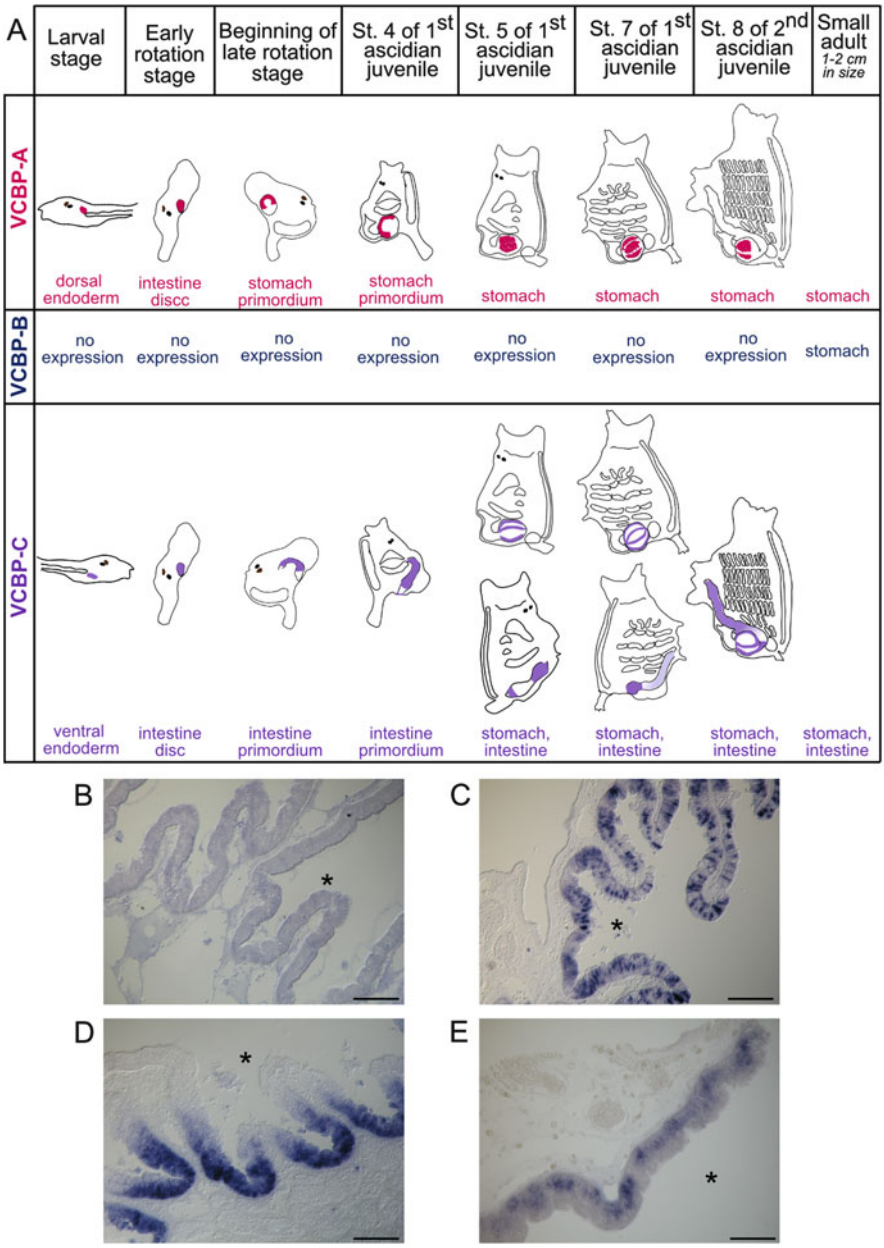


Fig. 3 VCBP expression in *Ciona intestinalis* gut. (**A**) *VCBP* expression patterns from larval stage to "small adult." In the adult stomach epithelium, (**B**) *VCBP-A* expression is not detected, while (**C**) *VCBP-B* and (**D**) *VCBP-C* are expressed differentially in scattered cells of the epithelium and in cells localized in the stomach crypts, respectively. (**E**) *VCBP-C* expression also is detected in the intestine. *Asterisk*, the lumen of the organ. *Scale bars*: (**B**)–(**D**) 100 μm; (**E**) 50 μm. Adapted from Liberti et al. (2014) and Dishaw et al. (2011)

the formation and the specification of distinctive areas of the stomach and intestine (Fig. 3A) (Liberti et al. 2014).

At the larval stage, *VCBP-A* and *VCBP-C* are localized in two different regions of the endoderm (Liberti et al. 2014), corresponding to the prospective regions in the adult that give rise to the esophagus, stomach, and intestine, respectively (Hirano and Nishida 2000). At the initiation of metamorphosis, *VCBP-A* and *VCBP-C* are expressed in the primordium of the gut, designated as intestine disk. From the beginning of late rotation (BLR) stage until the end of metamorphosis, *VCBP-A* and *VCBP-C* are expressed in different regions (Fig. 3A). Throughout metamorphosis, expression of *VCBP-A* is confined to scattered cells of the stomach epithelium (Liberti et al. 2014). At the BLR stage, *VCBP-C* is detected in one region of the developing intestine, whereas at stage 4 of 1st ascidian juvenile, corresponding to a further level of intestinal differentiation, gene expression is localized to a small ring of cells at the border between the stomach and intestine as well as to a more extended region of the intestine identified as the hindgut (Fig. 3A). The expression patterns of *VCBP-C* may reflect early compartmentalization of functions in the developing intestine. At the onset of feeding, at stage 5 of 1st ascidian juvenile, *VCBP-C* also is expressed in the stomach, suggesting a link between the gene expression in this organ and the contact with the external environment (Fig. 3A). Expression of *VCBP-C* in the stomach and intestine is maintained until adulthood (Liberti et al. 2014).

At the end of metamorphosis, *VCBP-A* and *VCBP-C* exhibit different expression patterns. *VCBP-A* is expressed in cells scattered in the epithelium of the stomach, whereas *VCBP-C* expression is localized in groups of cells at the base of the ridges, which are a feature of the adult stomach. The expression patterns of these genes suggest they represent either different cell types or different stages of differentiation of the same cell type (Liberti et al. 2014). At the small adult stage, corresponding to animals 1–2 cm in size, *VCBP-A* expression appears to switch off and is replaced by expression of *VCBP-B* that occurs in scattered cells of the stomach epithelium (Liberti et al. 2014).

Differences in the expression patterns of *VCBP-A* and *VCBP-C* suggest they may play a dual role in development and immunity. The timing and patterns of the expression suggest a role in mediating initial microbial colonization of the gut. It is well recognized that the gut epithelium in vertebrates is the first line of defense against invading pathogens. Gut innate immune molecules recognize microbial signatures and mediate signaling to the adaptive immune system of an impending attack. In both vertebrates and invertebrates, a protective mucus layer, colonized extensively by mostly beneficial microbes that are important for host health and homeostasis, covers the gut epithelium. Deciphering beneficial from pathogenic microorganisms is a critical task as microbial heterogeneity likely accounts for extensive variation in immune molecules of gut tissue. However, a direct role for VCBP molecules in gut differentiation, in some ways reminiscent of the duality of function seen in the DSCAM system (Boehm 2007), cannot be ruled out. Effective knockdown strategies of VCBPs could clarify whether or not these genes are

important in specifying territories of the digestive tract and/or in the differentiation of the various functional tracts of the gut.

In the adult, *VCBP-A*, *VCBP-B*, and *VCBP-C* transcripts are localized in immune competent tissues, including the digestive tract (Fig. 3B–E) and the blood cells (Dishaw et al. 2011). *VCBP-B* expression in the gut is localized in cells scattered throughout the stomach epithelium (Fig. 3E); in contrast, *VCBP-C* is expressed in the cell types that reside at the base of each epithelial fold (Fig. 3D) (Dishaw et al. 2011; Liberti et al. 2014). VCBP-C protein is localized both in the granules of zymogenic cells, which morphologically are analogous to mammalian stomach and pancreatic cells (Burighel 1997), and as soluble protein in the stomach lumen associated with the microvilli of the stomach cells (Dishaw et al. 2011). *VCBP-C* also can be detected in the intestine (Dishaw et al. 2011) where it maintains the regionalized expression observed during the differentiation of the organ at the juvenile stage (Liberti et al. 2014). The expression of both *VCBP-A* and *VCBP-C* has been detected in granular amoebocytes localized in the connective tissue surrounding the stomach epithelium, the so-called lamina propria, as well as in circulating blood cells (Dishaw et al. 2011; Liberti et al. 2014). *VCBP-A* is expressed in the blood cells, but not in the stomach epithelium, where VCBP-A protein has been detected. VCBP-A protein can be detected in the cytoplasm and large vacuoles of the absorptive cells as well as in the *lamina propria*, both as a secreted molecule in the lymph and in the granular amoebocytes. This latter group of cells has been observed to be associated closely with the basement and plasma membranes of the stomach cells. Although further investigation will be necessary to clarify the dynamics of the localization of VCBP-A in the stomach cells, these data support the hypothesis that a non-endogenous origin of VCBP-A, likely from the adjacent blood cell-rich tissue, accounts for localization in the stomach epithelial cells (Liberti et al. 2014). The expression of *VCBP-D* has not yet been characterized.

2.2.2 VCBPs in Recognition of Bacteria

Although the precise function of VCBP molecules has not yet been determined, extensive data indicate that these proteins act as secreted immune-type molecules (Dishaw et al. 2011, 2012). Experiments in which *Ciona* were fed either Gram-positive (*Bacillus cereus*) or Gram-negative (*Escherichia coli*) bacteria indicate that endogenous VCBP-C binds to the bacteria present in the stomach lumen (Fig. 4a, b) (Dishaw et al. 2011). Furthermore, affinity-purified VCBP-C protein acts as an opsonin, increasing the capacity of granular amoebocytes to phagocytose FITC-labeled *B. cereus* in vitro. An increase in phagocytosis also is observed when the bacteria are preincubated with recombinant VCBP-C protein fragment containing only the V-type domains. The full-length form of VCBP-C is only slightly more effective as an opsonin than is the truncated recombinant form lacking the CBD (Dishaw et al. 2011). VCBP-C is secreted by the stomach epithelium into the lumen, where it binds ingested bacteria. At the end of

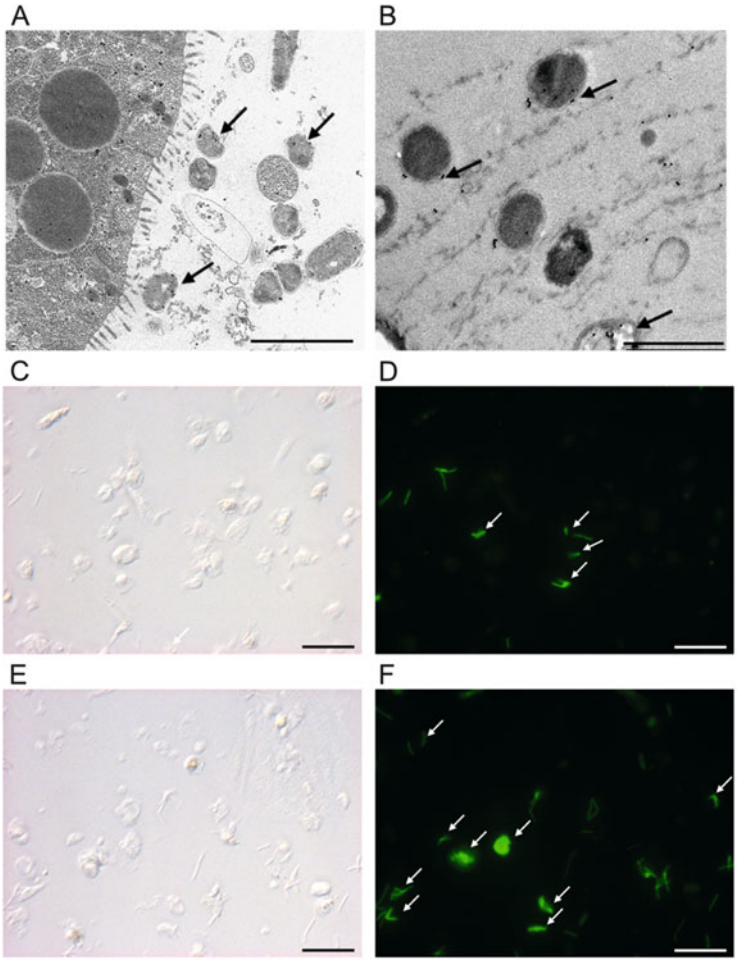


Fig. 4 VCBP-C functions as an immune-type molecule. VCBP-C immuno-localization on stomach sections of bacteria-fed animals showing gold particles localized on the surface (*arrows*) of both (**A**) *B. cereus* and (**B**) *E. coli*, indicating the ability of secreted VCBP-C to bind to the bacteria in the organ lumen. Phagocytic activity of *Ciona* hemocytes to FITC-labeled *B. cer*eus (**C** and **D**). When the FITC-labeled *B. cereus* (*arrows*) are preincubated with the affinity-purified VCBP-C protein, an increased number of phagocytic granular amoebocytes (*arrows*) is observed, consistent with an opsonic function for VCBP-C (**E** and **F**). (**C** and **E**) fluorescence microscopy and (**D** and **F**) Nomarski optics. *Scale bars*: (**A**) 2 μm, (**B**) 1 μm, and (**C**)–(**F**) 20 μm. Adapted from Dishaw et al. (2011)

metamorphosis (e.g., stage 7–8 juvenile), corresponding to the early phases of contact between the gut and the external environment, the ingestion of either Gram-positive *B. cereus* or Gram-negative *E. coli* influences the expression of *VCBP-A* and *VCBP-C*, both of which can be detected at this stage; this observation is consistent with the hypothesis that the VCBPs have a role at the onset of gut colonization by microbiota as well as in gut homeostasis and immunity (Dishaw

et al. 2011). *VCBP-A* and *VCBP-C* exhibit an opposite response to different bacteria strains (Liberti et al. 2014). The relative ease of generating germfree (gnotobiotic) *Ciona* will facilitate investigation of how the host-microbiota interaction affects VCBP expression and/or their localization in the gut lumen. By introducing selected bacteria strains in gnotobiotic animals, it will be possible to study which strains or communities of bacteria influence VCBP expression as well as how pathogenic bacteria influence these processes.

The role of the VCBPs in the complex interaction between the host and microorganisms across the gut epithelium is still under investigation. *Ciona* is ideally suited to such investigations as they possess numerous experimental advantages for studies of gut host-microbe interactions, a topic that is of particular current medical relevance in relation to bowel and other diseases in humans (Dishaw et al. 2012). The functional analogies that can be drawn between VCBPs and other components of gut immunity in invertebrates and gut immunity in higher vertebrates are striking. In mammals it is known that specialized epithelial cells, such as enterocytes and goblet cells, secrete mucins, which are glycoproteins that assemble into a viscous gel-like mucus layer that provides the frontline protections against microorganisms in the gut lumen (Atuma et al. 2001; Hooper and Macpherson 2010; Johansson et al. 2008). The gel-forming mucins are a recognized element of innate immunity and are well conserved during evolution, from the earliest metazoans (Lang et al. 2007). Mucin-related proteins have been identified in the *Ciona* genome but have not been characterized in detail (Lang et al. 2007). It is essential to study the structure of the mucus layer within the *Ciona* gut and determine if it is organized as in mammals with an inner, firmly adherent epithelium-associated layer devoid of bacteria and an outer loose, non-adherent layer, rich with bacteria (Johansson et al. 2008). Determining how the VCBP molecules are distributed in each compartment of the gut, if they are immobilized in some way to the mucus layers, how they attach to lumen bacteria, if specificity exists in such interactions, and what factors influence the distribution of these molecules will help delineate the role of VCBPs in the complex dynamics that define the colonization of the gut by resident microbiota and defense against pathogens. The diverse locations in which VCBPs are expressed and secreted could indicate a differential role played by the VCBP molecules during microbiota colonization in the distinct compartments of the digestive tract. Although the core community of bacteria is known in *Ciona* (Dishaw et al. 2014b), it is particularly important to elucidate the distribution of these microbiota among the stomach, midgut, and hindgut and determine what, if any, relationship this has to the patterned expression and secretion of VCBPs.

2.2.3 VCBP and IgA

Increasing interest has been focused on primary investigations of both innate and adaptive immunity in the digestive tract, which essentially constitutes an external environment within an animal. Even the simplest animals with a gut appear to host complex microbiota (Dishaw et al. 2014a). Separation of the gut lumen from the

host via a single layer of epithelium is a common feature of all animal forms. The epithelial layer is fully immunocompetent and secretes substances that initiate and maintain barriers. In addition, these barriers also support the growth of the host-specific microbiomes to which the animal in general does not mount detrimental responses (Hooper and Gordon 2001). While these barriers are maintained primarily via innate immune responses, the gut has sustained conditions that have driven and shaped the evolution of different forms of immunity over tens to hundreds of millions of years. Adaptive immunity in jawed and jawless vertebrates may have evolved to negotiate the intrinsic complexity of microbiomes and to regulate tolerance to beneficial flora (Lee and Mazmanian 2010; McFall-Ngai 2007). Adaptive immune systems, over millions of years, have coevolved with signals from symbiotic microbiota in ways that have resulted in protection of both the host and microbiota from pathogens (Lee and Mazmanian 2010). In vertebrates, the developmental maturation of the gut is coupled to symbiotic interactions that also are responsible for sustaining homeostasis. We speculate that while complex animal forms evolved vasculature within the gut to mobilize nutrition through larger, sub-compartmentalized bodies (as seen in most vertebrates), the adaptive immune system and functionally diverse immunocytes provide long-term protection from pathogens that may breach this limited barrier. Thus, the origins and evolution of adaptive immunity can be understood better by studying how diverse model systems (including those lacking adaptive immunity) negotiate with complex gut microbiota.

Although much is known about the genetics and structure of VCBPs, relatively little is known regarding the relationship between the Ig V regions and the C-terminal CBDs. In a broad sense, parallels exist between protochordate VCBPs and mammalian IgA (Holland et al. 2008), the predominant class of antibody produced in mucosal secretions of mammals. IgA is produced in a unique anatomical site localized within the gut lamina propria, i.e., in the gut-associated lymphoid tissue (GALT). The main site of production is in the Peyer's patches (Craig and Cebra 1971) and isolated lymphoid follicles (Fagarasan et al. 2010). Once produced, dimeric IgA is transported into the lumen via polymeric immunoglobulin receptor-dependent transcytosis through the intestinal epithelial cells (Mostov et al. 1986). The high-affinity, antigen-specific mode of IgA production within the gut lumen originally was recognized as an important factor protecting host integrity by preventing colonization and invasion by pathogenic microbes. IgA also may function in a low-affinity mode to confine the dense commensal microbiota within the intestinal lumen through "immune exclusion," which effectively eliminates microbes by coating them with the soluble antibody, thus preventing their adherence to the epithelial surface (Macpherson et al. 2008). Secretory IgA also may function within distinct compartments of the gut through the selective modulation of biofilms (Bollinger et al. 2003).

Structural and functional features of VCBPs strongly suggest parallels with IgA. Although vertebrate-type GALT has not been identified in *Ciona*, the localization of the VCBPs in different areas of the stomach epithelium implies that this organ could be the main site of production of immune molecules involved in gut

homeostasis. The VCBP-bacteria interaction should not be interpreted from the traditional perspective of vertebrate immunity but rather in the context of a complex symbiosis of host and microbes across the gut (Dishaw et al. 2011). VCBP-A, VCBP-B, and VCBP-C could mediate both immune exclusion and potentially other functions within the host-bacterial interface. Specifically, VCBP-C may function in the immune exclusion mechanism of microorganisms in the stomach lumen, and VCBP-A could have a defense role in at least one type of stomach cell that is in contact with microorganisms that have crossed the epithelial barrier. Although VCBPs can recognize microorganisms via the V-type domains (Dishaw et al. 2011), the precise structural mechanism of recognition likely is entirely different from that of the antigen-binding receptors in contemporary vertebrates. Protochordate VCBPs and higher vertebrate IgA present an interesting potential case of convergent evolution of gut immunity.

3 Conclusion

The presence and diversity of the V domain in VCBPs suggested initially that these molecules could be important in immune recognition; however, only recently has a role for them in mediating immune function been shown. A role for VCBPs in gut development and function underscores their broad involvement in immunological homeostasis. Although VCBPs do not exhibit the high degree of variability of somatically rearranged adaptive immune receptors, they likely mediate an array of functions, some of which may be shared with IgA, an important feature of adaptive immune systems of vertebrates. Irrespective of our less than complete understanding of VCBP function(s), they emphasize continuity between innate and adaptive immunity. Further investigations of VCBP function in the *Ciona* gut promise to define and clarify the precise role of the VCBPs in microbial gut colonization, innate immunity, and development of the gut.

Acknowledgments The authors would like to thank both the ASSEMBLE (Association of European Marine Biological Laboratories) program for collaborative support and Barbara Pryor for editorial assistance. LJD is supported by grants from the All Children's Hospital Foundation and the University of South Florida College of Medicine Sponsored Research; GWL is supported by NIH R01 Al 23338.

References

Atuma C, Strugala V, Allen A, Holm L (2001) The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. Am J Physiol Gastrointest Liver Physiol 280:G922–G929

Azumi K, De Santis R, De Tomaso A, Rigoutsos I, Yoshizaki F, Pinto MR, Marino R, Shida K, Ikeda M, Ikeda M, Arai M, Inoue Y, Shimizu T, Satoh N, Rokhsar DS, Pasquier D, Kasahara L, Satake M, Nonaka M (2003) Genomic analysis of immunity in a Urochordate and the

emergence of the vertebrate immune system: “waiting for Godot”. Immunogenetics 55:570–581
Bainard JD, Gregory TR (2013) Genome size evolution: patterns, mechanisms, and methodological advances. Genome 56:Vii–Viii
Berrill NJ (1947) The development and growth of *Ciona*. J Mar Biol Assoc UK 26:616–625
Boehm T (2007) Two in one: dual function of an invertebrate antigen receptor. Nat Immunol 8:1031–1033
Bollinger RR, Everett ML, Palestrant D, Love SD, Lin SS, Parker W (2003) Human secretory immunoglobulin A may contribute to biofilm formation in the gut. Immunology 109:580–587
Burighel P, Cloney RA (1997) Urochordata: Ascidiacea. In: Harrison FW, Ruppert EE (eds) Microscopic analysis of invertebrates. Wiley-Liss, New York, pp 221–347
Cannon JP, Haire RN, Litman GW (2002) Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. Nat Immunol 3:1200–1207
Cannon JP, Haire RN, Schnitker N, Mueller MG, Litman GW (2004) Individual protochordates have unique immune-type receptor repertoires. Curr Biol 14:R465–R466
Craig SW, Cebra JJ (1971) Peyer’s patches – enriched source of precursors for iga-producing immunocytes in rabbit. J Exp Med 134:188–200
Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439:965–968
Dishaw LJ, Mueller MG, Gwatney N, Cannon JP, Haire RN, Litman RT, Amemiya CT, Ota T, Rowen L, Glusman G, Litman GW (2008) Genomic complexity of the variable region-containing chitin-binding proteins in amphioxus. BMC Genet 9:78
Dishaw LJ, Ota T, Mueller MG, Cannon JP, Haire RN, Gwatney NR, Litman RT, Litman GW (2010) The basis for haplotype complexity in VCBPs, an immune-type receptor in amphioxus. Immunogenetics 62:623–631
Dishaw LJ, Giacomelli S, Melillo D, Zucchetti I, Haire RN, Natale L, Russo NA, De Santis R, Litman GW, Pinto MR (2011) A role for variable region-containing chitin-binding proteins (VCBPs) in host gut-bacteria interactions. Proc Natl Acad Sci USA 108:16747–16752
Dishaw LJ, Flores-Torres JA, Mueller MG, Karrer CR, Skapura DP, Melillo D, Zucchetti I, De Santis R, Pinto MR, Litman GW (2012) A basal chordate model for studies of gut microbial immune interactions. Front Immunol 3:96
Dishaw LJ, Cannon JP, Litman GW, Parker W (2014a) Immune-directed support of rich microbial communities in the gut has ancient roots. Dev Comp Immunol 47:36–51
Dishaw LJ, Flores-Torres J, Lax S, Gemayel K, Leigh B, Melillo D, Mueller MG, Natale L, Zucchetti I, De Santis R, Pinto MR, Litman GW, Gilbert JA (2014b) The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. PLoS One 9, e93386
Fagarasan S, Kawamoto S, Kanagawa O, Suzuki K (2010) Adaptive immune regulation in the gut: T cell-dependent and T cell-independent IgA synthesis. Annu Rev Immunol 28:243–273
Ghosh J, Lun CM, Majeske AJ, Sacchi S, Schrankel CS, Smith LC (2011) Invertebrate immune diversity. Dev Comp Immunol 35:959–974
Gough NM, Bernard O (1981) Sequences of the joining region genes for immunoglobulin heavy chains and their role in generation of antibody diversity. Proc Natl Acad Sci USA 78:509–513
Hernandez Prada JA, Haire RN, Allaire M, Jakoncic J, Stojanoff V, Cannon JP, Litman GW, Ostrov DA (2006) Ancient evolutionary origin of diversified variable regions demonstrated by crystal structures of an immune-type receptor in amphioxus. Nat Immunol 7:875–882
Hirano T, Nishida H (2000) Developmental fates of larval tissues after metamorphosis in ascidian *Halocynthia roretziI*. II. Origin of endodermal tissues of the juvenile. Dev Genes Evol 210:55–63
Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, Ferrier DE, Garcia-Fernandez J, Gibson-Brown JJ, Gissi C, Godzik A, Hallbook F, Hirose D, Hosomichi K, Ikuta T, Inoko H, Kasahara M, Kasamatsu J, Kawashima T, Kimura A, Kobayashi M, Kozmik Z, Kubokawa K, Laudet V, Litman GW, McHardy AC, Meulemans D, Nonaka M, Olinski RP, Pancer Z, Pennacchio LA, Pestarino M,

Rast JP, Rigoutsos I, Robinson-Rechavi M, Roch G, Saiga H, Sasakura Y, Satake M, Satou Y, Schubert M, Sherwood N, Shiina T, Takatori N, Tello J, Vopalensky P, Wada S, Xu A, Ye Y, Yoshida K, Yoshizaki F, Yu JK, Zhang Q, Zmasek CM, de Jong PJ, Osoegawa K, Putnam NH, Rokhsar DS, Satoh N, Holland PW (2008) The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res 18:1100–1111

Hooper LV, Gordon JI (2001) Commensal host-bacterial relationships in the gut. Science 292:1115–1118

Hooper LV, Macpherson AJ (2010) Immune adaptations that maintain homeostasis with the intestinal microbiota. Nat Rev Immunol 10:159–169

Jeuniaux C, Vossfoucart MF (1991) Chitin biomass and production in the marine-environment. Biochem Syst Ecol 19:347–356

Johansson MEV, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc Natl Acad Sci USA 105:15064–15069

Lang TA, Hansson GC, Samuelsson T (2007) Gel-forming mucins appeared early in metazoan evolution. Proc Natl Acad Sci USA 104:16209–16214

Lee CG (2009) Chitin, chitinases and chitinase-like proteins in allergic inflammation and tissue remodeling. Yonsei Med J 50:22–30

Lee YK, Mazmanian SK (2010) Has the microbiota played a critical role in the evolution of the adaptive immune system? Science 330:1768–1773

Liberti A, Melillo D, Zucchetti I, Natale L, Dishaw LJ, Litman GW, De Santis R, Pinto MR (2014) Expression of *Ciona intestinalis* variable region-containing chitin-binding proteins during development of the gastrointestinal tract and their role in host-microbe interactions. PLoS One 9, e94984

Macpherson AJ, Mccoy KD, Johansen FE, Brandtzaeg P (2008) The immune geography of IgA induction and function. Mucosal Immunol 1:11–22

McFall-Ngai M (2007) Adaptive immunity: care for the community. Nature 445:153

Mostov KE, de Bruyn KA, Deitcher DL (1986) Deletion of the cytoplasmic tail of the polymeric immunoglobulin receptor prevents basolateral localization and endocytosis. Cell 47:359–364

Satoh N (2003) The ascidian tadpole larva: comparative molecular development and genomics. Nat Rev Genet 4:285–295

Steward GF, Montiel JL, Azam F (2000) Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. Limnol Oceanogr 45:1697–1706

Variable Lymphocyte Receptors: A Current Overview

Masanori Kasahara

This chapter is dedicated to the memory of Dr. Zeev Pancer who passed away on April 20, 2014.

Abstract Jawless vertebrates represented by lampreys and hagfish mount antigen-specific immune responses using variable lymphocyte receptors. These receptors generate diversity comparable to that of T-cell and B-cell receptors by assembling multiple leucine-rich repeat modules with highly variable sequences. Although it is true that jawed and jawless vertebrates have structurally unrelated antigen receptors, their adaptive immune systems have much in common. Most notable is the conservation of lymphocyte lineages. It appears that specialized lymphocyte lineages emerged in a common vertebrate ancestor and that jawed and jawless vertebrates co-opted different antigen receptors within the context of such lymphocyte lineages.

1 Introduction

Lampreys and hagfish, also known as cyclostomes or agnathans, are the sole survivors of the once-flourishing jawless vertebrates (Janvier 2006; Shimeld and Donoghue 2012). Despite being a minority group comprising ~120 species in total, they occupy a critical phylogenetic position for understanding the evolution of adaptive immunity.

An initial investigation of the agnathan immune system began when Robert Good, William Hildemann, and their colleagues reported allograft skin rejection accompanied by immunologic memory in lampreys (Finstad and Good 1964) and

M. Kasahara (✉)
Department of Pathology, Hokkaido University Graduate School of Medicine, North 15 West 7, Sapporo 060-8638, Japan
e-mail: mkasaha@med.hokudai.ac.jp

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_8

hagfish (Hildemann and Thoenes 1969). Subsequent work showed that immunized lampreys and hagfish secrete antigen-specific agglutinins or bactericidal substances into the serum (Acton et al. 1969; Fujii et al. 1979; Linthicum and Hildemann 1970; Litman et al. 1970; Marchalonis and Edelman 1968; Pollara et al. 1970). These observations suggested strongly that, like jawed vertebrates, jawless vertebrates have the adaptive immune system (AIS) equipped with both humoral and cell-mediated immunity.

Assuming that jawless vertebrates have the AIS similar to that of jawed vertebrates, extensive efforts were made from the late 1980s to the early 2000s to search for agnathan genes coding for major histocompatibility complex (MHC) molecules, T-cell receptors (TCRs), B-cell receptors (BCRs), and recombination-activating gene (RAG) enzymes. However, such efforts have met with no success (Suzuki et al. 2004; Uinuk-Ool et al. 2002) (Fig. 1); this was in sharp contrast to the fact that the most primitive class of jawed vertebrates such as sharks, rays, and skates has a complete set of these genes (Du Pasquier 2000; Flajnik 2002; Flajnik and Kasahara 2001; Kasahara et al. 2004; Klein et al. 2000; Litman et al. 1999; Venkatesh et al. 2014) and was in apparent conflict with the observations pointing to the presence of the AIS in jawless vertebrates. This deadlock was eventually resolved by the discovery of agnathan antigen receptors currently known as variable lymphocyte receptors (VLRs) (Pancer et al. 2004a). The identification of VLRs, which are the members of leucine-rich repeat (LRR) family proteins and hence are structurally unrelated to TCRs or BCRs, provided convincing evidence that jawless vertebrates have an alternative form of AIS that does not rely on MHC, TCR, BCR, or RAG molecules (Boehm et al. 2012b; Cooper and Alder

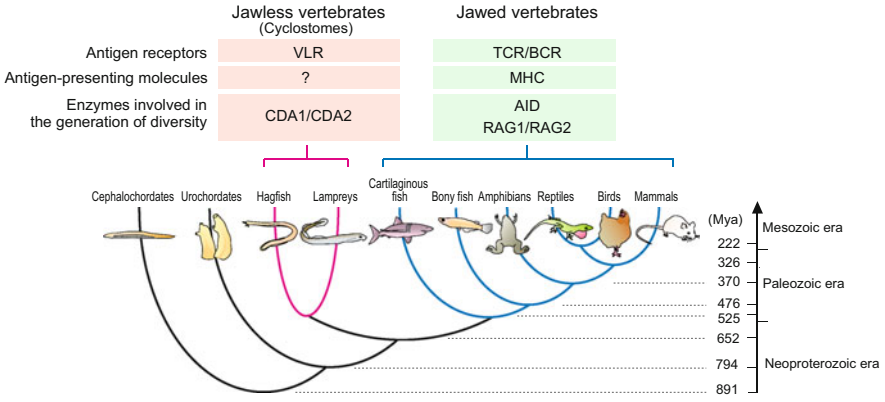


Fig. 1 Evolution of adaptive immunity. Jawless vertebrates do not have T-cell receptors (TCRs), B-cell receptors (BCRs), or major histocompatibility complex (MHC) molecules; they have neither the recombination-activating gene 1 (*RAG1*) nor 2 (*RAG2*). Instead of TCRs and BCRs, they use variable lymphocyte receptors (VLRs) as antigen receptors. Cytidine deaminases of the activation-induced deaminase (AID)/apolipoprotein B mRNA editing complex (APOBEC) are present in all vertebrates. It is not known whether jawless vertebrates have antigen-presenting molecules with functions equivalent to MHC class I or class II molecules. Mya, million years ago

2006; Flajnik 2014; Hirano et al. 2011; Hsu 2011; Kasahara and Sutoh 2014; Pancer and Cooper 2006).

Here, I review the current knowledge about the VLRs of jawless vertebrates, contrast the similarities and differences between the AISs of jawed and jawless vertebrates, and discuss the origin of adaptive immunity.

2 Structure and Function of VLRs

After the initial discovery of the lamprey *VLR* gene (Pancer et al. 2004a), two *VLR* genes were identified in hagfish (Pancer et al. 2005). One of them was clearly orthologous to the lamprey gene; this hagfish gene was named *VLRB* and the other was named *VLRA*. Subsequently, two more lamprey genes were identified: one was named *VLRA* (Rogozin et al. 2007) and the other *VLRC* (Kasamatsu et al. 2010). This initiated a search for a third hagfish gene. When it was identified, it turned out that the newly discovered hagfish gene is actually more closely related to the lamprey *VLRA* gene and that the hagfish gene previously named *VLRA* is the counterpart of the lamprey *VLRC* gene, thus necessitating a change in nomenclature for the hagfish genes. The presence of three corresponding VLRs in lampreys and hagfish indicates that a common ancestor of cyclostomes had VLRA, VLRB, and VLRC receptors.

Among the three receptors, VLRA and VLRC are more closely related to each other than they are to VLRB and occur only as membrane-bound proteins (Hirano et al. 2013; Kasamatsu et al. 2010; Li et al. 2013). These receptors are thought to have functions equivalent to those of gnathostome TCRs (Hirano et al. 2013). On the other hand, the VLRB receptor is an agnathan counterpart of gnathostome BCRs; it is a glycosylphosphatidylinositol-anchored membrane protein that occurs in both membrane-bound and secretory forms (Alder et al. 2005, 2008; Herrin et al. 2008). The domain organization of VLRA, VLRB, and VLRC receptors is basically identical (Fig. 2a). At the N-terminus is a capping module known as an N-terminal LRR (LRRNT). This is followed by an 18-residue N-terminal LRR module (LRR1), multiple 24-residue variable LRRs (LRRV), a 13-residue truncated LRR called the connecting peptide (CP), a C-terminal LRR (LRRCT), and an invariant domain containing a stalk region. The LRRV module has the consensus sequence XLXXLXXLXLXXNXLXXLPXXXFX (where X stands for any amino acid). The most C-terminal LRRV module, known as LRRVe, has a distinct sequence signature (Alder et al. 2005). Sequence diversity is found primarily in the 3′-part of LRRNT (3′-LRRNT), LRR1, LRRV, LRRVe, CP, and the 5′-part of LRRCT (5′-LRRCT).

Structural studies of VLRA and VLRB receptors showed that they bind antigens through their concave surface and a unique hypervariable loop in LRRCT, with the latter playing a major role in antigen binding (Deng et al. 2013). In TCRs and

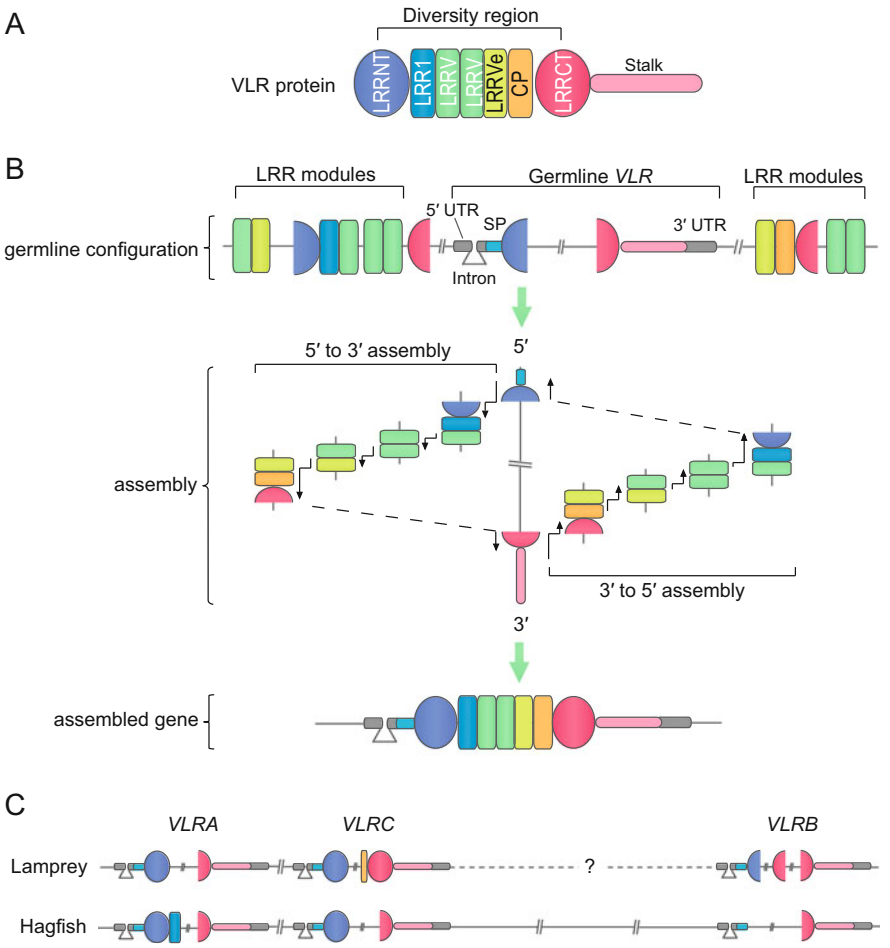


Fig. 2 VLR proteins and genes. (**a**) Domain organization of VLR molecules. Signal peptides are not depicted. All other modules are described in the text. (**b**) Organization of the germline *VLR* locus and gene assembly. The intervening sequence of the germline *VLR* gene is replaced by a gene conversion-like mechanism in a stepwise manner, beginning either from its 5′- or 3′-end, resulting in the formation of an assembled *VLR* gene capable of encoding functional proteins. Modules are color coded as in panel (**a**). *SP* signal peptide, *UTR* untranslated region. (**c**) Chromosomal localization of the three *VLR* genes. In lampreys, germline *VLRA* and *VLRC* genes are situated close to each other. The same is likely the case with hagfish *VLRA* and *VLRC* genes. In hagfish, the *VLRB* gene is located on the same chromosome, albeit physically apart from the *VLRA/C* locus. The order of *VLRA* and *VLRC* genes relative to the *VLRB* gene is drawn arbitrarily. Information is not available on the chromosomal localization of the lamprey *VLRB* gene. Panels (**a**) and (**b**) were modified from Kasahara and Sutoh (2014)

BCRs, the loops formed in the complementarity-determining region 3 (known as CDR3) account for the major fraction of variability in antigen-binding sites. Therefore, both jawed and jawless vertebrates use hypervariable loops as major

structural elements for antigen recognition. Interestingly, VLRC receptors lack a hypervariable loop in the LRRCT region (Holland et al. 2014; Kanda et al. 2014; Kasamatsu et al. 2010). Therefore, it has been suggested that they are likely to recognize antigen in a different manner (Kasamatsu et al. 2010).

In somatic cells other than lymphocytes, the *VLR* locus is in a germline configuration; a large number of LRR-encoding modules (LRRNT, LRR1, LRRV, CP, and LRRCT) are scattered around the germline *VLR* gene that lacks sequences coding for a contiguous variable region and hence is incapable of encoding any protein (Fig. 2b). During the development of lymphocytes, the intervening sequence of the germline *VLR* gene is replaced by a gene conversion-like mechanism in a stepwise fashion, beginning either from its 5′- or 3′-end, by adding flanking LRR-encoding modules, eventually forming a completely assembled *VLR* gene (Alder et al. 2005; Nagawa et al. 2007; Rogozin et al. 2007). Addition of a module appears to occur as an independent event with a constant success rate (Sutoh and Kasahara 2014). This assembly process is thought to be catalyzed by cytidine deaminases of the activation-induced cytidine deaminase (AID)-apolipoprotein B mRNA editing complex (APOBEC) family (Rogozin et al. 2007). Because the sequence of each LRR-encoding module is highly variable, VLRs can generate diversity comparable to that of TCRs and BCRs; the potential repertoire of VLRB receptors has been estimated as ~10^{14} (Alder et al. 2005).

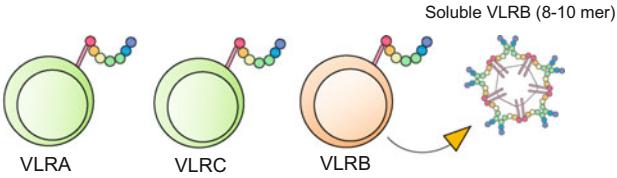
Interestingly, recent work has uncovered that, somewhat reminiscent of the gnathostome *TCRA*/*TCRD* locus, *VLRA* and *VLRC* germline genes are located close to each other in the lamprey genome (Fig. 2c). Consistent with this, assembled lamprey *VLRA* and *VLRC* genes often share variable LRR-encoding modules with identical sequences. Such sharing of modules is also observed in assembled hagfish *VLRA* and *VLRC* genes, suggesting that these two genes are also adjacent to each other in the hagfish genome. In contrast, neither lamprey nor hagfish VLRB receptors share completely identical LRR-encoding modules with the VLRA/VLRC receptors of respective species, suggesting that the germline *VLRB* gene is not situated close to the *VLRA*/*VLRC* locus. This is consistent with our earlier observation that hagfish *VLRB* and *VLRC* (formerly known as *VLRA*) genes yielded well-separated in situ hybridization signals, albeit on the same chromosome (Kasamatsu et al. 2007). An ancient *VLR* gene might have increased its copy number by tandem duplication; subsequently, chromosomal inversion or intrachromosomal translocation might have separated the *VLRB* gene from the *VLRA*/*VLRC* genes, facilitating the functional specialization of the receptor genes. It would be interesting to examine whether lamprey *VLRB* and *VLRA*/*C* genes also map to the same chromosome.

3 Three Lineages of Agnathan Lymphocytes

Similar to gnathostome lymphocytes that clonally express either TCR or BCR with defined specificity, each agnathan lymphocyte expresses only one of the three types of VLRs with a unique sequence (Guo et al. 2009; Hirano et al. 2013; Kishishita et al. 2010; Nagawa et al. 2007; Pancer et al. 2004a). Accumulated evidence indicates that agnathan lymphocytes expressing VLRB receptors resemble gnathostome B cells (Fig. 3). When challenged with antigen, $VLRB^+$ cells differentiate into plasma cell-like cells with well-developed endoplasmic reticulum and secrete VLRB molecules as antibodies (Alder et al. 2005, 2008). Secreted VLRB molecules, which represent the identity of the agglutinins or bactericidal substances described in the 1960s and 1970s, occur as a pentamer or tetramer of dimers, with 8–10 antigen-binding sites, thus resembling the subunit organization of IgM (Herrin et al. 2008). In addition, $VLRB^+$ cells express genes characteristically expressed in gnathostome B cells such as those coding for B-lymphocyte-induced maturation protein 1 (BLIMP-1), B-cell leukemia/lymphoma 6 (BCL6), paired box protein 5 (PAX5), E2A, B-cell adaptor for phosphoinositide 3-kinase (BCAP), and spleen tyrosine kinase (Syk), among others (Guo et al. 2009).

On the other hand, $VLRA^+$ cells and $VLRC^+$ cells resemble gnathostome T cells. Like T cells, they do not secrete receptors as antibodies. $VLRA^+$ cells express transcription factor genes characteristic of T cells such as *GATA2/3* and *HNF1A* (HNF1 homeobox A or T-cell factor 1), and the genes coding for cell surface molecules expressed on T cells such as NOTCH, AHR (aryl hydrocarbon receptor), CCR9/7 (chemokine (C–C motif) receptor 9/7), CD45, and CTLA4 (cytotoxic T-lymphocyte antigen 4) (Guo et al. 2009). $VLRC^+$ cells generally have a gene expression profile similar to that of $VLRA^+$ cells, but differ from $VLRA^+$ cells in that they express genes characteristically expressed in gnathostome γδ T cells, such as *SOX13* (SRY-box 13), *TLR3* (Toll-like receptor 3), *ITGA4* (integrin α4), *ITGB1* (integrin β1), and *ITGAL* (integrin αL) (Hirano et al. 2013). Interestingly, $VLRC^+$ cells occur abundantly in the epidermis and the epithelium of intestine and gill, reminiscent of mouse dendritic epidermal γδ T cells and mammalian intraepithelial lymphocytes (Girardi 2006; Hayday 2000). Also, epithelium-resident $VLRC^+$ cells show restricted antigen receptor diversity, again showing intriguing similarity to mouse dendritic epidermal T cells and intraepithelial lymphocytes (Vantourout and Hayday 2013). TLR3 is responsible for the reaction to viral dsRNA. Stimulation with poly (I:C), a synthetic analog of dsRNA, induced cell proliferation and production of IL-16 by $VLRC^+$ cells (Hirano et al. 2013). In jawed vertebrates, IL-16 is a chemoattractant acting on $CD4^+$ immune cells, including monocytes, eosinophils, and dendritic cells (Cruikshank et al. 2000). These observations suggest that TLR3 on $VLRC^+$ cells may be involved in the immune defense against RNA virus infection through the recruitment of inflammatory cells.

Taken together, accumulated evidence indicates that jawless vertebrates have a lymphocyte subset corresponding to gnathostome B cells and two lymphocyte subsets that might correspond to αβ T cells and γδ T cells, respectively (Guo



Immune function	Cellular immunity		Humoral immunity
Sites of development	Thymoid		Hematopoietic organ
Response to PHA	Yes		No
Response to poly I:C		Strong	Weak
Tissue distribution	Blood	Blood, Epithelium (skin, intestine)	Blood, intestine/typhlosole
Receptor diversity	High	High, but limited in skin	High
Enzymes involved in the generation of diversity	CDA1		CDA2
Cytokines	MIF, IL-17	IL-16, IL-17	IL-8
Cytokine receptors	IL-8R		IL-17R
Transcription factors	GATA2/3, TCF1	SOX13	BLIMP-1 BCL6, PAX5, E2A
Signaling molecules	LAT		BCAP, Syk
Cell surface molecules	NOTCH, AHR, CCR9/7, CD45, CTLA4	TLR3, ITGA4, ITGB1, ITGAL	

Fig. 3 Three lineages of agnathan lymphocytes. VLRA$^+$ cells and VLRC$^+$ cells are T-cell-like, whereas VLRB$^+$ cells are B -cell-like. Soluble VLRB molecules have subunit organization resembling that of IgM. This figure was modified from Kasahara (2013a). AHR, aryl hydrocarbon receptor; BCAP, B-cell adaptor for phosphoinositide 3-kinase; BCL6, B-cell leukemia/lymphoma 6; BLIMP-1, B-lymphocyte-induced maturation protein-1; CCR9/7, chemokine (C–C motif) receptor 9/7; CDA1, cytidine deaminase 1; CDA2, cytidine deaminase 2; CTLA4, cytotoxic T-lymphocyte-associated protein 4; GATA2/3, GATA binding protein 2/3; ITGB1, integrin β1; ITGA4, integrin α4; ITGAL, integrin αL; LAT, linker for activation of T cells; MIF, macrophage migration inhibitory factor; PAX5, paired box protein 5; SOX13, SRY (sex determining region Y)-box 13; Syk, spleen tyrosine kinase; TCF1, T-cell factor 1; and TLR3, Toll-like receptor 3

et al. 2009; Hirano et al. 2013; Kasahara 2013a). Notably, VLRB$^+$ cells express IL-17 receptors and IL-8, whereas IL-17 is expressed by VLRA$^+$ cells and VLRC$^+$ cells and IL-8 receptors by VLRA$^+$ cells. Therefore, VLRA$^+$ or VLRC$^+$ cells might interact with VLRB$^+$ cells in a manner analogous to T-B-cell collaboration in jawed vertebrates.

4 Lymphocyte Development

The discovery that VLRA$^+$ cells are T-cell-like (Guo et al. 2009) initiated a renewed search for a thymus equivalent in lamprey larvae, leading to the identification of the "thymoid" (Bajoghli et al. 2011). "Thymoids" are located at the tips of the gill filaments in the gill basket, occurring not as a single organ but as a constellation of specialized lymphoid tissues with no obvious corticomedullary differentiation. The "thymoid" expresses the chemokine receptor CXCR4 as well as a homolog of FOXN1, a transcription factor specifically expressed in thymic epithelia (Bajoghli et al. 2009). In thymoids, incompletely assembled *VLRA* and *VLRC* genes occur frequently although they are rarely found in the peripheral blood. Furthermore, lymphocytes in the "thymoid" express cytidine deaminase 1 (CDA1), an enzyme of the AID/APOBEC family involved in the assembly of *VLRA* and *VLRC* genes (Guo et al. 2009). These observations suggest that VLRA$^+$ cells and VLRC$^+$ cells develop in the "thymoid" (Fig. 3).

Interestingly, recent work suggests that positive selection, which might be akin to thymic selection in gnathostomes, might operate on developing VLRA$^+$ cells and VLRC$^+$ cells. Thus, the second LRRV module shows distinctive sequence signatures in VLRA and VLRC, but not VLRB receptors, suggesting that lymphocytes with VLRA or VLRC receptors with an appropriate second LRRV module might be selected (Sutoh and Kasahara 2014). Also, the copy number of LRRV modules deviates from the binominal distribution only in VLRA and VLRC receptors, suggesting that VLRA or VLRC receptors with two or less LRRV modules are selected against (Sutoh and Kasahara 2014). Indeed, analysis of VLRC transcripts has shown that the distribution of the copy number of LRR modules differs in "thymoids" and periphery, suggesting that VLRC$^+$ cells undergo selection in the "thymoid" (Holland et al. 2014).

Although T cells develop in the thymus in all classes of jawed vertebrates, the site of B-cell development differs depending on animal classes (Boehm et al. 2012a). Earlier histological studies of lamprey larvae suggested the typhlosole, an inner wall structure of the intestine, and kidney as major hematopoietic organs (Ardavin and Zapata 1987). Cytidine deaminase 2 (CDA2), an enzyme involved in *VLRB* assembly, is expressed by VLRB$^+$ cells in these organs (Bajoghli et al. 2011). Thus, lamprey VLRB$^+$ cells appear to develop in the typhlosole and kidney.

5 Similarities and Differences Between the Two Forms of AIS

The AISs of jawed and jawless vertebrates have much in common despite the fact that structurally unrelated molecules are used as antigen receptors and the major molecular mechanisms involved in the generation of diversity are different (Fig. 4).

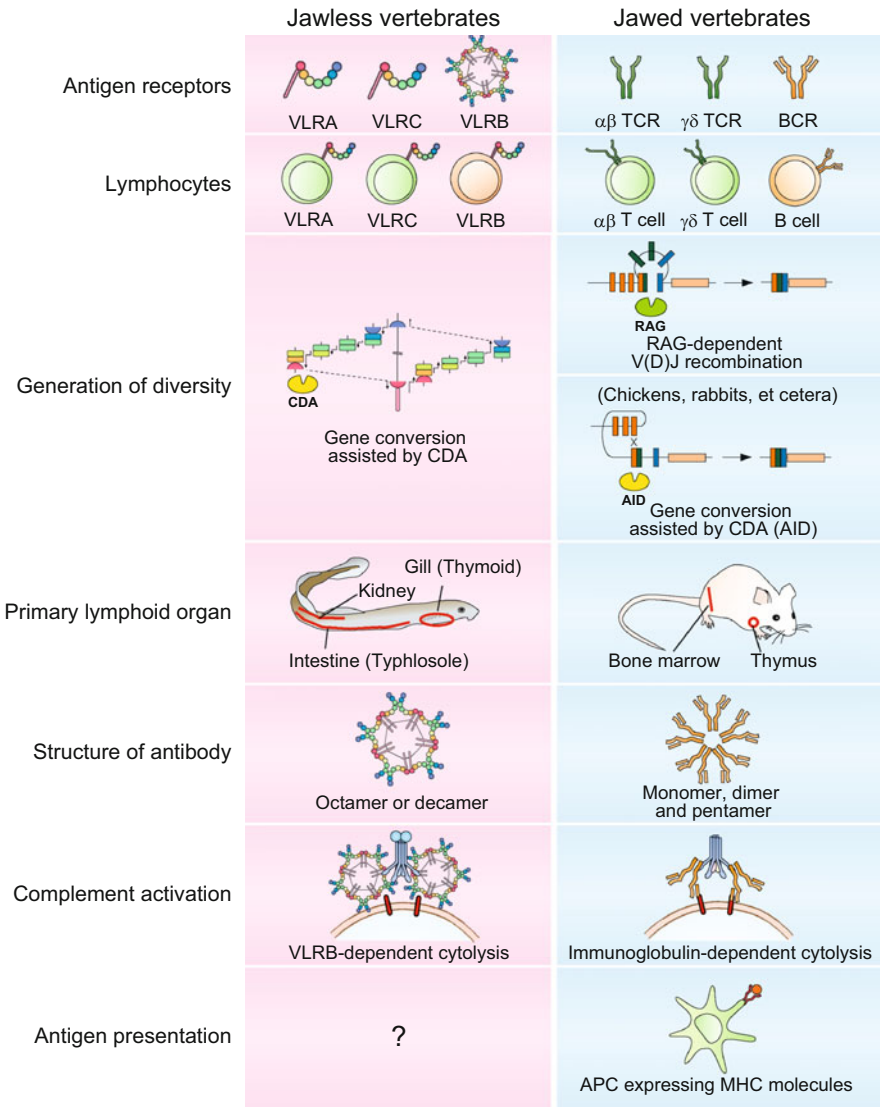


Fig. 4 Similarities and differences between the AISs of jawed and jawless vertebrates. This figure was modified from Kasahara and Sutoh (2014). In the immunoglobulin genes of mammals such as chickens and rabbits, V, D, and J gene segments are rearranged by the RAG enzyme, but their rearranged V-region sequences display little diversity. Diversity is formed mainly by gene conversion and somatic hypermutation mediated by the CDA named AID. *APC* antigen-presenting cells, *CDA* cytidine deaminases

A particularly important point of similarity is that both jawed and jawless vertebrates have lymphocyte-based AISs and apparently have the same lymphocyte lineages. Thus, the two forms of AIS share features such as clonal expression of a

single type of receptors with allelic exclusion, clonal proliferation of antigen-stimulated lymphocytes, and the dichotomy of AIS into thymus (or "thymoid")-dependent and humoral arms (Boehm 2011; Flajnik and Kasahara 2010; Hirano et al. 2011). Given the overall similarity, it is reasonable to assume that the basic design of AIS was established in a common ancestor of jawed and jawless vertebrates. Most likely, a vertebrate ancestor had both T-cell-like and B-cell-like lymphocyte lineages, and jawed and jawless vertebrates co-opted different molecules as their antigen receptors within the context of specialized lymphocyte lineages.

Recent evidence indicates that jawless vertebrates use VLRB molecules for complement activation. In lampreys, binding of specific VLRB antibodies to cell surface antigen recruits C1q-like protein (Matsushita et al. 2004) and MASP-A (mannose-binding lectin-associated protein A), a component of the lectin pathway of complement activation, leading to the lysis of target cells (Wu et al. 2013). Because the depletion of C1q-like protein, VLRB, or C3 inhibited cytolysis, all of these molecules appear to be required for cytolysis. Since hagfish also have this C1q-like protein (Yamaguchi et al. 2014), they likely have a similar system for complement activation. Therefore, contrary to the traditional view that the classical pathway is restricted to jawed vertebrates (Cerenius et al. 2010; Fujita 2002; Nonaka and Kimura 2006), jawless vertebrates appear to have a primitive classical pathway that uses VLRBs instead of immunoglobulins (and a component of the lectin pathway) for its activation, thus further extending the similarity of the two forms of AIS.

A key issue that remains unresolved is whether jawless vertebrates have antigen-presenting molecules with MHC-like functions. Although recombinantly produced VLRA receptors can bind antigen directly in at least some cases (Deng et al. 2010), it is not known whether such direct recognition is the norm in vivo. Failure of immunized VLRA$^+$, but not VLRB$^+$, cells to bind to anthrax spores or *Escherichia coli* suggests that, like TCRs, VLRA receptors might recognize processed rather than native antigen (Guo et al. 2009). In this regard, it is interesting to note that a highly polymorphic membrane protein expressed on hagfish leukocytes was shown to elicit natural VLRB antibody responses and function as a major allogenic antigen (Takaba et al. 2013). This molecule, originally identified by us as NICIR3 (*n*ovel *i*mmunoreceptor tyrosine-based activation motif-*c*ontaining *Ig* superfamily *r*eceptors 3) (Haruta et al. 2006) and named ALA (allogenic leukocyte antigen) by Takaba and colleagues (2013), is predominantly expressed in phagocytic leukocytes and is associated with phagocytosed protein antigens, suggesting a potential role in antigen presentation. Whether NICIR3/ALA has any MHC-like functions remains to be determined.

6 How Well Are Ohnologs Conserved Between Jawed and Jawless Vertebrates? Implications for the Evolution of Adaptive Immunity

It is known that a common ancestor of jawed vertebrates experienced two rounds of whole-genome duplication (WGD) (Furlong and Holland 2002; Kasahara 2007, 2010, 2013b; Ohno 1970, 1999). In contrast, the genomes of basal chordates such as urochordates and cephalochordates show no evidence of such duplication (Dehal and Boore 2005; Putnam et al. 2008). Thus, both the first and second rounds of WGD (commonly abbreviated as "1R" and "2R") must have taken place after the emergence of urochordates and before the divergence of jawed vertebrate classes. Jawless vertebrates occupy an intermediate phylogenetic position between urochordates and jawed vertebrates. Therefore, it is important to know the timing of "1R" and "2R" relative to the emergence of jawless vertebrates. Given that many genes involved in gnathostome adaptive immunity are ohnologs (paralogous genes that arose by two rounds of WGD) (Flajnik and Kasahara 2010; Kasahara 2010), this issue is important for assessing the extent of similarity between the AISs of jawed and jawless vertebrates.

Initially, it was suggested that jawless vertebrates experienced only "1R" and probably additional agnathan-specific WGDs (Fried et al. 2003; Holland et al. 1994; Stadler et al. 2004). Recently, this view was challenged by Kuraku and coworkers who argued that both "1R" and "2R" took place in a common ancestor of jawed and jawless vertebrates after its separation from the invertebrate lineage (Kuraku 2013; Kuraku et al. 2009; Smith et al. 2013). However, more recent analysis of the lamprey *HOX* gene clusters using bacterial artificial chromosome cloning suggests another scenario (Mehta et al. 2013). This analysis, which is so far the most informative, indicates that lampreys have at least six *HOX* clusters, suggesting that the lamprey genome underwent at least three rounds of WGD. Importantly, this work suggested that the WGDs including "1R" and "2R" occurred independently in the lamprey and gnathostome lineages.

Phylogenetic analysis of transcription factors such as Spi and Ikaros indicates that lampreys lack genes orthologous to the members of the gnathostome gene family (Anderson et al. 2001; Haire et al. 2000; Mayer et al. 2002; Shintani et al. 2000). In hagfish, a *GATA* family of genes equally related to *GATA2* and *GATA3* has been identified, suggesting that hagfish lack a genetic ortholog of *GATA3*, an important regulator of T-cell development (Suzuki et al. 2004). These observations are consistent with the hypothesis of independent WGD in jawless vertebrates and argue against the idea that the genomes of jawed and jawless vertebrates have undergone the same two rounds of duplication.

If WGDs occurred independently in the lamprey and gnathostome lineages as suggested by Mehta et al. (2013), their genomes will not share corresponding ohnologs. In this case, the immune functions mediated by ohnologs are unlikely to be well conserved between jawed and jawless vertebrates, although non-orthologous ohnologs may sometimes function as functional homologs.

7 Which Came First, VLR or TCR/BCR?

Conservation of lymphocyte lineages between jawed and jawless vertebrates indicates that specialized lymphocyte subsets emerged in a common vertebrate ancestor (Fig. 5). This suggests that a common ancestor of vertebrates presumably possessed lymphocyte-based adaptive immunity with both humoral and cellular arms. Were the antigen receptors used by such an ancestor VLR-like or TCR/BCR-like? It has been proposed that an ancestor of VLR emerged from a GPIbα-like protein, a component of the platelet glycoprotein receptor complex conserved in all vertebrates (Rogozin et al. 2007). Likewise, it has been proposed that a common ancestor of vertebrates had V-type immunoglobulin-like domains that could be converted to those of TCRs and BCRs (Boehm et al. 2012a; Du Pasquier et al. 2004; Flajnik and

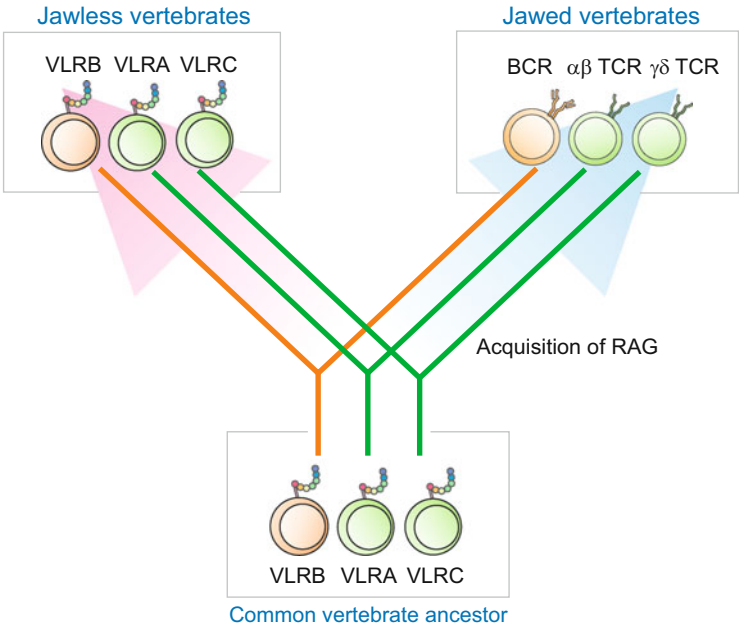


Fig. 5 Evolution of lymphocyte lineages and antigen receptors. Functional specialization of lymphocytes into B-like and T-like cells predated the divergence of jawed and jawless vertebrates. Thus, a common ancestor of vertebrates probably had both humoral and cell-mediated immunity. This ancestor most likely used VLR-like rather than TCR/BCR-like molecules as antigen receptors. Although jawless vertebrates did not change their receptors significantly, a common ancestor of jawed vertebrates, which acquired *RAG* transposons, switched their receptors to the members of the immunoglobulin superfamily. Functional diversification of gnathostome antigen receptors into BCR, αβ TCR, and γδ TCR probably occurred within the context of specialized lymphocyte lineages. As discussed in the text, $VLRA^+$ cells and $VLRC^+$ cells resemble αβ T cells and γδ T cells, respectively, in gene expression profiles. However, direct evidence that $VLRA^+$ cells and $VLRC^+$ cells share common ancestry with αβ T cells and γδ T cells, respectively, is not available. Thus, this part of the figure should be regarded as speculative

Kasahara 2010; Kasahara et al. 2008; Pancer et al. 2004b; Suzuki et al. 2005). Therefore, a common vertebrate ancestor most likely had building blocks for both VLRs and TCRs/BCRs. Hence, it is reasonable to assume that a common ancestor of vertebrates had the potential of developing both VLR-based and TCR/BCR-based adaptive immunity.

A key observation that helps deduce the nature of primordial receptors in a common vertebrate ancestor is that all vertebrates have cytidine deaminases of the AID/APOPEC family (Rogozin et al. 2007). In contrast, *RAG* genes are the product of horizontal gene transfer that took place in a jawed vertebrate lineage, and only jawed vertebrates have functional *RAG* genes (Fugmann 2010; Kapitonov and Jurka 2005). In jawed vertebrates, AID is involved in gene conversion, class switch recombination, and somatic hypermutation (Honjo et al. 2004), thus having functions overlapping with those of CDA1 and CDA2. These considerations suggest that a common vertebrate ancestor probably possessed VLR-like receptors and used AID/APOBEC homologs for generating their diversity. Presumably, VLRs lost the position of antigen receptors in a jawed vertebrate lineage when its common ancestor acquired *RAG* transposons and developed a more efficient and powerful TCR/BCR-based antigen-recognition system (Kasahara and Sutoh 2014; Kato et al. 2012). It is likely that the functional specialization of a gnathostome antigen receptor into TCRs and BCRs and then of the former into αβ TCRs and γδ TCRs occurred under the strong functional constraint of the by then well-established, three lineages of functionally specialized lymphocytes.

8 Concluding Remarks

Accumulated evidence indicates strongly that the two lineages of lymphocytes, one dedicated to humoral immunity and the other to cellular immunity, emerged in a common ancestor of jawed and jawless vertebrates. Surprisingly, like jawed vertebrates, jawless vertebrates have two lineages of T-like lymphocytes. Furthermore, $VLRA^+$ cells and $VLRC^+$ cells resemble gnathostome αβ T cells and γδ T cells, respectively. Is this just a coincidence or does it reflect common ancestry and functional homology? This is certainly one of the key questions that need to be addressed. Specifically, it would be important to examine whether agnathans have antigen-presenting molecules with MHC-like functions and to understand the nature of antigen recognized by $VLRA^+$ cells and $VLRC^+$ cells. When such information becomes available, we should be able to tell more confidently whether $VLRA^+$ cells and $VLRC^+$ cells indeed share common ancestry with αβ T cells and γδ T cells, respectively.

Acknowledgements Experimental work in my laboratory has been supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan. I thank Dr. Yoichi Sutoh for his kind help with the preparation of figures.

References

Acton RT, Weinheimer PF, Hildemann WH, Evans EE (1969) Induced bactericidal response in the hagfish. J Bacteriol 99:626–628

Alder MN, Rogozin IB, Iyer LM, Glazko GV, Cooper MD, Pancer Z (2005) Diversity and function of adaptive immune receptors in a jawless vertebrate. Science 310:1970–1973

Alder MN, Herrin BR, Sadlonova A, Stockard CR, Grizzle WE, Gartland LA, Gartland GL, Boydston JA, Turnbough CL Jr, Cooper MD (2008) Antibody responses of variable lymphocyte receptors in the lamprey. Nat Immunol 9:319–327

Anderson MK, Sun X, Miracle AL, Litman GW, Rothenberg EV (2001) Evolution of hematopoiesis: Three members of the PU.1 transcription factor family in a cartilaginous fish, *Raja eglanteria*. Proc Natl Acad Sci USA 98:553–558

Ardavin CF, Zapata A (1987) Ultrastructure and changes during metamorphosis of the lymphohemopoietic tissue of the larval anadromous sea lamprey *Petromyzon marinus*. Dev Comp Immunol 11:79–93

Bajoghli B, Aghaallaei N, Hess I, Rode I, Netuschil N, Tay BH, Venkatesh B, Yu JK, Kaltenbach SL, Holland ND, Diekhoff D, Happe C, Schorpp M, Boehm T (2009) Evolution of genetic networks underlying the emergence of thymopoiesis in vertebrates. Cell 138:186–197

Bajoghli B, Guo P, Aghaallaei N, Hirano M, Strohmeier C, McCurley N, Bockman DE, Schorpp M, Cooper MD, Boehm T (2011) A thymus candidate in lampreys. Nature 470:90–94

Boehm T (2011) Design principles of adaptive immune systems. Nat Rev Immunol 11:307–317

Boehm T, Iwanami N, Hess I (2012a) Evolution of the immune system in the lower vertebrates. Annu Rev Genomics Hum Genet 13:127–149

Boehm T, McCurley N, Sutoh Y, Schorpp M, Kasahara M, Cooper MD (2012b) VLR-based adaptive immunity. Annu Rev Immunol 30:203–220

Cerenius L, Kawabata S, Lee BL, Nonaka M, Soderhall K (2010) Proteolytic cascades and their involvement in invertebrate immunity. Trends Biochem Sci 35:575–583

Cooper MD, Alder MN (2006) The evolution of adaptive immune systems. Cell 124:815–822

Cruikshank WW, Kornfeld H, Center DM (2000) Interleukin-16. J Leukoc Biol 67:757–766

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3, e314

Deng L, Velikovsky CA, Xu G, Iyer LM, Tasumi S, Kerzic MC, Flajnik MF, Aravind L, Pancer Z, Mariuzza RA (2010) A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey. Proc Natl Acad Sci USA 107:13408–13413

Deng L, Luo M, Velikovsky A, Mariuzza RA (2013) Structural insights into the evolution of the adaptive immune system. Annu Rev Biophys 42:191–215

Du Pasquier L (2000) The phylogenetic origin of antigen-specific receptors. Curr Top Microbiol Immunol 248:160–185

Du Pasquier L, Zucchetti I, De Santis R (2004) Immunoglobulin superfamily receptors in protochordates: before RAG time. Immunol Rev 198:233–248

Finstad J, Good RA (1964) The evolution of the immune response. III. Immunologic responses in the lamprey. J Exp Med 120:1151–1168

Flajnik MF (2002) Comparative analyses of immunoglobulin genes: surprises and portents. Nat Rev Immunol 2:688–698

Flajnik MF (2014) Re-evaluation of the immunological big bang. Curr Biol 24:R1060–R1065

Flajnik MF, Kasahara M (2001) Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity 15:351–362

Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11:47–59

Fried C, Prohaska SJ, Stadler PF (2003) Independent *Hox*-cluster duplications in lampreys. J Exp Zool B Mol Dev Evol 299:18–25

Fugmann SD (2010) The origins of the *Rag* genes—from transposition to V(D)J recombination. Semin Immunol 22:10–16

Fujii T, Nakagawa H, Murakawa S (1979) Immunity in lamprey. I. Production of haemolytic and haemagglutinating antibody to sheep red blood cells in Japanese lampreys. Dev Comp Immunol 3:441–451
Fujita T (2002) Evolution of the lectin-complement pathway and its role in innate immunity. Nat Rev Immunol 2:346–353
Furlong RF, Holland PW (2002) Were vertebrates octoploid? Philos Trans R Soc Lond B Biol Sci 357:531–544
Girardi M (2006) Immunosurveillance and immunoregulation by γδ T cells. J Invest Dermatol 126:25–31
Guo P, Hirano M, Herrin BR, Li J, Yu C, Sadlonova A, Cooper MD (2009) Dual nature of the adaptive immune system in lampreys. Nature 459:796–802
Haire RN, Miracle AL, Rast JP, Litman GW (2000) Members of the Ikaros gene family are present in early representative vertebrates. J Immunol 165:306–312
Haruta C, Suzuki T, Kasahara M (2006) Variable domains in hagfish: NICIR is a polymorphic multigene family expressed preferentially in leukocytes and is related to lamprey *TCR-like*. Immunogenetics 58:216–225
Hayday AC (2000) γδ cells: a right time and a right place for a conserved third way of protection. Annu Rev Immunol 18:975–1026
Herrin BR, Alder MN, Roux KH, Sina C, Ehrhardt GR, Boydston JA, Turnbough CL Jr, Cooper MD (2008) Structure and specificity of lamprey monoclonal antibodies. Proc Natl Acad Sci USA 105:2040–2045
Hildemann WH, Thoenes GH (1969) Immunological responses of Pacific hagfish. I. Skin transplantation immunity. Transplantation 7:506–521
Hirano M, Das S, Guo P, Cooper MD (2011) The evolution of adaptive immunity in vertebrates. Adv Immunol 109:125–157
Hirano M, Guo P, McCurley N, Schorpp M, Das S, Boehm T, Cooper MD (2013) Evolutionary implications of a third lymphocyte lineage in lampreys. Nature 501:435–438
Holland PWH, Garcia-Fernandez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. Development 1994(Suppl):125–133
Holland SJ, Gao M, Hirano M, Iyer LM, Luo M, Schorpp M, Cooper MD, Aravind L, Mariuzza RA, Boehm T (2014) Selection of the lamprey VLRC antigen receptor repertoire. Proc Natl Acad Sci USA 111:14834–14839
Honjo T, Muramatsu M, Fagarasan S (2004) AID: How does it aid antibody diversity? Immunity 20:659–668
Hsu E (2011) The invention of lymphocytes. Curr Opin Immunol 23:156–162
Janvier P (2006) Palaeontology: modern look for ancient lamprey. Nature 443:921–924
Kanda R, Sutoh Y, Kasamatsu J, Maenaka K, Kasahara M, Ose T (2014) Crystal structure of the lamprey variable lymphocyte receptor C reveals an unusual feature in its N-terminal capping module. PLoS One 9, e85875
Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 3, e181
Kasahara M (2007) The 2R hypothesis: an update. Curr Opin Immunol 19:547–552
Kasahara M (2010) Genome duplication and T cell immunity. Prog Mol Biol Transl Sci 92:7–36
Kasahara M (2013a) *Deja vu*: the identity of a third lineage of lymphocytes in lampreys. Immunol Cell Biol 91:599–600
Kasahara M (2013b) Impact of whole-genome duplication on vertebrate development and evolution. Semin Cell Dev Biol 24:81–82
Kasahara M, Sutoh Y (2014) Two forms of adaptive immunity in vertebrates: similarities and differences. Adv Immunol 122:59–90
Kasahara M, Suzuki T, DuPasquier L (2004) On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. Trends Immunol 25:105–111
Kasahara M, Kasamatsu J, Sutoh Y (2008) Two types of antigen receptor systems in vertebrates. Zool Sci 25:969–975

Kasamatsu J, Suzuki T, Ishijima J, Matsuda Y, Kasahara M (2007) Two variable lymphocyte receptor genes of the inshore hagfish are located far apart on the same chromosome. Immunogenetics 59:329–331

Kasamatsu J, Sutoh Y, Fugo K, Otsuka N, Iwabuchi K, Kasahara M (2010) Identification of a third variable lymphocyte receptor in the lamprey. Proc Natl Acad Sci USA 107:14304–14308

Kato L, Stanlie A, Begum NA, Kobayashi M, Aida M, Honjo T (2012) An evolutionary view of the mechanism for immune and genome diversity. J Immunol 188:3559–3566

Kishishita N, Matsuno T, Takahashi Y, Takaba H, Nishizumi H, Nagawa F (2010) Regulation of antigen-receptor gene assembly in hagfish. EMBO Rep 11:126–132

Klein J, Sato A, Mayer WE (2000) Jaws and AIS. In: Kasahara M (ed) Major histocompatibility complex: evolution, structure, and function. Springer, Tokyo, pp 3–26

Kuraku S (2013) Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. Semin Cell Dev Biol 24:119–127

Kuraku S, Meyer A, Kuratani S (2009) Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? Mol Biol Evol 26:47–59

Li J, Das S, Herrin BR, Hirano M, Cooper MD (2013) Definition of a third *VLR* gene in hagfish. Proc Natl Acad Sci USA 110:15013–15018

Linthicum DS, Hildemann WH (1970) Immunologic responses of Pacific hagfish. III. Serum antibodies to cellular antigens. J Immunol 105:912–918

Litman GW, Finstad FJ, Howell J, Pollara BW, Good RA (1970) The evolution of the immune response. III. Structural studies of the lamprey immunoglobulin. J Immunol 105:1278–1285

Litman GW, Anderson MK, Rast JP (1999) Evolution of antigen binding receptors. Annu Rev Immunol 17:109–147

Marchalonis JJ, Edelman GM (1968) Phylogenetic origins of antibody structure. III. Antibodies in the primary immune response of the sea lamprey, *Petromyzon marinus*. J Exp Med 127:891–914

Matsushita M, Matsushita A, Endo Y, Nakata M, Kojima N, Mizuochi T, Fujita T (2004) Origin of the classical complement pathway: lamprey orthologue of mammalian C1q acts as a lectin. Proc Natl Acad Sci USA 101:10127–10131

Mayer WE, O'hUigin C, Tichy H, Terzic J, Saraga-Babic M (2002) Identification of two Ikaros-like transcription factors in lamprey. Scand J Immunol 55:162–170

Mehta TK, Ravi V, Yamasaki S, Lee AP, Lian MM, Tay BH, Tohari S, Yanai S, Tay A, Brenner S, Venkatesh B (2013) Evidence for at least six *Hox* clusters in the Japanese lamprey (*Lethenteron japonicum*). Proc Natl Acad Sci USA 110:16044–16049

Nagawa F, Kishishita N, Shimizu K, Hirose S, Miyoshi M, Nezu J, Nishimura T, Nishizumi H, Takahashi Y, Hashimoto S, Takeuchi M, Miyajima A, Takemori T, Otsuka AJ, Sakano H (2007) Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. Nat Immunol 8:206–213

Nonaka M, Kimura A (2006) Genomic view of the evolution of the complement system. Immunogenetics 58:701–713

Ohno S (1970) Evolution by gene duplication. Springer, New York

Ohno S (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. Semin Cell Dev Biol 10:517–522

Pancer Z, Cooper MD (2006) The evolution of adaptive immunity. Annu Rev Immunol 24:497–518

Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD (2004a) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 430:174–180

Pancer Z, Mayer WE, Klein J, Cooper MD (2004b) Prototypic T cell receptor and CD4-like coreceptor are expressed by lymphocytes in the agnathan sea lamprey. Proc Natl Acad Sci USA 101:13273–13278

Pancer Z, Saha NR, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, Cooper MD (2005) Variable lymphocyte receptors in hagfish. Proc Natl Acad Sci USA 102:9224–9229

Pollara B, Litman GW, Finstad J, Howell J, Good RA (1970) The evolution of the immune response. VII. Antibody to human "O" cells and properties of the immunoglobulin in lamprey. J Immunol 105:738–745

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071

Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, Aravind L, Pancer Z (2007) Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat Immunol 8:647–656

Shimeld SM, Donoghue PC (2012) Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). Development 139:2091–2099

Shintani S, Terzic J, Sato A, Saraga-Babic M, O'hUigin C, Tichy H, Klein J (2000) Do lampreys have lymphocytes? The Spi evidence. Proc Natl Acad Sci USA 97:7417–7422

Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, Morgan JR, Buxbaum JD, Sachidanandam R, Sims C, Garruss AS, Cook M, Krumlauf R, Wiedemann LM, Sower SA, Decatur WA, Hall JA, Amemiya CT, Saha NR, Buckley KM, Rast JP, Das S, Hirano M, McCurley N, Guo P, Rohner N, Tabin CJ, Piccinelli P, Elgar G, Ruffier M, Aken BL, Searle SM, Muffato M, Pignatelli M, Herrero J, Jones M, Brown CT, Chung-Davidson YW, Nanlohy KG, Libants SV, Yeh CY, McCauley DW, Langeland JA, Pancer Z, Fritzsch B, de Jong PJ, Zhu B, Fulton LL, Theising B, Flicek P, Bronner ME, Warren WC, Clifton SW, Wilson RK, Li W (2013) Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat Genet 45:415–421

Stadler PF, Fried C, Prohaska SJ, Bailey WJ, Misof BY, Ruddle FH, Wagner GP (2004) Evidence for independent *Hox* gene duplications in the hagfish lineage: a PCR-based gene inventory of *Eptatretus stoutii*. Mol Phylogenet Evol 32:686–694

Sutoh Y, Kasahara M (2014) Copy number and sequence variation of leucine-rich repeat modules suggests distinct functional constraints operating on variable lymphocyte receptors expressed by agnathan T cell-like and B cell-like lymphocytes. Immunogenetics 66:403–409

Suzuki T, Shin IT, Kohara Y, Kasahara M (2004) Transcriptome analysis of hagfish leukocytes: a framework for understanding the immune system of jawless fishes. Dev Comp Immunol 28:993–1003

Suzuki T, Shin-I T, Fujiyama A, Kohara Y, Kasahara M (2005) Hagfish leukocytes express a paired receptor family with a variable domain resembling those of antigen receptors. J Immunol 174:2885–2891

Takaba H, Imai T, Miki S, Morishita Y, Miyashita A, Ishikawa N, Nishizumi H, Sakano H (2013) A major allogenic leukocyte antigen in the agnathan hagfish. Sci Rep 3:1716

Uinuk-Ool T, Mayer WE, Sato A, Dongak R, Cooper MD, Klein J (2002) Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. Proc Natl Acad Sci USA 99:14356–14361

Vantourout P, Hayday A (2013) Six-of-the-best: unique contributions of γδ T cells to immunology. Nat Rev Immunol 13:88–100

Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay BH, Tohari S, Kong KW, Ho S, Lorente-Galdos B, Quilez J, Marques-Bonet T, Raney BJ, Ingham PW, Tay A, Hillier LW, Minx P, Boehm T, Wilson RK, Brenner S, Warren WC (2014) Elephant shark genome provides unique insights into gnathostome evolution. Nature 505:174–179

Wu F, Chen L, Liu X, Wang H, Su P, Han Y, Feng B, Qiao X, Zhao J, Ma N, Liu H, Zheng Z, Li Q (2013) Lamprey variable lymphocyte receptors mediate complement-dependent cytotoxicity. J Immunol 190:922–930

Yamaguchi T, Takamune K, Kondo M, Takahashi Y, Kato-Unoki Y, Nakao M, Sano N, Fujii T (2014) Hagfish C1q: Its unique binding property. Dev Comp Immunol 43:47–53

Antibody Repertoires in Fish

Eva Bengtén and Melanie Wilson

> *What attracted me to immunology was that the whole thing seemed to revolve around a very simple experiment: take two different antibody molecules and compare their primary sequences. The secret of antibody diversity would emerge from that. Fortunately at the time I was sufficiently ignorant of the subject not to realize how naive I was being.*
>
> César Milstein (1984)

Abstract As in mammals, cartilaginous and teleost fishes possess adaptive immune systems based on antigen recognition by immunoglobulins (Ig), T cell receptors (TCR), and major histocompatibility complex molecules (MHC) I and MHC II molecules. Also it is well established that fish B cells and mammalian B cells share many similarities, including Ig gene rearrangements, and production of membrane Ig and secreted Ig forms. This chapter provides an overview of the IgH and IgL chains in cartilaginous and bony fish, including their gene organizations, expression, diversity of their isotypes, and development of the primary repertoire. Furthermore, when possible, we have included summaries of key studies on immune mechanisms such as allelic exclusion, somatic hypermutation, affinity maturation, class switching, and mucosal immune responses.

1 Introduction

The human immune system is capable of synthesizing many millions of different antibodies that are in place before it encounters any antigen. This enormous diversity and preparedness intrigued scientists for more than a century. It is now 40 years since Susumu Tonegawa discovered "the genetic principle for generation of antibody diversity" for which he was later awarded the Nobel Prize in Medicine in 1987. His work elegantly demonstrated that the exons encoding the variable

E. Bengtén (✉) • M. Wilson (✉)
Department of Microbiology and Immunology, University of Mississippi Medical Center, Jackson, MS 39216-4505, USA
e-mail: ebengten@umc.edu; mwilson@umc.edu

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_9

(V) regions of immunoglobulin (Ig) heavy (H) and light (L) chain genes are assembled by somatic gene rearrangement from an array of multiple V, (D), and J segments during B cell development. This process allows each B cell to produce its unique antibody. Later after B cell activation and clonal selection, the specificity of the V regions is fine-tuned through somatic hypermutation to increase their affinity. This process is initiated by the enzyme activation-induced cytidine deaminase (AID) that also mediates class switch recombination, which generates isotypes of different effector functions.

Not all tetrapods, however generate their Ig repertoire in this manner which is so well defined in humans and mice (see Butler 2006). In birds, rabbits, cows, and sheep, there is little or no germline diversity, and the initial B cell repertoire is very similar in most immature B cells. In these species, the immature B cells must move to a specialized microenvironment, e.g., the bursa of Fabricius in the chicken and the appendix in rabbits. There the Ig genes undergo somatic diversification. In birds and rabbits, this process occurs mainly by gene conversion; short sequences are replaced with sequences from an upstream pseudogene V gene segment. Gene conversion is also mediated by AID. Comparatively, in sheep and cows, this diversification takes place in the ileal Peyer's patches and is mainly driven by T-independent somatic hypermutation prior to antigen stimulation. Here in this chapter we examine to what extent the "textbook knowledge" on how germline diversity, combinatorial diversity through VDJ recombination, junctional diversity (nucleotide insertions and imprecise joining), and random association of IgH and IgL chains contribute to the primary repertoire in cartilaginous and bony fish and how this repertoire is further diversified after B cell activation. Our goal is to present an overview of the different Ig isotypes and focus on the major points by presenting data from what we consider to be key experiments from representative elasmobranch and teleost species. Also, please keep in mind that when comparing the "immunologic status" used by elasmobranchs and teleosts while there are some major differences, it seems that quite a few simply involve variations on a theme.

2 The Cartilaginous Fish

2.1 Why We Study Them

The cartilaginous fish (Chondrichthyes) arose from a common ancestor approximately 450 million years ago (Smith et al. 2015) and are divided into two subclasses the Holocephali or chimeras and the more "modern" Elasmobranchii, i.e., the sharks, skates, and rays. Importantly, these fish, represented by approximately 1200 species, are the earliest vertebrates with an adaptive immune system similar to the one in mammals, that is, one based on antigen recognition by immunoglobulins (Ig), T cell receptors (TCR), and major histocompatibility complex (MHC) I and MHC II molecules (reviewed in Flajnik and Du Pasquier 2013; Pettinello and

Dooley 2014). Even though IgH and IgL gene clusters have been sequenced in several shark species and rays, currently how their immune system functions is by far best understood in the nurse shark, *Ginglymostoma cirratum*. This is largely due to work pioneered by Bill Clem and Parker Small in the 1970s (Clem et al. 1967; Clem and Leslie 1971; Fidler et al. 1969; Klapper et al. 1971; Leslie and Clem 1970; McCumber and Clem 1976; Small et al. 1970) and the more recent and excellent studies carried out in the laboratories of Martin Flajnik, Ellen Hsu, and their collaborators, which have provided us with the most complete picture of how B cells function in any fish species (Dooley and Flajnik 2006; Hsu et al. 2006; Pettinello and Dooley 2014; Smith et al. 2015).

2.2 Their IgH Chains

The IgH chain genes in all cartilaginous fish studied to date are arranged in distinct clusters (or mini-loci) as opposed to translocon organization of multiple V, D, and J gene segments followed by the CH region genes encoding the different IgH isotypes, as defined in mammals and found in other vertebrates. The clusters can be repeated as many as a hundred times in the genome, depending upon the species, and each spans approximately 10–20 kb. Each cluster consists of one V, two or three Ds, and one J, followed by one set of constant region exons for a specific isotype. These IgH gene clusters are scattered over long distances on the chromosome, as much as 120 kb apart. Moreover, the clusters encoding the different IgH chain isotypes are not linked on one chromosome, and in some species, the same isotypes can be found on different chromosomes. Also, the number of clusters varies both between the isotypes and the species (see below; see Table 1). For example, it is estimated that there are approximately 100 independent clusters encoding the Igμ heavy chain in the spiny dogfish, *Squalus acanthias*, as compared to 15 in the nurse shark (Lee et al. 2008; Smith et al. 2012). Another unique feature of cartilaginous immunoglobulin genes is that the V, D, and J segments are often found germline-joined either completely (VDDJ) or to varying degrees (VD–DJ, VDD–J, V–DDJ; Hinds and Litman 1986; Lee et al. 2008; Litman et al. 1993; Yoder and Litman 2000). To date three Ig classes have been identified in cartilaginous fish, IgM, IgW, and IgNAR, and while all three of these IgH isotypes are present in the elasmobranchs, only IgM is found in the chimeras (Holocephali).

2.3 IgM

Serum IgM was first isolated from the spiny dogfish 50 years ago where it was demonstrated to occur in two forms, a 19S pentamer and a 7S monomer (Fig. 1, Marchalonis and Edelman 1965, 1966). Later, as shown by Clem and Small in their work with adult lemon sharks, *Negaprion brevirostris*, and nurse sharks, it was

Table 1 IgH chain isotypes expressed in cartilaginous fish

	IgM clusters	IgM VH families	IgW clusters	IgW VH families	IgNAR clusters
Nurse Shark *Ginglymostoma cirratum*	12[a] nonrearranged 1[a] germline-joined VDDJ–C	1	2[a]nonrearranged 4[a] germline-joined VD–D–J–C 2[a] germline-joined VDD–J–C	4	3[a] nonrearranged 1[a] germline-joined VDD–D–J
Horned shark *Heterodontus francisci*	~100[b] germline-joined ~100[b] nonrearranged	2	7[b] nonrearranged	1	Yes, indirect evidence
Sandbar shark *Carcharhinus plumbeus*	cDNA	6[c]	Yes, number unknown	1	
Spiny dogfish *Squalus acanthias*	100[b]	1	10–20[b]	1	10–20[b] VD–D–J–C
Little skate *Raja erinacea*	Germline-joined Nonrearranged	1	~6[b] nonrearranged	1	
Spotted ratfish *Hydrolagus colliei*	5–10 Germline-joined 50 nonrearranged 90 % pseudogenes	2	ND		ND
Elephant shark *Callorhinchus milii*	~30	2	ND		ND

ND not documented

[a]Number is based on genomic sequencing

[b]Number of clusters is estimated by Southern blot

[c]Defined by the authors as sharing >80 % nucleotide identity

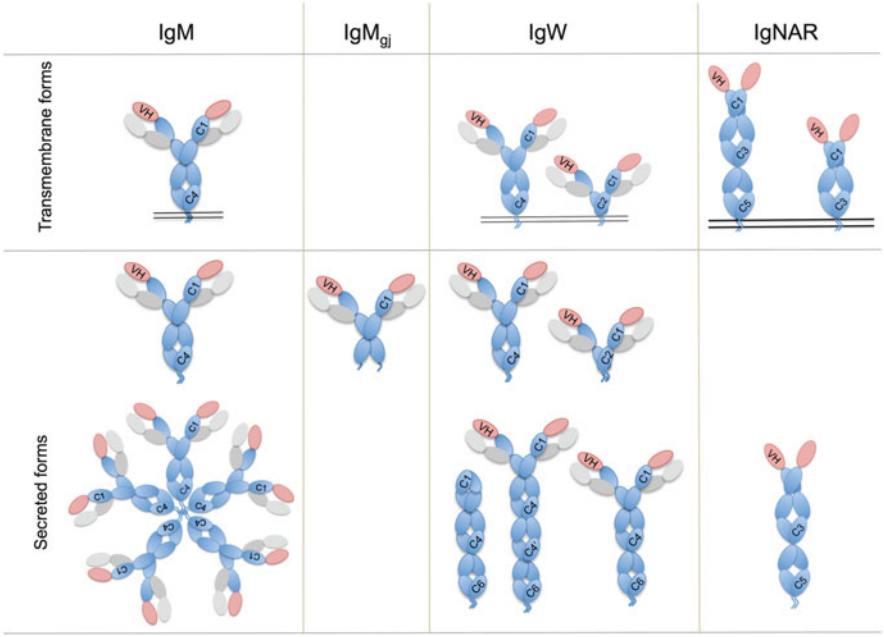


Fig. 1 Schematic representations of the transmembrane and secreted forms of the different shark immunoglobulins. IgH C domains are *dark shaded*, and IgL C domains are *lighter shaded*

found that both forms were expressed at roughly the same levels, and together the 19S and 7S IgM forms were secreted at a rate of 100 mg/day/kg and made up half (>20 mg/ml) of the total serum protein (Clem and Small 1967; Small et al. 1970). Also, the 7S form was predicted to function similarly to IgG since it was (1) present both intra- and extravascularly, (2) not produced by neonatal sharks, and (3) found represented in specific antibodies produced considerably later in the immune response as compared to the 19S pentameric form. Importantly, it was proven in vivo that 19S shark IgM does not degrade into 7S shark IgM nor is the 7S monomer an extracellular precursor for the 19S pentamer. Thus, it was proposed that the two forms are produced independently, perhaps by different B cells. This work as well as studies from groups within other disciplines led to the hypothesis that since sharks lack albumin, IgM takes on the role to regulate osmotic pressure in the blood (Flajnik and Rumfelt 2000). Almost 30 years later, this suggestion that pentameric IgM serves as a first line of defense while 7S is produced later presumably by a T-dependent immune response was confirmed elegantly by Flajnik et al. (Dooley and Flajnik 2005; Rumfelt et al. 2002, and see below).

The first immunoglobulin genes sequenced in an ectothermic vertebrate were isolated from a horned shark, *Heterodontus francisci*, cDNA library by cross-hybridization screening with a mouse VH probe (Litman et al. 1985), and this work quickly changed the field of comparative immunology. Briefly, for the sharks, it was subsequently estimated that the horn shark genome contained approximately

200 VH genes with 50 % of them joined in germline, which may (or may not) limit the diversity of the primary repertoire (Kokubu et al. 1987, 1988; reviewed in Litman et al. 1999). Today even though VH genes in most cartilaginous fish have diverse CDR1 and CDR2 sequences, they can be classified in one family based on the sharing of >80 % nucleotide identity (Kokubu et al. 1988; Litman et al. 1993). An exception in the literature is from a study using sandbar shark, *Carcharhinus plumbeus*, where the VHs were divided into six closely related families exhibiting slightly less than 80 % nucleotide identities with each other (Shen et al. 1996). Today we would consider these different Vμ genes to belong to subfamilies. More recently, because of the limited germline diversity and the prediction that rearrangement would occur only within clusters, Hsu and Flajnik set out to determine to what extent junctional diversity contributes to the primary repertoire using nurse sharks, as their model. By sequencing one locus with a unique restriction site in the Cμ2 exon and comparing its sequence to the sequences of 22 cDNAs obtained from the epigonal organ (the primary lymphoid tissue) of the same individual shark, it was demonstrated that considerable diversity occurred at the junctions in CDR3. Both nucleotide deletions and additions of N-nucleotides, on average of 15 nucleotides per sequence, were found, as well as D segments used in all three reading frames. In addition, the D2 segment was found in reverse orientation in a few cDNAs (Malecek et al. 2005), i.e. D2 is flanked by recombination signal sequences (RSS) with 12 bp spacers at both the 5′ and 3′ end, which allows for recombination through inversion (Fig. 2).

Further sequencing in nurse sharks revealed that each individual had from 9 to 12 functional Igμ loci, 3–4 nonfunctional loci due to deletions in Igμ or a J gene segment, and one mini-locus that contains a germline-rearranged VDJ with a short CDR3 and a short Igμ gene missing the Cμ2 exon (Lee et al. 2008). Importantly, having determined the sequence of all the nurse shark Igμ germline mini-loci, the authors were then able to match the VDJ of 135 cDNA sequences to a specific IgM

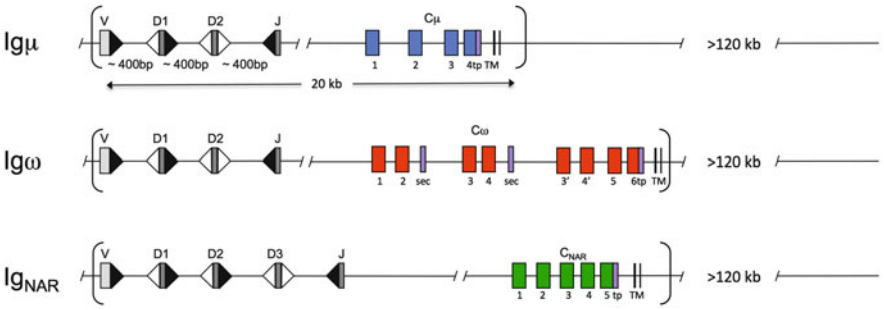


Fig. 2 Schematic representations of elasmobranch IgH clusters. The VH, D, JH and Cμ, Cω, and C_{NAR} gene segments are *color shaded* and *labeled*. The C exons are *numbered*. *Black triangles* indicate RSS with 22- or 23-bp spacers, and *white triangles* indicate RSS with 12-bp spacers. The TM exons are *labeled*, the secretory exon is labeled "sec", and "tp" indicates a secreted tailpiece. Note that different germline-joined configurations not shown here exist for each isotype. Drawing is not to scale

cluster. In other words, no rearrangements occurred between clusters. In addition, sequencing of the mini-locus confirmed that the truncated form of IgM, termed IgM_{1gj}, isolated from neonatal nurse shark plasma by Rumfelt et al. in 2001 and also identified as cDNA, originated from this cluster (Rumfelt et al. 2001). Finally, phylogenetic analyses and domain-by-domain nucleotide comparisons from two unrelated nurse sharks revealed that all five IgH exons (VH, Cμ1, Cμ2, Cμ3, Cμ4) evolved at a similar rate. However, the ratio of non-synonymous to synonymous nucleotide substitutions suggested that both the V and the Cμ2 exons were under strong positive selection for amino acid diversity as compared to the other Cμ exons. Thus, it was speculated that the differences in Cμ2 domains may be associated with different effector functions, for example, ability to bind to FCR (Lee et al. 2008). Here it can also be speculated that the Cμ4 exon is under strong stabilizing selection, perhaps because of the Cμ4 domain interaction with the CD79 B cell signaling molecules.

The holocephalan elephant shark, *Callorhinchus milii*, is said to have the slowest evolving vertebrate genome based on the extensive conserved synteny observed in a recent whole-genome analysis (Venkatesh et al. 2014). In this study investigators found only IgM genes arranged in a cluster-type organization, and there were no evidence for genes encoding the IgW and IgNAR isotypes. The VH genes were similar to those of the spotted ratfish, *Hydrolagus colliei*. Ten conventional VH genes and one single chain V_HH-like gene were reported. This single chain gene resembles the one V_HH-like gene found in *Hydrolagus*, which is spliced to an Igμ gene that lacks a true Cμ1 exon and the amino acids that are necessary for association with IgL chains. Instead the first CH exon resembles a Cμ2 exon and appears to be the result of an exon duplication event (Rast et al. 1998). While reminiscent of the camelid IgG V_HH genes, these unique V_HH-like genes would allow for the production of single-chained IgM antibodies. These types of V_HH-like Igμ genes have not been found in elasmobranchs.

2.4 IgW

The second class of serum Ig in cartilaginous fish is IgW, and this immunoglobulin was previously termed IgNARC, IgX, and IgR in different species of sharks and skates (reviewed in Flajnik and Rumfelt 2000). Briefly, IgR was identified in the serum of the skate *Raja kenojei* as a low molecular weight (LMW) short Ig antigenically distinct from IgM (HMW), and through the use of polyclonal antisera combined with immunohistochemistry, it was shown that this LMW Ig was produced independently from IgM in cells most frequently found in the gut and spleen and to a lesser extent in the Leydig organ, epigonal organ, and the liver (Kobayashi et al. 1984; Tomonaga and Kobayashi 1985; Tomonaga et al. 1984). Later both long and short forms of this second isotype were found as cDNAs in a second species of skate, *Raja eglanteria*, and this isotype was termed IgX (Anderson et al. 1999; Harding et al. 1990). Similarly, longer forms of cDNAs related to IgX were identified at the cDNA level in nurse shark and sandbar shark as IgNARC and

IgW, respectively (Berstein et al. 1996; Greenberg et al. 1996). Interestingly, IgNARC was first described as a chimeric Ig class since the first CH domain showed a clear relationship with Cμ1 domains of tetrapods and sharks. Today it is clear that IgW is orthologous to IgD described in bony fish and tetrapods (Ohta and Flajnik 2006), where it is found encoded immediately 3′ the Igμ gene. This location was important in the classification of IgD in teleost since the Cμ1 exon is used to encode the first CH region of teleost IgD though alternative splicing (Wilson et al. 1997). Even so, since the elasmobranch IgW genes are expressed independently of IgM, it is justified, i.e. makes sense, to use the IgW nomenclature (Zhang et al. 2013).

In the past 10 years, sequencing of cDNAs and the Igω loci in the nurse shark has demonstrated that in contrast to IgM, multiple splice forms of IgW other than the original long form (6 Cω domains) occur. In 2004, Rumfelt et al. identified both transmembrane and secreted IgW forms with four and two Cω domains (Rumfelt et al. 2004b). The shorter secreted IgW transcripts ended with Cω2 followed by a unique secretory sequence containing either 4 or 5 cysteine residues. This secretory exon was predicted to be located directly 3′ of Cω2. As for tissue expression, northern blots show that both long and short transmembrane and secreted IgW forms are expressed in the spleen; however in the pancreas, the dominant membrane form is the long IgW, and the only secreted IgW form is the short form. Recently, the nurse shark Igω loci were sequenced, and this combined with cDNA analyses revealed Igω splicing patterns were even more complex. In total, the nurse shark possesses eight Igω clusters; each cluster contains V, D, and J and six to eight Cω domains (Zhang et al. 2013). Five of the clusters have exon duplications of Cω3 and Cω4 and six clusters contain germline-joined VD. From this study, six secreted isoforms of distinct length and domain composition were identified as cDNAs; four of these forms are shown in Fig. 1. Briefly, IgW-secreted forms can be produced by multiple mRNA splicing patterns, including alternative splicing of Cω exons and the use of secretory exons (sec) or tail pieces (tp, see Fig. 2). Five of the Igω clusters contain two interspersed secretory exons, one located 3′ of Cω2 and one located 3′ of Cω4. All the Cω6 exons end in a secretory tp which resembles the secreted tp of Cμ4 (Zhang et al. 2013). In addition, a V-less IgW form (IgWΔV), with the leader sequence spliced directly to Cω1, has been described in both the spiny dogfish and the nurse shark. In spiny dogfish these V-less transcripts represented 8 % of the IgW transcripts analyzed (Smith et al. 2012), while the V-less transcript levels in nurse shark could only be detected using RT-PCR. In this regard, the functional significance of the different IgW forms can only be determined once IgW-specific reagents, such as mAbs become available.

2.5 IgNAR

The third heavy chain isotype IgNAR (new antigen receptor) was serendipitously isolated from nurse shark serum during the characterization of a panel of anti-IgM mAb (Greenberg et al. 1995). One mAb immunoprecipitated a homodimer of two

covalently attached proteins slightly larger in size than the Igμ chain, but devoid of light chains. Based on peptide sequencing and the use of degenerate primers, the IgNAR cDNA was subsequently identified and found to encode an N-terminal VH domain and five CH domains. The last four CH domains were found to be most closely related to the IgW Cω3–Cω6 domains (Greenberg et al. 1996). IgNAR VH domains are unusual in that they are more similar to TCR and VL domains than to other immunoglobulin VH domains. Plus they are small and contain only two complementarity-determining regions corresponding to CDR1 and CDR3. The CDR2 is replaced by a short hypervariable strand termed HV2 (Stanfield et al. 2004). Early on, EM studies of IgNAR showed that the VH domains are attached to the CH1 domains by flexible hinges and do not form dimers, but instead function as independent, flexible domains that bind antigen in the same manner as camel IgG V_HH (Roux et al. 1998). The electron microscopy also revealed a flexible kink (hinge-like region) within the Fc portion of IgNAR which allows the molecule to bend up to 90° between the CH3 and CH4 domains. These hinge-like regions at C3/C4 and V/C1 were also shown to be susceptible to proteolysis which likely contributes to the observed low IgNAR serum concentrations (0.1–1 mg/ml; Dooley and Flajnik 2005). In the nurse shark, there are four IgNAR loci, three consists of V–D–D–D–J–C and require four gene rearrangements for expression (Greenberg et al. 1995). The fourth locus is mainly expressed in neonatal sharks and has a germline-joined (V–DD–D–J; Diaz et al. 2002). The junctional diversity created by the four rearrangement events results in long and variable CDR3 regions. The IgNAR VH domains have been classified into three "types" based on the presence and location of extra noncanonical cysteine residues, and it is the placement of these extra cysteines that create the different IgNAR conformations. For example, in Type II IgNAR, the CDR3 loop protrudes in a "fingerlike" structure that potentially can bind recessed epitopes or block enzymatic clefts (Stanfield et al. 2007; Streltsov et al. 2005). Because of their structure, IgNAR antigen-binding sites or vNAR, like the camelid V_HH domains, have become antigen-binding scaffold candidates for clinical and biotechnical applications (Kovalenko et al. 2013; reviewed in Criscitiello 2014 and Zielonka et al. 2015).

IgNAR resembles IgW in that the long and short forms are produced by alternative splicing. The long transmembrane and secreted forms contain five CH domains, and the short transmembrane form is produced by alternative splicing of the TM exons to CH3. This shorter form is predicted to form a more stable BCR since it lacks the proteolytically sensitive CH3/CH4 kink (Rumfelt et al. 2004a). More recently IgNAR was sequenced as cDNA in the spiny dogfish and small-spotted catshark, *Scyliorhinus canicula* (Crouch et al. 2013; Smith et al. 2012), and while both species were found to have the long secreted forms (five CH domains) as seen in the nurse shark, there were some differences. First, neither the spiny dogfish nor the spotted catshark expressed Type I NAR VH regions. Second, the spiny dogfish contained 10–20 IgNAR loci as estimated by Southern blot and genomic sequencing indicated that at the most there are two D segments in each clusters and half of the clusters have the VH segment germline-joined to the D1. Also, like the nurse shark, the spiny dogfish expresses both the long and the short transmembrane

forms, while only short TM forms have been found in spotted catshark. Third, as estimated by Southern blot, the spotted catshark has only two or three IgNAR clusters, and sequencing of two clusters from three individuals revealed that each of these clusters contained three D segments. Thus, as in the nurse shark, four gene rearrangements would need to occur to produce a functional VH region.

Interestingly, serum from spiny dogfish and spotted catshark contain a multimeric form of IgNAR as assessed by nonreducing SDS-PAGE combined with western blot and mass spectroscopy, respectively (Crouch et al. 2013; Smith et al. 2012). Thus, IgNAR forms can vary between species.

2.6 Their IgL Chains

Four major types of IgL chains have been classified in the elasmobranchs (reviewed in Criscitiello and Flajnik 2007). The first IgL isotype was identified in the horned shark by Shamblott and Litman in 1989 by genomic sequencing, and IgL chains of this isotype are now designated σ-cart. The σ-cart IgL chains were previously classified as Type I or NS5 IgL chains, and until recently they were thought to be unique to cartilaginous fish since orthologs outside of this class had not been found (Anderson et al. 1995; Criscitiello and Flajnik 2007; Fleurant et al. 2004; Shamblott and Litman 1989b). However, in 2014, using the available coelacanth BAC sequences and transcriptome assemblies, Saha et al. identified orthologs of the σ-cart type in these lobe-finned fish (Saha et al. 2014; discussed in Sect. 3.5). The shark λ IgL chains (Type II/NS3) and κ IgL chains (Type III/NS4) are considered orthologs to Igλ and Igκ chains of the tetrapods (Greenberg et al. 1993; Hohman et al. 1993; Lee et al. 2000a; Rast et al. 1994; Schluter et al. 1989). The fourth elasmobranch IgL isotype was first described in nurse shark, spiny dogfish shark, and in the little skate, *Raja erinacea*, in 2007. These IgL chains were classified as IgLσ since their VLs exhibited 40–56 % amino acid identity with *Xenopus* Igσ and teleost Igσ VL sequences, and their CL were clearly distinct from other elasmobranch CL domains (Criscitiello and Flajnik 2007). Also, phylogenetic analyses of the CL domains of the four isotypes from representative vertebrates demonstrated that elasmobranch Igσ and Igσ-cart CL domains clustered with the Igσ CL domains of *Xenopus* and representative teleost. In this regard, the Igσ and Igσ-cart isotypes could further be distinguished by their CDR1 and CDR2 lengths. The CDR2 regions in both Igσ and Igσ-cart were 3–5 amino acids longer than the CDR2 of the Igλ and Igκ isotypes, while the Igσ CDR1 regions were relatively short. The opposite was observed for Igλ and Igκ, where the CDR1s were longer than the CDR2. In contrast to the elasmobranchs, very few IgL studies have been performed in chimeras. Early on, IgL λ chains were sequenced in the spotted ratfish (Rast et al. 1994), and in 2014 the recent survey of the elephant shark genome identified representatives of three IgL isotypes, Igλ, Igκ, and Igσ-cart; Igσ was not found (Venkatesh et al. 2014).

In all cartilaginous fish examined to date, the IgL chain genes, like the IgH chain genes, are arranged in distinct genomic clusters of $(V–J–C)_n$ with the gene segments

in the same transcriptional orientation. However, the number of clusters for each isotype can vary widely depending upon the species and the extent to which IgL genes are germline-joined also differs among species (Table 2, Fig. 3 and reviewed in Edholm et al. 2011b). All of cartilaginous fish Ig λ clusters examined to date, including those of the elephant shark, are germline-joined. Igκ and Igσ-cart gene clusters are found as either nonrearranged or germline-joined forms. For example, in the nurse shark genome, there are six germline-joined and more than 50 nonrearranged Igκ clusters, and in the little skate, all 89 of the Igσ-cart genes are germline-joined (Anderson et al. 1995; Fleurant et al. 2004; Lee et al. 2000b; Shamblott and Litman 1989b). Currently, elasmobranchs appear to have very few Igσ genes, and they are not germline-joined (Criscitiello and Flajnik 2007).

To date very few studies have addressed how germline IgL (VJ) genes impact the diversity of the primary repertoire. In 1992, Hohman et al. reported the sequencing of four germline-joined Igλ genes in the sandbar shark and showed their CDR3 regions varied in length and encoded for 5–11 amino acids (Hohman et al. 1992). This variability was comparable to that observed for human Igλ CDR3 regions which vary from 2 to 7 amino acids in length (Rock et al. 1994). More recently, and by taking advantage of the low number of Igσ-cart genes in nurse shark, where two IgL genes are unrearranged and one is germline-joined, it was demonstrated that approximately 90 % of the 500 Igσ-cart sequences obtained by RT-PCR or cDNA sequencing contained N-regions of 1–10 bp, with an average length of 4 bp (Fleurant et al. 2004). Moreover, these numbers are higher than what is observed in humans and mice. In humans 58 % of Igκ and 38 % Igλ chains contain from 1 to 4 N-nucleotides. In the mouse 11 % of the Igκ gene rearrangements contain 1 or 2 N-nucleotides, while Igλ rearrangements do not have N-nucleotides. Hence, in the elasmobranchs, TdT is the major contributor to IgL, as well as IgH diversity (see above).

In general, the relative expression of IgL isotypes in the different species correlates with the number of IgL genes present in their genome (Dooley and Flajnik 2006). For example, in the nurse shark, Igκ is the most highly expressed IgL isotype, and it is the isotype that has the largest number of IgL genes (Greenberg et al. 1993; Sledge et al. 1974). In the sandbar shark and horned shark, Igλ and Igσ-cart are the predominantly expressed IgL isotypes, respectively (Hohman et al. 1993, 1995; Rast et al. 1994; Shamblott and Litman 1989a).

2.7 Answers to Important Questions Concerning Germline Joining and Allelic - Isotype Exclusion

When shark IgH chain genes (1987) and IgL genes (1993) were first identified in their germline-joined configuration, many investigators began to question as to when and how this "joining" had occurred and whether perhaps the germline-joined genes were ancestral (?) or did they actually rearrange in germ cells. In 2000, Ellen Hsu and her collaborators found a way to address this issue in the nurse shark. As discussed above, nurse sharks possess both germline-joined and unrearranged Igκ

Table 2 IgL chain isotypes expressed in cartilaginous fish

	IgL σ-cart Type I/NS5	IgL λ Type II/NS3	IgLκ Type III/NS4	IgLσ Type IV	References
Horn shark *Heterodontus francisci*	40 loci[a] Unrearranged 2 genomic clones sequenced	10 loci[a] Germline-joined	Number of clusters unknown 2 cDNA sequenced 1 genomic clone sequenced Unrearranged	3 loci[a]	σ-cart: Rast et al. (1994) λ: Shamblott and Litman (1989b) κ: Rast et al. (1994) σ: Criscitiello and Flajnik (2007)
Nurse shark *Ginglymostoma cirratum*	2 loci[b] unrearranged 1 locus germline-joined 1 pseudogene	6 loci[a] Germline-joined	~50 loci[a] Unrearranged 6 loci Germline-joined	3 loci[a]	σ-cart: Fleurant et al. (2004) λ: Lee et al. (2000b) κ: Greenberg et al. (1993) and Lee et al. (2000b) σ: Criscitiello and Flajnik (2007)
Sandbar shark *Carcharhinus plumbeus*		~200 loci[c] Germline-joined 5 loci sequenced 12 loci mapped			λ: Hohman et al. (1993, 1995)
Little skate *Raja erinacea*	89 loci[a] Germline-joined	15 loci[a] Germline-joined		2 loci[a]	σ-cart: Rast et al. (1994) λ: Anderson et al. (1995) σ: Criscitiello and Flajnik (2007)
Ratfish *Hydrolagus colliei*		30 loci[a] Germline-joined 21 loci mapped 2 loci sequenced 2 VL families			λ: Rast et al. (1994)
Elephant shark *Callorhinchus milii*	1 CL exon identified[b]	20 loci[b] Germline-joined	2 loci[b]	ND	Venkatesh et al. (2014)

ND not documented
[a]Number of clusters is estimated by Southern blot
[b]Genomic sequencing
[c]Calculated based on the number of positive phage clones in a library

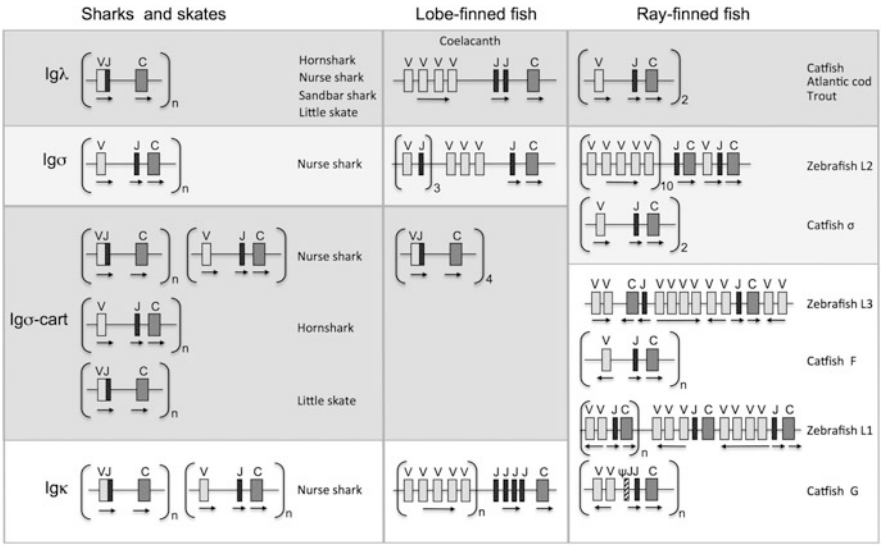


Fig. 3 Schematic comparisons of the different IgL isotype cluster organizations from representative cartilaginous species, teleosts, and the lobe-finned coelacanth. *White boxes* indicate VL genes, *black boxes* indicate JL genes, and *gray boxes* indicate CL genes. The *hatched box* indicates a JL pseudogene. Clusters are denoted by *brackets*, and *arrows* indicate transcriptional orientation. The Igλ and Igσ-cart genes are on *gray backgrounds* for clarity. Drawing is not to scale

genes; six of the Igκ gene clusters are germline-joined and more than 50 are unrearranged (Lee et al. 2000a). Since all the Igκ genes are closely related, phylogenetic analyses could be performed to determine the relationship between the germline-joined and unrearranged genes. In the resulting phylogenetic tree, the germline-joined Igκ genes did not cluster together, but instead clustered among the different unrearranged Igκ genes. This finding suggested that the VJ joining in the germline occurred independently at multiple isolated times. Based on sequence divergence, it was calculated that in nurse sharks, different germline Igκ rearrangements occurred at approximately 49 million years ago (Myr), 11 Myr and 7 Myr. Furthermore, the presence of P nucleotides indicated processing of hairpins associated with RAG-mediated recombination. Because germline-joined genes were apparently derived from the unrearranged genes by a process that may have involved an intermediate characteristic of RAG reactions, it was suggested that recombination events occurred in the germ cells of cartilaginous fish. The functional significance of genes generated this way, however is unclear.

The second question concerns how allelic exclusion and locus exclusion can be maintained in a species with a clustered Ig gene organization. Two studies, again using the nurse shark as a model, have begun to solve this puzzle (Malecek et al. 2008; Zhu et al. 2011). Briefly, by using IgH cluster-specific primers in single-cell genomic PCR, the authors were able to determine the status of each Igμ mini-locus. Altogether 49 surface IgM^+ B cells were examined, and the status

of each Igμ, rearranged or not, was determined. The combined results demonstrated that while one to three rearrangements per B cell could occur, in most cases only one VDJ was in-frame. In two cells where there were two functional rearrangements, whether both could be expressed is unknown; however in mammals IgH allelic inclusion at the DNA level does not always translate into IgH allelic inclusion at the protein level (Brady et al. 2010). Moreover, the shark Igμ genes were activated independently of their linkage or allelic location, and since partially rearranged genes were rarely observed in the B cells, it was concluded that the short distance between the V–D–D–J allowed for efficient recombination within a cluster. Also, because only a limited number of loci were activated in each B cell, it was suggested that both the time during which VDJ rearrangement can occur and the nuclear factors that initiate chromatin activation are limiting. From these studies, a model was proposed where Ig rearrangements at individual loci are attempted sequentially during a limited time window, and a functional VDJ could cause feedback signaling (inhibition) to additional recombination events.

Notably, even though the exact B cell-specific nuclear factors have yet to be identified, clearly the IgH gene organization (a compact ~2 kb germline V–D–D–J separated from the neighboring loci by more than 120 kb) and a short rearrangement window are responsible for IgH chain exclusion in sharks.

2.8 Immune Repertoire Studies

Nurse sharks are ovoviviparous animal, and the embryos develop within an egg case in the uterus where they hatch and live for several weeks before being released (Smith et al. 2015). As with humans and mice, the neonatal or fetal nurse shark antibody repertoire is restricted. However since their primary repertoire is dependent on N-nucleotides contributing to their Ig junctional diversity, neonatal sharks express TdT from birth (humans do not). Despite the TdT expression, analyses of cDNAs from newborn nurse shark epigonal organs (their primary lymphoid tissue) revealed that the VDJ junctions from all three IgH chain isotypes (IgM, IgW, IgNAR) were significantly shorter in newborn as compared to adult sharks (Rumfelt et al. 2004b). Similar studies in the clearnose skate and spiny dogfish, however, did not detect any significant differences in CDR3 lengths between adults and neonates (Miracle et al. 2001; Smith et al. 2012).

Two other studies from the same group further demonstrated that neonatal nurse sharks have different immune responses than adults. During neonatal life, the two serum proteins that are predominantly expressed are IgM1gj and Type III IgNAR. Briefly, newborn nurse shark pups expressed very low amounts of 19S IgM, instead the predominant IgM in their plasma is IgM1gj that possesses a very short CDR3 encoded by its germline-joined VDJ (Rumfelt et al. 2001). However at 1 month, the 19S IgM reaches adult levels, and equal amounts of IgM1gj and 19S IgM are found. IgM1gj preferentially associated with an Igσ-cart IgL (NS5), which is also expressed from a germline-joined cluster. IgM1gj lacks a Cμ2 domain and

resembles mammalian IgG (or IgA?), with both monomeric and dimeric forms in serum. Since no membrane form has been identified and the specificity is "predetermined," it was speculated that IgM1gj may function more as a pattern recognition molecule than a true antibody (Dooley and Flajnik 2006). Interestingly in a parallel study, the Type III IgNAR cluster was shown to contain two D gene segments which can rearrange, even so the CDR3s of Type III IgNAR were very restricted; they exhibited the same length (16 amino acids) and almost identical amino acid compositions. Also as the shark pups aged, both the IgM1gj and Type III NAR expression declined; however, it was still detectable in the epigonal organ in adults (Diaz et al. 2002). Notably, at five months after birth, well-defined white pulp zones, presumably containing T cells and dendritic cells, were observed in the spleen (Rumfelt et al. 2002). Also it was shown at this time that the concentrations of serum 7S IgM and Type I and Type II IgNAR began to increase and reach adult levels at approximately six months of age, and in their review summarizing much of this work, Hsu and Flajnik proposed that these observations mimicked the appearance of specific Ig after immunization (Hsu et al. 2006). We believe these studies as well as the one discussed below additionally tie back with the Clem and Small studies of shark 19S and 7S IgM.

In 2005, Dooley and Flajnik defined shark immune responses by the use of size exclusion chromatography combined with ELISA using a panel of isotype-specific mAbs to IgM and IgNAR. In a three year-long immunization study, six adult sharks were immunized with hen egg lysozyme (HEL), and serum samples were collected at different time points (Dooley and Flajnik 2005).

This study demonstrated that:

1. A small increase in antigen binding by the pentameric IgM fraction could be observed after 3–4 monthly immunizations. As expected, this Ab was cross-reactive with the control antigen.
2. A prominent increase in binding of monomeric IgM occurred at 6 months after the primary immunization. This response was highly specific to HEL with very low cross-reactivity. Also, a significant increase in the titer of antigen-specific IgNAR was also observed.
3. Once the Ab response had reached a plateau, the antigen-specific monomeric IgM and IgNAR titers remained high from 10 months to 2 years dependent on the number of immunizations given.
4. Further immunizations did not increase the response once the plateau was reached.
5. A memory response was demonstrated for both monomeric IgM and IgNAR by the presence of a shorter lag-phase of 4–6 weeks as compared to 4–5 months in the primary response.
6. Affinity maturation was demonstrated with monomeric IgM, but not with pentameric IgM.

Since IgNAR is sensitive to proteolysis and degrades during storage, an increase in IgNAR could not be examined in this study. Instead, the authors used a phage display library based on cDNA obtained from a HEL-immunized shark to identify clonal

lineages and to reconstruct hypermutation pathways from the ancestral to the most mutated family members. The binding affinity of the single-chained antibodies were also analyzed by BIAcore (Dooley and Flajnik 2006), and because the crystal structure of IgNAR in complex with HEL had been determined (Stanfield et al. 2004, 2007; Streltsov et al. 2004), it was possible to map which mutations were important for increasing affinity. These experiments demonstrated that the selected high-affinity (nM) primary repertoire ancestral clones only increased their affinity ~ ten-fold during the immune response. This is in contrast to what is observed in mammalian systems where in the primary repertoire the antibodies are low-affinity (μM) binders and only after somatic hypermutation and clonal selection/expansion in germinal centers is there affinity maturation, i.e., >1000-fold increased binding affinity.

We believe these to be seminal studies that unequivocally prove that immunologic memory and affinity maturation exist in sharks.

As for how somatic hypermutation mechanisms contribute to the antibody repertoire, studies in the nurse shark began with the identification of IgNAR. First, Greenberg et al. 1995 found high number of mutations in IgNAR cDNAs isolated from the spleen, and later Diaz in 1998 asked the question, "Does somatic hypermutation contribute to the primary repertoire (as observed in sheep) or is it mainly responsible for diversifying the secondary repertoire in antigen-driven immune responses?" For this study, cDNA sequences encoding transmembrane ($IgNAR_{Tm}$) and secreted ($IgNAR_{Sec}$) IgNAR were isolated from peripheral blood leukocytes (PBL) from three adult sharks and compared to IgNAR germline sequences with the reasoning that the $IgNAR_{Tm}$ would reflect the primary repertoire, while the $IgNAR_{Sec}$ sequences would come from activated cells. The author's analyses revealed that >90 % of cDNA clones represented Type I and Type II IgNAR, and >40 % of the $IgNAR_{Sec}$ cDNAs contained more than five mutations in their VH regions. While <11 % of $IgNAR_{Tm}$ cDNAs had more than five mutations in their VH regions. Basically, on average $IgNAR_{Sec}$ contained 5.14 mutations per VH, and $IgNAR_{Tm}$ cDNAs contained 1.65 mutations per VH. Furthermore, the mutations were targeted to the HV2 in the Type I $IgNAR_{Sec}$ and to the CDR1 in Type II $IgNAR_{Sec}$ cDNAs with a bias toward replacement substitutions. Also, there was no significant bias toward replacement substitutions in $IgNAR_{Tm}$ cDNAs. In addition, analyses of $IgNAR_{Tm}$ and $IgNAR_{Sec}$ cDNAs from a neonatal nurse shark showed that only a few mutations occurred for both forms. Based on these results, it was concluded that somatic hypermutation does not generate the primary IgNAR repertoire, but only occurs after antigenic stimulation (Diaz et al. 1998). In a follow-up study of IgNAR mutants, Diaz et al. 1999 demonstrated that the mutations in IgNAR occurred mainly in doublets and triplets, and it was proposed that in B cells an error-prone DNA polymerase is responsible for the sequential insertions (Diaz et al. 1999).

Comparative studies also examined somatic hypermutations in IgL λ V regions since these clusters are germline-joined in the nurse shark, and as above, both adult and neonatal sequences were examined (Lee et al. 2002). While Vλ cDNAs were readily isolated from four different IgL λ clusters in the neonatal spleen and epigonal organs, no mutations were observed. However, almost all of the PBL cDNAs from adult sharks could be shown to be derived from a single IgL λ cluster

(a different one for each shark), and it was speculated that this skewing might represent an ongoing immune response. The overall mutation frequency was 1.1 % with 4.45 ± 2.5 substitutions per sequence, albeit a higher frequency of mutations was observed in cDNAs isolated from the spleen (10.1 ± 5.5 per transcript). More than half of the mutations occurred in blocks of 2–4 bp similar to that observed with IgNAR, and the rest were single point mutations. This same type of mutation pattern was also demonstrated for nurse shark IgM VH (Malecek et al. 2005) and later Igσ-cart IgL (Zhu and Hsu 2010). Importantly, by comparing somatic mutations in functional and nonfunctional (unproductive rearrangements) Igσ-cart IgL cDNA as well as in IgM VH, transcripts, it was determined that there is selection for replacement mutations in the CDRs, however the selection is not as strong as compared to that observed in humans. This second study mapped the mutations and determined that they were not limited to the VDJ but extended ~1.5 kb into the J–C intron. Also from the analyses of nurse shark RAG rearrangements in Zhu et al. (2011), it could be concluded that somatic hypermutations are limited to rearranged VDJ regardless of functionality since there was no evidence of mutations in germline VH regions (Zhu et al. 2011).

An unexpected finding during cDNA analyses of VDJ regions known to be linked to a certain Igμ cluster was that the same VDJ regions were also associated with another Cμ region. Upon closer examination, it was determined that this occurred through class switch recombination and that switching was increased following immunization and occurred together with somatic hypermutation. This class switch was observed in 25 out of 108 Cμ (23 %) cDNAs (Zhu et al. 2012); it occurred in the J–C intron and was not dependent on a "switch region." Importantly this class switch is reciprocal, nondirectional, and independent of gene location. This switching was observed to occur between IgCμ genes and IgCω genes, i.e., VDJ switching from one Cμ to another Cμ and VDJ switching from one Cω to another Cω. Also rearranged VDJ linked to Cω could switch to Cμ and vice versa (Zhang et al. 2013; Zhu et al. 2012). Thus, from these studies in her laboratory, Ellen Hsu suggested that in sharks, somatic mutation and class switch recombination are contiguous and dependent since both occur in proximity to the VH promoter and are due to the repair that occurs following AID-mediated lesions.

3 The Bony Fish (Osteichthyes)

3.1 Why We Study Them and Overview of What We Know

Osteichthyes are fish that have a bone skeleton, and they form the largest class of vertebrates with more than 31,000 diverse species occupying both marine and freshwater habitats. This superclass is divided into the ray-finned fishes (Actinopterygii) and lobe-finned fishes (Sarcopterygii). The ray-finned fish are further divided into the Chondrostei (sturgeons, Acipenseridae, and paddlefish,

Polyodontidae), Holostei (gar, Lepisosteidae, and bowfin, *Amia calva*), and the Teleostei that includes the majority of living fishes. The lobe-finned fish (*Sarcopterygii*) diverged from the ray-finned fishes approximately 400 million years ago and are comprised of the coelacanths, *Latimeria*, and the lungfishes, *Dipnoi*. This section should be considered an overview of what is known about the different immunoglobulins and their expression in teleosts, with an emphasis on the varied IgH chain and IgL gene organizations from representative species. We have also included schematics of the coelacanth Ig gene organizations for comparisons (Fig. 4).

Three IgH isotypes have been identified in teleosts: IgM, IgD, and IgT, and in teleosts, tetrameric IgM (μ_2L_2) is the most prevalent form found in serum, although IgM monomers were isolated in giant grouper, *Epinephelus itajara*, and sheepshead, *Archosargus probatocephalus* (Clem 1971; Lobb and Clem 1981a, 1981b). Overall IgM is the most well-characterized Ig isotype in bony fish (with studies beginning over 50 years ago), and it is accepted to be a structural and functional homolog of mammalian IgM (reviewed in Bengtén et al. 2000; Fillatreau et al. 2013; Flajnik and Du Pasquier 2013). Notably in all teleost species of bony

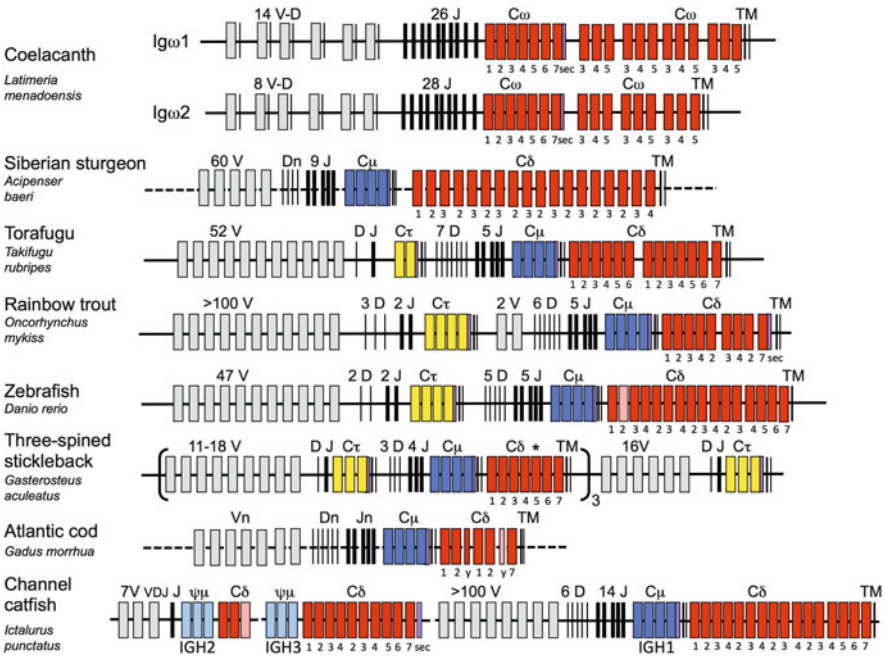


Fig. 4 Schematic comparison of the IgH genomic organization in the lobe-finned coelacanth, sturgeon, and representative teleosts. The VH, D, JH and Cμ, Cδ, and Cτ gene segments are *color shaded* and *labeled*. The Cω and Cδ exons are numbered; *ψ* indicates a pseudogene and *y* indicates a mini-exon. The duplicated IgH chain loci of the three-spined stickleback are indicated by *brackets*, and *asterisk* indicates that in one of the duplications this exon is a pseudogene. *Dashed lines* in the sturgeon and *Atlantic cod* indicate the gene organizations were inferred from sequencing and Southern blots. Drawing is not to scale

fish sequenced to date the Igμ gene always consists of four Cμ exons and two transmembrane (TM) exons. However, unlike in other vertebrates, teleost Igμ transmembrane forms are produced by alternative splicing of the TM exons directly into the donor splice site of the Cμ3 exon since the cryptic splice site within the Cμ4 exon is missing. This results in shorter IgM receptors on the B cell surface (consisting of VH–Cμ1–Cμ2–Cμ3–TM), and this lack of a Cμ4 does not affect their ability to associate with CD79 signaling molecules (Sahoo et al. 2008). In addition, this unusual pathway of Igμ RNA processing is evolutionarily ancient, as it also occurs in the holosteans, and modifications of this splicing also occur in some Antarctic fish (Quiniou et al. 2011; Wilson et al. 1995).

In contrast to IgM, IgD, and IgT, functional studies have only been performed in a few of the model teleost species, namely, channel catfish, *Ictalurus punctatus*, and rainbow trout, *Oncorhynchus mykiss*, respectively. Even so, the data from Oriel Sunyer and his collaborators demonstrating that IgT in rainbow trout functions analogously to mammalian IgA by protecting mucosal surfaces is impressive (Gomez et al. 2013; Sunyer 2013; Xu et al. 2013; Zhang et al. 2010), and it was determined that the levels of IgT in gut mucus is almost twice that of what is observed in serum (7.1 μg/ml vs. 3.7 μg/ml). Also, the number of papers in the "IgT" field is rapidly increasing! Since most of these studies can be found summarized in this edition, please see Magadan et al. (2015) in this edition for an overview.

Catfish IgD was originally identified as a chimeric cDNA consisting of a rearranged VDJ spliced to the Cμ1 domain, followed by seven Cδ domains, a TM, and a positively charged short cytoplasmic tail (CYT). This finding demonstrated that catfish IgD would be produced by alternative processing of a primary RNA transcript (Wilson et al. 1997), and this IgD processing pathway has been found in all teleosts studied to date. Interestingly, while only V-less Igδ cDNA transcripts could be found by cDNA screening and 5′-RACE, a secreted IgD form was readily detected in catfish serum using an anti-IgD mAb specific for the secreted tail. Consequently, it was proven that membrane Igδ and secreted Igδ forms were encoded by different genes located ~700 kb apart, and the gene encoding secreted Igδ was linked to a pseudoIgμ gene (Fig. 4 and Bengten et al. 2002, 2006b). Notably, in 2009, three different IgD^+ leukocyte populations were identified in catfish PBL: small and agranular double-positive (IgM^+/IgD^+) B cells; larger single-positive IgM^-/IgD^+ B cells that resembled plasmablasts; and granular cells that were armed with exogenous IgD via a putative IgD-binding receptor. Furthermore, it was inferred that this mechanism of IgD being bound to a granulocyte by an IgD-binding receptor is evolutionary conserved since secreted IgD is found on human basophils. Briefly, Chen et al. demonstrated that in humans, IgM^-/IgD^+ plasmablasts associated with the upper respiratory tract secrete IgD, which in turn is bound to basophils through a calcium-mobilizing receptor. Moreover, cross-linking of the IgD bound to the basophils resulted in the release of antimicrobial, opsonizing, pro-inflammatory, and B cell-stimulating factors, including cathelicidin, pentraxin-3, interleukin 1 (IL-1), and B cell-activating factor of TNF family (BAFF; Chen et al. 2009).

In a parallel study published in 2010 (Edholm et al. 2010), through the use of cell sorting, flow cytometry, and RT-PCR, it was demonstrated that catfish IgM^-/IgD^+ B cells preferentially utilized IgL σ chains, and the IgM^-/IgD^+ B cell populations varied between individual catfish. For example, the IgM^-/IgD^+ B cells could represent as much as 72 % of the PBL in some individuals. Also and again, all of the secreted IgD transcripts that were sequenced from these cells were V-less and contained a leader sequence spliced to the Cδ1 domain, and it seems likely that these transcripts encode for the secreted form of IgD previously found in catfish serum. Importantly, the Igδ leader sequence was (1) mapped to a region 5′ of the secreted Igδ gene and (2) shown to mediate IgD protein secretion when transfected into catfish B cells. More recently using anti-rainbow trout Cδ2 mAbs, Ramirez-Gomez et al. identified populations of small IgM^+/IgD^{dim} B cells and IgM^+/IgD^{bright} blasting cells in rainbow trout (Ramirez-Gomez et al. 2012). Importantly, these mAbs were also used to identify two heavily glycosylated secreted forms of IgD (associated with IgL chains) in rainbow trout serum. The IgD forms when analyzed under nonreducing conditions in SDS-PAGE were ~380 and 240 kDa in size, and under reducing conditions (after deglycosylations), the two IgD heavy chain proteins were estimated to be 95 kDa and 70 kDa in size, respectively. In this same study, IgD serum concentrations were shown to vary between 2 and 80 μg/ml depending upon the individual fish, and ELISA spot assays revealed that IgD secreting cells were present in the anterior and posterior kidneys, spleen, and gills. The ratio of IgM to IgD was five times higher in the kidney and spleen but was close to one in the gills. In this context, it is important to understand that the authors proved that the two IgD secreted forms are produced by alternative splicing by using RT-PCR coupled with genomic sequencing. Their results showed that exon Cδ7 contains a tail piece of 33 nucleotides that is eliminated when the TM exons are spliced into a cryptic splice site within the Cδ7 exon to produce the membrane form. Transcripts encoding the 95 kDa form consist of VDJ–Cμ1–Cδ1–Cδ2–Cδ3–Cδ4–Cδ2–Cδ7 s, while the one encoding the 70 kDa protein consists of VDJ–Cμ1–Cδ1–Cδ2–Cδ7 s. Also, V-less forms were not found in rainbow trout.

Notably, while IgM^-/IgD^+ B cells were not detected in rainbow trout PBL, a population of IgM^-/IgD^+ B cells was detected in the gills of rainbow trout by Castro et al. in 2014. These IgM^-/IgD^+ B cells made up ~9 % of the gill leukocyte population and expressed message for membrane IgD, secreted IgD, and the chemokine receptor CCR7. In fact, 75 % of the cells expressing CCR7 were IgM^-/IgD^+ B cells. In mammals, CCR7 is the receptor for the CCL19 and CCL21 chemokines that control trafficking of naïve T cells, B cells, and mature dendritic cells to secondary lymphoid tissues (Castro et al. 2014). Thus, while progress has been made in characterizing teleost IgD at both the DNA and protein levels in a few teleosts, we have yet to tie these bits and pieces together and more work needs to be done.

3.2 Their IgH Chains

Teleost IgH chain genes are organized in a translocon configuration similar to those of mammals and other vertebrates (reviewed in Fillatreau et al. 2013; Hikima et al. 2011). However, depending on the species, differences from the typical organization of multiple VH, D, and J gene segments followed by genes encoding the different IgH isotypes do occur. As in mammals where species unique IgH variations are well documented, the different teleost IgH loci have undergone multiple rounds of duplication and deletion. Briefly, the bony fish IgH genes are highly diverse both in terms of the number of duplications and the types of duplications they have acquired. There are duplications of individual VH, D, or J segments and tandem duplications of exons encoding CH domains such as the (δ2–δ3–δ4) duplications found in catfish, zebrafish, *Danio rerio*, and Atlantic salmon, *Salmo salar* (Danilova et al. 2005; Yasuike et al. 2010). Also tandem duplications of entire IgH loci can occur, as seen in the stickleback, *Gasterosteus aculeatus*, and medaka, *Oryzias latipes* (Gambon-Deza et al. 2010; Magadan-Mompo et al. 2011). In the coelacanth IgH locus, there are tandem duplications of VH–D units upstream of multiple duplicated JH (see below). Thus, as expected, the IgH gene duplications in bony fish greatly contribute to diversity since duplicated genes can drift and perhaps acquire new functions. Figure 4 shows the different IgH gene organization from representative bony fish and highlights some of the various types of gene duplications that have been found, for example, the numbers of Igτ and Igδ exons, the lack of Igτ in catfish, and the existence of only Igω (the Igδ ortholog) in the coelacanth. In rainbow trout and zebrafish, the Igτ gene is located 5′ of the D and J gene segments associated with the Igμ and Igδ genes, and in these species, these Igτ genes have their own set of 5′ D and J gene segments. In both species, the Igτ genes consist of four Cτ exons and two τTM exons, and of course, this gene encodes both the secreted and transmembrane forms by differential splicing of primary RNA. While the torafugu, *Takifugu rubripes*, Igτ locus is organized the same, the Igτ gene contains only two Cτ exons (Fu et al. 2015; Savan et al. 2005). In contrast, in the three-spined stickleback, *Gasterosteus aculeatus* (a model species for evolutionary studies), there are three repeated units of (V_n–D–J–Cτ–D_3–J_4–Cμ–Cδ) followed by an additional repeat of (V_6–D–J–Cτ). Such an organization results in four copies of Igτ, each having three Cτ exons, and here it should be noted that the stickleback IgH chain locus (IGH) is only ~175–200 kb in length (Bao et al. 2010; Gambon-Deza et al. 2010). Interestingly, Igτ genes have not been found in catfish and medaka, and in these species, it may have been lost when the IgH loci were duplicated or perhaps it was never there to begin with (Bengten et al. 2006b; Magadan-Mompo et al. 2011)!

As in all tetrapods and teleosts, the genes encoding IgD are directly 3′ of the Igμ gene. For example, the catfish IGH locus contains multiple Ig gene duplications and transpositions and spans ~1 Mbp. It contains three Igδ genes (IGHD1, IGHD2, and IGHD3), and each is individually linked to either a functional Igμ gene or an Igμ pseudogene. The Igδ that encodes the IgD transmembrane form is directly 3′ of the functional Igμ gene. The other Igμ genes found in the IGH2 and IGH3 loci are

pseudogenes, but the Igδ genes linked to them are intact. Furthermore, the catfish IgD secreted form is encoded by the Igδ linked to IGH3. In comparison, the zebrafish IGH locus is 175 kb in length, and it contains only single copies of Igμ, Igδ, and Igτ (Danilova et al. 2005; Zimmerman et al. 2011a). That teleost Igδ cDNAs can vary greatly is obvious from the sequencing of Igδ cDNAs from multiple species. In catfish, the Igδ transmembrane form consists of VDJ–Cμ1–Cδ1–Cδ2–Cδ3–Cδ4–Cδ5–Cδ6–Cδ7–TM, and in rainbow trout, the Igδ transmembrane form consists VDJ–Cμ1–Cδ1–Cδ2–Cδ3–Cδ4–Cδ2–Cδ7–TM1–TM2 (Hansen et al. 2005). Comparatively, in Atlantic cod, the Igδ transmembrane form consists of VDJ–Cμ1–Cδ1–Cδ2–δy–Cδ1–Cδ2–Cδ7–TM1–TM2 (Stenvik and Jorgensen 2000), and in the sturgeon *Acipenser baeri*, the most prevalent Igδ membrane transcripts consist of VDJ–Cμ1–Cδ1–Cδ2–Cδ3–Cδ4–TM. However, multiple Igδ splice variants containing duplication of (Cδ2–Cδ3) also occur in sturgeons, and the longest transcripts contain eight such duplications (Zhu et al. 2014).

To date, comprehensive germline assemblies for zebrafish, stickleback, Atlantic salmon, and torafugu are available. Table 3 summarizes the current number of VH families and the VH, D, and JH gene numbers in eight representative teleost species compared to those in the coelacanth and human. In the zebrafish there are 47 VH gene segments; 39 of these are functional, and these are divided into 13 VH families based on members sharing at least 70 % nucleotide identity (Danilova et al. 2005). In comparison, based on VH family members sharing >80 % nucleotide identity, the stickleback possesses 49 VH gene segments divided into four VH families (Bao et al. 2010), and the torafugu have 34 functional VH gene segments that form 3 families (Fu et al. 2015). In the catfish there are at least 14 VH families of substantial size, and this is based on the current available sequencing data since the sequencing and annotation of the catfish genome are still ongoing (Bengten et al. 2006a; reviewed in Bengten et al. 2006b). As for the D and J gene segments associated with Igτ, the numbers are low, ranging from one to three in all the species that have been examined. In contrast and as represented in Table 3, the numbers of D and J gene segments associated with Igμ are larger, e.g., three to nine Dμ genes and four to 12 Jμ genes.

3.3 *The Coelacanth IgH Chain Loci and Lungfish IgH Chains*

The coelacanth IgH locus exhibits a very unusual organization, and in both the *Latimeria* species, there are two very similar IgW loci (IgW1 and IgW2) that are on different scaffolds in the genome assembly. The IgW1 locus contains 14 closely linked VH and D gene segments and 26 diverse JH gene segments upstream of the exons that encode the Igω C region. The first seven exons are unique and encode the secreted IgW. These exons are followed by four tandem duplications of (Cω3–Cω4–Cω5) and two exons encoding the TM (Saha et al. 2014). Currently, since

Table 3 The IgH chain germline repertoires in ray-finned and lobe-finned fish

	Zebrafish	Catfish[a]	Rainbow trout	Atlantic salmon[b]	Torafugu	Tuna	Stickleback	*Latimeria menadoensis*		Human	
VH families	13[c] 1[d]	14	13	18	3 2[d]	4	4	VH families	9	VH families	7
Sharing >80 % nucleotide identity											
VH segments	39 8[d]	28 27[d]	>100	300	34 16[d]	ND	49	VHω1 VHω2	10 + 4[d] 7 + 1[d]	VH segments	65
Dτ	2	ND	3	2_A 2_B	1	1	1×4[e]	Dω1	13		
Jτ	2	ND	2	2_A 2_B	1	1	1×4[e]	Jω1	26		
Dμ	5	6	6	9_A 6_B	7	ND	3×3[e]	Dω2	6	D	27
Jμ	5	12	5	5_A 5_B	5	10	4×3[e]	Jω2	28	J	6

[a]The numbers are based on sequencing 25 % of the catfish IgH locus

[b]The Atlantic salmon contain two independent IgH loci termed IGH_A and IGH_B, these are located on separate chromosomes; both loci contain several Igτ genes or pseudogenes and their number are not included here

[c]Defined by the authors as sharing >70 % nucleotide identity

[d]Pseudogenes may be functional in some individuals

[e]The Igτ locus is duplicated four times and the Igμ locus three times

References: Zebrafish: (Danilova et al. 2005), Catfish: (Bengten et al. 2006b), reviewed in (Bengten et al. 2006a) Rainbow trout: (Brown et al. 2006; Hansen et al. 2005; Roman et al. 1996), Salmon: (Yasuike et al. 2010), Torafugu: (Fu et al. 2015) Tuna: (Mashoof et al. 2014), Stickleback: (Bao et al. 2010; Gambon-Deza et al. 2010) *Latimeria*: (Saha et al. 2014) Human: reviewed in (Saada et al. 2007)

RNA from coelacanth lymphoid tissues has not been available, the contribution of junctional diversity to the repertoire cannot be examined.

The IgH chains have been identified in three species of lungfish, *Protopterus aethiopicus*, *P. annectens*, and *P. dolloi*. In sharp contrast to the coelacanth, the lungfish species express multiple diverse IgM- and IgW-like isotypes, and the number of IgM and IgW genes differs among the various species (Ota et al. 2003; Zhang et al. 2014). The diploid *P. annectens* has three IgM isotypes, while the tetraploid *P. dolloi* has two. Similarly, IgW has undergone several duplications: all three species have a long form of IgW, consisting of seven Cω domains and one short form consisting of two Cω domains as in the sharks and skates. In addition there are three unique isotypes termed IgN that are related to the lungfish IgW, but they are clearly distinct since these sequences share ~30 % amino acid identity. *P. annectens* expresses all three IgN isotypes, while *P. dolloi* expresses only one IgN. Also, a fourth isotype termed IgQ was found in *P. annectens*. Phylogenetic analyses using the Bayesian method showed that lungfish IgW and IgN sequences branch with the coelacanth IgW and tetrapod IgD. Interestingly, it appears that IgQ is more related to sturgeon and teleost IgD and the IgW of sharks, than it is to tetrapod IgD. However, due to the large genome size of the lungfish, it is impossible to sort out the IgH chain gene organization. Whether this suggests that several IgD-like genes were present in the common ancestor to the coelacanth is certainly debatable, and it seems ironic that the coelacanth has survived on this enigmatic isotype. As a historic note, in the 1990s, studies using transgenic (knockout) mice that were germline deficient for either IgD or IgM demonstrated that IgM or IgD, each by itself, was capable of functioning as a BCR and inducing normal B cell immune responses (Lutz et al. 1998; Nitschke et al. 1993; Roes and Rajewsky 1993).

3.4 Their IgL Chains

To date, four IgL isotypes, orthologous to κ, σ, λ, and σ-2, have been identified in bony fish, and these are the designations used in this overview. The σ-2 isotype (formerly known as σ-cart) was recently identified in the coelacanth, and the name change was suggested by the authors (Saha et al. 2014). Early on, as with the elasmobranchs, the relationship between teleost IgL chains and mammalian Igκ and Igλ chains was unclear. Often teleost IgL sequences were named to emphasize that they were neither Igκ nor Igλ, and in some species, the IgL chains were named in the order they were identified. To make matters worse, the relation between the different isotypes within a species was also blurred, most likely due to the fact that very few IgL chains had been sequenced in teleost fish or in other ectothermic vertebrates (reviewed in Criscitiello and Flajnik 2007; Edholm et al. 2011b). Reasons for the Ig κ, σ, and λ designations can be summarized in brief. First, while the teleost IgL L1 and L3 chains clearly represented distinct isotypes, i.e., their C regions share ~30 % amino acid identity with each other, phylogenetic analyses of V as well as C domains clearly demonstrated that both were orthologs of IgL κ

chains. Second, the rainbow trout L2 σ chains were shown to be orthologous to the σ isotype found in *Xenopus* (Partula et al. 1996). Third, the λ classification of the new IgL chain sequences in the catfish and Atlantic cod in 2009 was supported not only by phylogenetic analyses, but also by their longer CDR1 regions and the presence of conserved RSS with the same orientations as in the λ genes in chicken and mammals. Basically, the Vλ RSS consisted of a heptamer and nonamer separated by a 23-bp spacer, and the Jλ RSS consisted of a heptamer and nonamer separated by a 12-bp spacer (Edholm et al. 2009).

As in cartilaginous fish, teleost IgL genes are arranged in (V–J–C) genomic clusters; however, there are distinct differences. In teleost the IgL clusters are more closely linked and have a more varied organization, and the Igκ V genes are in opposite transcription orientation with respect to their J and C genes. This reverse orientation implies that teleost Igκ genes rearrange via inversion rather than deletion and was first suggested by Ghaffari and Lobb in 1997 and later proven by Zimmerman in 2008 (Ghaffari and Lobb 1997; Zimmerman et al. 2008). Also, the numbers of clusters vary in the different species and between the different IgL isotypes (see Fig. 3 and Table 4). The IgL gene organizations have been determined either directly by genomic sequencing and/or inferred from Southern blot analyses in several different teleost species (reviewed in Edholm et al. 2011b). For the purpose of this review, we have included data from only a few representative species, including the catfish (Edholm et al. 2009; Ghaffari and Lobb 1993, 1997; Jones et al. 2004), zebrafish (Zimmerman et al. 2008), rainbow trout, Atlantic cod (Daggfeldt et al. 1993; Partula et al. 1996; Timmusk et al. 2000; Edholm et al. 2009), stickleback (Bao et al. 2010), and for comparison the coelacanth (Saha et al. 2014).

Most teleost IgL κ gene clusters, as represented by the catfish, contain one or two V genes, a single J gene, and a single C gene (Fig. 3). From Southern blots, catfish

Table 4 IgL isotype clusters in teleosts

	Channel catfish	Zebrafish				Rainbow trout	Three-spined stickleback		
Igλ									
V	2					2[a]			
J–C	2	ND				2[a]	ND		
Igκ (L1)		*Chr:*	*3*	*19*	*24*		*Chr:*	*10*	*15*
V	30[a]		5	8	19	40[b]		43	6
J–C	15[a]		2	5	5	10[a]		23	6
Igκ (L3)		*Chr: 25*							
V	50[a]	10				40[b]			
J–C	50[a]	2				15[a]			
Igσ		*Chr: 3*					*Chr: 11*		
V	2	11				8–11[c]	24		
J–C	2	2				3[a]	8		

[a]Number of clusters is estimated by Southern blot
[b]Igκ L1 and L3 share VL families and may be linked on the same chromosome
[c]Several Igσ V families are present

are estimated to have ~65 Igκ genes (50 Igκ F and 15 Igκ G), and sequencing has demonstrated that the Igκ G clusters contain a pseudo J gene (Jones et al. 2004). In contrast, genomic sequencing and Southern blots show that there are only two Igσ and Igλ genes in the catfish. Also, to date the catfish, Atlantic cod, and rainbow trout are the only teleost where Igλ genes have been identified (Edholm et al. 2009). Interestingly, in the zebrafish the Igκ gene clusters are longer due to multiple VL gene duplications (Zimmerman et al. 2011b). All together there are 12 Igκ L1 genes scattered on three different chromosomes, and the different Igκ clusters can contain from one to four V. The zebrafish IgLκ L3 genes are organized slightly differently in two clusters of (V_4–J–C–V_2) and (V_2–J–C–V_2), i.e., there are 10 V genes. In contrast, the zebrafish Igσ locus contains 10 V followed by two tandem copies J–C with a V in between. Whether this is a translocon organization can be debated. Except for in the stickleback, all Igσ and Igλ V are all in the same transcriptional orientation as their corresponding J and C genes. For a comparison, from preliminary studies in the Siberian sturgeon and sterlet using Southern blots, it appears that both species possess at least 20 Igκ V genes (Lundqvist et al. 1996).

3.5 The Coelacanth IgL Chain Loci

Finally, the recent analyses of the coelacanth genome identified all of the four IgL isotypes previously found in ectotherms, Igκ, Igλ, Igσ, and Igσ-cart. The presence of Igσ-cart gene clusters in the coelacanth genome was unexpected and again (!) prompted a change in the IgL chain nomenclature. The Igσ-cart genes, now termed Igσ-2, resemble the cartilaginous fish IgL genes; a total of four VJ germline-joined Igσ-2 gene clusters were found (Fig. 4). In contrast, the Igλ, Igκ, and Igσ loci have translocon-type organizations. The Igλ locus contained four V genes, two J genes, and a single C gene, and the Igκ locus contained four V genes, four J genes, and a single C on one scaffold. The IgLσ locus contained three tandem duplications of V–J followed by three V and single J and C, and these were on a single scaffold as well. All of the coelacanth genes were in the same transcriptional orientation. Even so, due to the number of assembly gaps and ambiguous regions, the sequencing of the IgL chain genes is still not complete. Interestingly (and fun), in phylogenetic analyses the coelacanth Vκ and Vλ genes and their teleost counterparts branched together in a neighboring joining tree, while the Vσ genes branched with both the *Xenopus* Vσ and teleost Vσ genes (Saha et al. 2014). As expected and per definition, the Vσ-2 clustered with the cartilaginous Vσ-cart.

3.6 Diversification of the Primary Repertoire

Post-rearrangement diversification for Ig genes include: (1) receptor editing, (2) receptor revision, and (3) somatic hypermutation, and evidence for two have been found in certain species of teleosts, e.g., catfish and zebrafish:

1. **Receptor editing** is a process through which an unwanted gene rearrangement can be eliminated by a secondary gene rearrangement. This occurs through RAG-mediated recombination. Basically, an upstream RSS of the IgH chain can partner with a cryptic RSS located near the end of the FR3 of the VDJ rearrangement. This second rearrangement results in a new VDJ with a longer CDR3. The 3′ cryptic RSSs are known to be conserved in jawed vertebrates including teleosts (Sun et al. 2012). However to what extent this type of editing occurs has not been examined in the teleosts.

   Here it should also be noted that the teleost IgL genomic organization with closely linked clusters of (V–J–C) offers several recombination possibilities for receptor editing and locus silencing (Hsu and Criscitiello 2006; Zimmerman et al. 2011b).
2. **Receptor revision** occurs in activated B cells that have undergone clonal expansion and somatic mutation. The revised VDJ sequences are hybrids that consist of two separate germline VH genes joined at a homologous sequence. Revisions can be attributed to either AID-mediated gene conversions or RAG-mediated recombinations using cryptic RSS pairs embedded in both VH genes (Nemazee and Weigert 2000; Wilson et al. 2000). In 2009, Lange et al. provided strong evidence for receptor revision of catfish IgH chains (Lange et al. 2009). Briefly, >300 cDNA clones were obtained by parallel screening of seven different tissue libraries obtained from one individual catfish using VH family specific probes. At least eight clonal sets or ancestral lineages could be identified. Some clonal sets were represented in PBL, anterior kidney, spleen, and intestine, while others were only found in the intestine. In total 13 hybrid VH clones were classified, six sequences were hybrids of different VH segments from the same family, and seven of the hybrids had a donor sequence of a different VH family. Also, the frequency of hybrids within a clonal set varied. For example, one large clonal set was represented by 60 sequences and did not contain any VH hybrids, while another set of 10 sequences represented in four tissues contained four different hybrids and were the result of multiple sequential VH replacement events. Importantly, all the hybrid clones could be re-isolated from the same cDNA libraries, and these were also verified by RT-PCR from total RNA obtained from the corresponding original tissues used to make the libraries. That the authors could find the same VH hybrid in the cDNA libraries from different tissues gives credence to their study. Notably, since heptamers of cryptic RSS could be identified in the predicted VH donor and recipient sequences, the authors suggested that the receptor revision in catfish is RAG mediated and occurred via hybrid joint formation.
3. **Somatic hypermutation** has been studied in catfish IgH chains and zebrafish IgL chains (Marianes and Zimmerman 2011; Yang et al. 2006). These studies demonstrated that the mutations preferentially target G and C nucleotides within WRCH/DGYW nucleotide motifs, which are the common targets for AID in mice and humans. At this time and based on these few studies, it seems as the mutation/repair mechanisms following cytidine deamination may differ between teleosts and sharks since no block substitutions could be identified (Barreto and

Magor 2011). However, as in sharks, the somatic mutations in teleost are not accompanied by a substantial increase in antibody affinity. Thus, the primary role of somatic mutation in ectothermic vertebrates may be to diversify and broaden the antibody repertoire, perhaps allowing their immune systems to recognize the microbial antigenic variations that can occur during infection.

3.7 *Summary Points of What We Know About Primary Repertoire Development and Evolution of Ig Isotypes*

Table 5 provides a comparison of the known mechanisms that generate the antibody repertoires associated with the different isotypes found in teleosts, cartilaginous fish, and the coelacanths. Both teleosts and sharks have three IgH chain isotypes, presumably each of them has different effector functions. IgM is the primordial immunoglobulin present in all species, with the exception of the coelacanth. In all fish species examined, it is the most prevalent immunoglobulin in serum, and with exception of the zebrafish, it is the first Ig isotype to be expressed during development. Functionally, fish IgM has been shown to mediate opsonization, antibody-dependent cell-mediated cytotoxicity (ADCC), and complement activation (Haynes et al. 1988; McKinney and Flajnik 1997; Smith 1998; Lobb and Hayman 1989; Shen et al. 2003; Boshra et al. 2006). As for Igω/Igδ, it seems likely that this isotype first originated in cartilaginous fish, perhaps from a fusion of duplicated Igμ loci.

Table 5 Mechanisms contributing to the antibody repertoire in fish

	Teleosts	Coelacanths	Elasmobranchs
Isotypes	IgM, IgD (IgT)[a]	IgW	IgM, IgW, IgNAR
Germline diversity	+++	+++	+
Combinatorial diversity	Yes	Yes	No
Junctional diversity	+	Assumed	+++
Pairing of heavy and light chains	+++	++ IgW: four IgL types	IgM19S: + IgM 7S: ++ IgW: + IgNAR: −
Somatic hypermutation	IgM, IgL ++	Unknown	IgNAR +++ IgM, IgW, IgL ++
Affinity maturation	5×	Unknown	10×
B cell memory[b]	Yes, in some species	Unknown	Yes
Class switch	Unknown	Unknown	Yes
Neonatal immunity		Unknown	IgM19S short CDR3 IgM1gj Type III IgNAR
Mucosal immunity	IgT/IgD	Unknown	Possibly IgW

[a]Present in the majority of teleosts
[b]Defined by a shorter lag time in the secondary response

Also, it is clear that Igω/Igδ has remained as an evolutionary labile Ig gene in vertebrates since exon duplications seem to be its hallmark. For example, in nurse sharks the Cδ3 and Cδ4 exons are duplicated; in the coelacanths, block duplications of exons (Cδ3–Cδ4–Cδ5) are found between the Cδ7, which also encodes the secreted tail piece, and the TM exons, and in the Siberian sturgeon, there are eight repeats of (Cδ2–Cδ3). In teleosts different types of duplications have been observed (see above); the most common is (Cδ2–Cδ3–Cδ4). In most species, these duplicated exons are differently expressed as the result of alternative splicing. Interestingly and as an aside, IgD in *Xenopus*, leopard gecko, and platypus is also long, consisting of VDJ and 8, 11, and 10 Igδ domains, respectively. In all of these species, the IgD lacks a hinge region, and it is proposed that the length of the molecule may allow for some flexibility (Edholm et al. 2011a).

A second, albeit related, feature of Igω/Igδ is the expression of long and short forms of Igω in sharks, rays, and lungfish (see Fig. 1) and the shorter and longer serum IgD forms observed in rainbow trout and catfish. It is tempting but maybe premature to speculate that the different IgW/IgD forms have differed effector functions. Perhaps the shorter Igω forms, like the duck IgYΔFc, function as neutralizing antibodies without activating complement (Rumfelt et al. 2004a; Magor 2011). The presence of V-less IgD forms in catfish serum, as well as V-less Igω cDNAs in spiny dogfish and nurse shark, is intriguing and suggests that perhaps this isotype has coevolved with certain respective pathogens to bind conserved structures via its Fc (reviewed in Edholm et al. 2011a). If so, IgW/IgD would function similarly to the IgD found on plasmablasts present in the human upper respiratory mucosa where it specifically binds *Moraxella catarrhalis* and *Haemophilus influenzae* via its Cδ1 domain (Chen et al. 2009). Thus, even though we know much about Igω/Igδ gene structure in many cold-blood vertebrates and mammals, its function still remains elusive, and as suggested by Hsu and Flajnik, the IgD locus is the one that evolution "plays with" (Hsu et al. 2006). In elasmobranchs, this playing with the Igω locus has clearly resulted in the unique IgNAR isotype, which lacks IgL chains. It is predicted that the IgNAR resulted from a translocation of a NAR-TCR V region to an Igω cluster (Criscitiello et al. 2006), and as discussed above, IgNAR immune responses can be highly specific even in the absence of germinal centers. This is not only due to the high levels of somatic hypermutation that occur in the CDRs, but also to those in the framework regions, HV2 and HV4. These mutations, specifically those of HV2, stabilize the IgNAR binding sites by increasing their number of antigen contact residues, and antigen binding occurs sideways in relation to the antibody framework (Kovalenko et al. 2013).

IgT is an isotype that is unique to teleosts, and studies show that it has evolved to a dedicated mucosal antibody (reviewed in Sunyer 2013; and see below). However, the organization of the teleost IgH locus implies that the IgT repertoire is limited since there are only a few Dτ and Jτ segments (Table 3). For example, in zebrafish there are 156 possible recombinations for Igτ (39 VH × 2 D × 2 JH) compared to the 975 possible Igμ recombinations. Furthermore, the location of Igτ 5′ of Igμ precludes intrachromosomal class switch from IgM to IgT. Since teleost AID is able to mediate class switch in murine cell lines (Barreto et al. 2005; Wakae et al. 2006), it would be

interesting to examine if class switching can occur in teleosts. Transchromosomal recombinations between different Igμ and Igω genes have been observed in nurse sharks, but how or whether this is regulated is unknown (see above).

3.8 *Beyond Sequencing, Natural Antibodies, and Specific Immune Responses in Teleost*

The use of high-throughput sequencing has made it possible to examine antibody repertoires in many vertebrates; however, to date zebrafish is the only fish species where such studies have been performed (Jiang et al. 2011; Weinstein et al. 2009). For example, in order to examine how the zebrafish repertoire changes as fish age and mature, Jiang et al. sequenced 82,000 VDJ junctions from each of 51 individual fish representing five wild-type families and five developmental time points. Young fish (two weeks post fertilization) expressed a limited repertoire and utilized ~20 % of the possible 975 VDJμ combinations. The most frequently used VDJμ combinations were found to be shared among these young fish since six or seven combinations dominated their repertoire. Such results strongly suggest that VDJμ expression is developmentally controlled. Furthermore, as the fish matured, their repertoires became more diverse; at two months of age, ~70 % of the possible combinations were used, and this level of diversity was also observed in 1-year-old fish. Consequently, the most abundant VDJ combinations previously observed were no longer shared between individuals. Presumably, this reflected the clonal expansion and increased transcription of unique activated B cells. Here one can speculate that the restricted early IgM repertoire in zebrafish and in nurse shark (see Sect. 2.8) functions in an innate fashion like natural antibodies. In humans and mice, natural antibodies are polyreactive and bind with low affinity to phosphorylcholine and carbohydrates on microbes (Benedict and Kearney 1999). These low-affinity interactions have proven essential in preventing widespread systemic infection and in priming the acquired immune responses by trapping and concentrating antigens in secondary lymphoid tissues (Ochsenbein et al. 1999). Natural antibodies also bind nucleic acids and other autoantigens and play a role in clearing damaged and apoptotic cells (reviewed in Panda and Ding 2015). One of the first clues concerning natural antibodies in ectothermic vertebrates came from studies performed in nurse and lemon sharks. In 1970 Leslie and Clem isolated anti-DNP serum antibodies from unimmunized sharks and demonstrated that these antibodies would agglutinate sheep erythrocytes (Leslie and Clem 1970). Later extensive cross-reactivity was found with anti-TNP antibodies isolated from the serum of multiple species of unimmunized sharks, sturgeons, and teleosts; besides TNP, these antibodies also bound to six other antigens including actin, myosin, thyroglobulin, and DNA (Gonzalez et al. 1988). In addition, the authors showed that cross-reactivity varied significantly among the different teleost, implying that proportions of natural antibodies will vary depending on the species. A classic example

of this is the absent or minimal specific antibody responses noted after immunizations of Atlantic cod (Magnadottir et al. 2009; Pilstrom et al. 2005). We believe that the lack of specific antibody responses combined with the well-documented absence of MHC II in Atlantic cod (Star et al. 2011; Star and Jentoft 2012) strongly suggests that natural antibodies comprise a large fraction of their serum IgM, i.e., the smaller portion is acquired T-independent IgM.

Initially, the high concentration of polyreactive antibodies in serum from both sharks and teleost made it difficult to study and detect affinity maturation by traditional ELISA. As discussed above for nurse shark (see Sect. 2.8), the details of an antibody response after immunization could only be understood once the serum immunoglobulins had been separated by size exclusion chromatography. Similarly in rainbow trout, not until affinity partitioning was employed could the direct dynamics of affinity maturation and antigen-driven selection be determined (Kaattari et al. 2002). Briefly, immune serum collected at 5, 12, 21, 27, and 34 weeks after immunization with TNP-KLH was fractionated into high-affinity and low-affinity aliquots by consecutive absorption on TNP_1-BSA sepharose beads followed by TNP_{13}-BSA sepharose beads. As measured by ELISA-based partitioning, there was a clear shift in the distribution of affinities during the immune response starting at week 5. As a result of this clonal expansion, there was a relative increase in concentration of antibodies with high affinities. This was followed later by emergence of new higher-affinity subpopulations presumably resulting from selection of somatic mutations at 12 and 27 weeks. Interestingly, the higher-affinity IgM was accompanied by a higher degree of disulfide polymerization of the tetrameric IgM as compared to the lower-affinity IgM antibodies (Kaattari et al. 2002). In 2010 Ye et al. demonstrated by in vitro antigen-driven selection that high-affinity antigen binding by the BCR resulted in more extensive disulfide polymerization of the secreted IgM. In this same study, the authors showed that fully cross-linked antibodies (tetramers) had a longer half-life than partially cross-linked tetramers (Ye et al. 2010). From this, it was postulated that higher-affinity BCR interaction leads to posttranslational modifications that result in both increased polymerization and glycosylation. In turn, these modifications could change IgM effector function, i.e., complement fixation, mucosal transport, and antibody-dependent cellular cytotoxicity (see Ye et al. 2011a). Here it should also be stated that this variability in inter-IgH chain disulfide polymerization is not unique to rainbow trout. It was initially discovered in the sheepshead in 1981 and has since been reported in at least 17 different species of teleosts (Lobb and Clem 1981a). Importantly, in mouse and humans, posttranslations modifications, such as glycosylation, are known to regulate immune responses. For example, sialylated and galactosylated IgG antibodies mediate inhibitory or anti-inflammatory responses, while non-sialylated or non-galactosylated IgG molecules are pro-inflammatory (Nimmerjahn and Ravetch 2008; Anthony et al. 2012).

Currently, mucosal immune studies are of great interest to comparative immunologists since an increased understanding of fish mucosal immunity will lead to vaccines that offer more effective protection at the site of pathogen entry. In teleosts, B cell differentiation and compartmentalization have been best studied in the

rainbow trout (reviewed in Ye et al. 2011b). The anterior kidney constitutes the primary lymphoid tissue and is equivalent to bone marrow in that it harbors long-lived plasma cells (Bromage et al. 2004). The spleen and the posterior kidney are secondary lymphoid tissues and sites of B cell activation (Zwollo et al. 2005), and memory B cells can be found in the blood and spleen (Ma et al. 2013). Compartmentalization of the rainbow trout immune response was also recently demonstrated by examining expression of IgM, IgD, and IgT following immunization with an attenuated *Flavobacterium psychrophilum* strain. In this study the authors used three different immunization routes, intra-coelomic, anal intubation, and immersion (Makesh et al. 2015), and showed the levels of IgM as measured by ELISA were significantly increased in serum, gill, and skin mucus 28 days following intra-coelomic immunization. In comparison, the transcription of secretory Igδ and Igτ was significantly upregulated in gills of fish immunized by the immersion route and in the intestines of fish immunized by anal intubation. These findings confirmed that IgT plays a role in mucosal immunity and support such a role for IgD.

The importance of IgT in gut mucosal responses has been well documented in rainbow trout where survival of natural infection with the myxozoan parasite *Ceratomyxa shasta* was accompanied by a significant increase in the number of IgT^+ B cells in the lamina propria and elevated titers of *C. shasta*-specific IgT in the gut mucus and *C. shasta*-specific IgM in the serum (Zhang et al. 2010). Similarly, an increase in specific IgT and IgT^+ B cells was observed in skin mucus and skin, respectively, in fish infected with the ciliated parasite *Ichthyophthirius multifiliis* (Ich), while the concentrations of IgM- and IgM-producing B cells remained the same. Again a high titer of specific serum IgM was observed in the surviving fish. In this context, it is important to note that both IgM and IgT polymers are transported to skin and gut mucus by a polyIg receptor, and at these sites, the immunoglobulins were found associated with a 38 kDa protein secretory component (Zhang et al. 2010; Xu et al. 2013). Notably, IgM concentrations were higher than IgT concentrations in both skin and gut mucus of normal (control) fish; however, no increase in total or specific mucus IgM could be observed in infected fish. From these results, one could speculate that in rainbow trout, the mucus IgM may function as a first line of defense to immobilize microbes until a specific IgT response is mounted. In contrast in catfish, clearly tetrameric IgM is the specific antibody in cutaneous mucus. Following intra-coelomic immunization with *I. multifiliis* antigens or skin infection, there was an observed 20-fold increase in the number of IgM-secreting B cells, and *I. multifiliis*-specific IgM could be detected 3–9 weeks after immunization. These antibodies were shown to mediate long-term protection against reinfection (Zhao et al. 2008), and importantly memory persisted 2–3 years after initial immunization. This memory was demonstrated by (1) a faster antibody response since specific antibodies were present 7 days after Ich exposure and (2) protection from or rapid clearance of the infection (Findly et al. 2013).

Currently and for comparison, little is known about mucosal immunity in sharks. However, based on the finding that Igω mRNA is expressed in the pancreas of nurse shark, spiny dogfish, and banded houndshark, *Triakis scyllium*, it has been

speculated that IgW is secreted with pancreatic juice into the lumen of the gut, and in this way, it may play a role in protecting mucosal surfaces (Rumfelt et al. 2004a; Smith et al. 2012; Honda et al. 2010). Interestingly, in situ hybridization demonstrated that Igω is co-expressed with J-chain in the adult epigonal tissue, suggesting that IgW may be expressed as a multimer (Castro et al. 2013). Thus, it would be interesting to know if such co-expression also occurs in the pancreas.

4 Remarks, Where to Go from Here

During the past 20 years, there has been substantial advancement in our understanding of the diverse immunoglobulins present in fish, their Ig gene loci and repertoires. Plus in teleosts, we have made significant progress in our understanding of B cell development and in defining Ig promoters, enhancers, and their transcription factors. Furthermore, the development of isotype-specific mAbs has and will continue to aid in defining the specialized effector functions of the unique Ig classes found in fish. Even so, important questions remain. For example, what are the "species-specific" functions of IgW/IgD? How many B cell lineages are there? Do multiple T-independent pathways for B cell activation exist? Of these three questions, this review has already touched on IgW/IgD. As for B cell lineages, IgT B cells are a distinct linage, at least in zebrafish, since they are dependent on the Ikaros transcription factor for their development (Schorpp et al. 2006). However, the relationship between 19S and 7S IgM-producing B cells in sharks is more complex. Castro et al. has proposed two models; the first suggests that 19S IgM-secreting B cells represent a separate linage akin to the B-1 cells in mice. Like natural antibody-producing B-1 B cells, shark 19S IgM-secreting B cells are present early in ontogeny and do not express the transcription factor Blimp-1. In the second model, the 19S IgM-secreting B cells represent a plasmablast differentiation stage that after antigenic stimulation will cease to express J-chain and differentiate into Blimp-1 expressing 7S IgM-secreting B cells (Castro et al. 2013). Perhaps as proposed by Steve Kaattari for rainbow trout, it is the affinity of the antigen-BCR interaction and the signaling strength that determines if the 19S B cell will differentiate into a 7S B cell in the shark. Predictably, cell separation and in vitro culture coupled with transcriptome analyses could resolve this issue. In addition, because genomic sequencing has revealed an extensive repertoire of innate receptors in fish species, we anticipate that future studies will uncover the interplay between innate and adaptive receptors in the regulation of immune cells and their effector functions.

Finally, for those who are interested, we have listed a few key references that have examined teleost antibody immune responses.

IgM redox forms in rainbow trout	– Ye et al. (2010, 2011a)
B cell development and activation in zebrafish	– Page et al. (2013)
B cell memory following infection in catfish	– Findly et al. (2013)
	– Dickerson and Findly (2014)
Antigen-presenting cells in zebrafish	– Lewis et al. (2014)
Mucosal immune responses	– Rombout et al. (2014)
	– Magadan et al. (2015)
"Poor" antibody responses in Atlantic cod	– Magnadottir et al. (2009)
	– Star and Jentoft (2012)

Acknowledgements This work was supported by grants from the US Department of Agriculture (2006-35204-16880) and (2009-651-19-05672).

References

Anderson MK, Shamblott MJ, Litman RT, Litman GW (1995) Generation of immunoglobulin light chain gene diversity in *Raja erinacea* is not associated with somatic rearrangement, an exception to a central paradigm of B cell immunity. J Exp Med 182:109–119

Anderson MK, Strong SJ, Litman RT, Luer CA, Amemiya CT, Rast JP, Litman GW (1999) A long form of the skate IgX gene exhibits a striking resemblance to the new shark IgW and IgNARC genes. Immunogenetics 49:56–67

Anthony RM, Wermeling F, Ravetch JV (2012) Novel roles for the IgG Fc glycan. Ann N Y Acad Sci 1253:170–180

Bao Y, Wang T, Guo Y, Zhao Z, Li N, Zhao Y (2010) The immunoglobulin gene loci in the teleost *Gasterosteus aculeatus*. Fish Shellfish Immunol 28:40–48

Barreto VM, Magor BG (2011) Activation-induced cytidine deaminase structure and functions: a species comparative view. Dev Comp Immunol 35:991–1007

Barreto VM, Pan-Hammarstrom Q, Zhao Y, Hammarstrom L, Misulovin Z, Nussenzweig MC (2005) AID from bony fish catalyzes class switch recombination. J Exp Med 202:733–738

Benedict CL, Kearney JF (1999) Increased junctional diversity in fetal B cells results in a loss of protective anti-phosphorylcholine antibodies in adult mice. Immunity 10:607–617

Bengtén E, Wilson M, Miller N, Clem LW, Pilström L, Warr GW (2000) Immunoglobulin isotypes: structure, function, and genetics. Curr Top Microbiol Immunol 248:189–219

Bengten E, Quiniou SM, Stuge TB, Katagiri T, Miller NW, Clem LW, Warr GW, Wilson M (2002) The IgH locus of the channel catfish, *Ictalurus punctatus*, contains multiple constant region gene sequences: different genes encode heavy chains of membrane and secreted IgD. J Immunol 169:2488–2497

Bengten E, Clem LW, Miller NW, Warr GW, Wilson M (2006a) Channel catfish immunoglobulins: repertoire and expression. Dev Comp Immunol 30:77–92

Bengten E, Quiniou S, Hikima J, Waldbieser G, Warr GW, Miller NW, Wilson M (2006b) Structure of the catfish IGH locus: analysis of the region including the single functional IGHM gene. Immunogenetics 58:831–844

Berstein RM, Schluter SF, Shen S, Marchalonis JJ (1996) A new high molecular weight immunoglobulin class from the carcharhine shark: implications for the properties of the primordial immunoglobulin. Proc Natl Acad Sci USA 93:3289–3293

Boshra H, Li J, Sunyer JO (2006) Recent advances on the complement system of teleost fish. Fish Shellfish Immunol 20:239–262

Brady BL, Steinel NC, Bassing CH (2010) Antigen receptor allelic exclusion: an update and reappraisal. J Immunol 185:3801–3808

Bromage ES, Kaattari IM, Zwollo P, Kaattari SL (2004) Plasmablast and plasma cell production and distribution in trout immune tissues. J Immunol 173:7317–7323

Brown GD, Kaattari IM, Kaattari SL (2006) Two new Ig VH gene families in *Oncorhynchus mykiss*. Immunogenetics 58:933–936

Butler JE (2006) Special Issue: Antibody repertoire development. Dev Comp Immunol 30:1–247

Castro CD, Ohta Y, Dooley H, Flajnik MF (2013) Noncoordinate expression of J-chain and Blimp-1 define nurse shark plasma cell populations during ontogeny. Eur J Immunol 43:3061–3075

Castro R, Bromage E, Abos B, Pignatelli J, Gonzalez Granja A, Luque A, Tafalla C (2014) CCR7 is mainly expressed in teleost gills, where it defines an IgD + IgM-B lymphocyte subset. J Immunol 192:1257–1266

Chen K, Xu W, Wilson M, He B, Miller N, Bengten E, Edholm E-S, Santini P, Rath P, Chiu A, Cattalini M, Litzman J, Bussel J, Huang B, Meini A, Riesbeck K, Cunningham-Rundles C, Plebani A, Cerutti A (2009) Immunoglobulin D enhances immune surveillance by activation antimicrobial, pro-inflammatory and B cell-stimulating programs in basophils. Nat Immunol 10(8):889–98

Clem LW (1971) Phylogeny of immunoglobulin structure and function. IV. Immunoglobulins of the giant grouper, *Epinephelus itaira*. J Biol Chem 246:9–15

Clem LW, Leslie GA (1971) Production of 19S IgM antibodies with restricted heterogeneity from sharks. Proc Natl Acad Sci USA 68:139–141

Clem LW, Small PA Jr (1967) Phylogeny of immunoglobulin structure and function. I. Immunoglobulins of the lemon shark. J Exp Med 125:893–920

Clem IW, De Boutaud F, Sigel MM (1967) Phylogeny of immunoglobulin structure and function. II. Immunoglobulins of the nurse shark. J Immunol 99:1226–1235

Criscitiello MF (2014) What the shark immune system can and cannot provide for the expanding design landscape of immunotherapy. Expert Opin Drug Discov 9:725–739

Criscitiello MF, Flajnik MF (2007) Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates. Eur J Immunol 37:2683–2694

Criscitiello MF, Saltis M, Flajnik MF (2006) An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. Proc Natl Acad Sci USA 103: 5036–5041

Crouch K, Smith LE, Williams R, Cao W, Lee M, Jensen A, Dooley H (2013) Humoral immune response of the small-spotted catshark, *Scyliorhinus canicula*. Fish Shellfish Immunol 34: 1158–1169

Daggfeldt A, Bengtén E, Pilström L (1993) A cluster type organization of the loci of the immunoglobulin light chain in Atlantic cod (*Gadus morhua* L.) and rainbow trout (*Oncorhynchus mykiss* Walbaum) indicated by nucleotide sequences of cDNAs and hybridization analysis. Immunogenetics 38:199–209

Danilova N, Bussmann J, Jekosch K, Steiner LA (2005) The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat Immunol 6:295–302

Diaz M, Greenberg AS, Flajnik MF (1998) Somatic hypermutation of the new antigen receptor gene (NAR) in the nurse shark does not generate the repertoire: possible role in antigen-driven reactions in the absence of germinal centers. Proc Natl Acad Sci USA 95:14343–14348

Diaz M, Velez J, Singh M, Cerny J, Flajnik MF (1999) Mutational pattern of the nurse shark antigen receptor gene (NAR) is similar to that of mammalian Ig genes and to spontaneous mutations in evolution: the translesion synthesis model of somatic hypermutation. Int Immunol 11:825–833

Diaz M, Stanfield RL, Greenberg AS, Flajnik MF (2002) Structural analysis, selection, and ontogeny of the shark new antigen receptor (IgNAR): identification of a new locus preferentially expressed in early development. Immunogenetics 54:501–512

Dickerson HW, Findly RC (2014) Immunity to Ichthyophthirius infections in fish: a synopsis. Dev Comp Immunol 43:290–299

Dooley H, Flajnik MF (2005) Shark immunity bites back: affinity maturation and memory response in the nurse shark, Ginglymostoma cirratum. Eur J Immunol 35:936–945

Dooley H, Flajnik MF (2006) Antibody epertoire development in cartilaginous fish. Dev Comp Immunol 30:43–56

Edholm ES, Wilson M, Sahoo M, Miller NW, Pilström L, Wermenstam NE, Bengtén E (2009) Identification of Igsigma and Iglambda in channel catfish, *Ictalurus punctatus*, and Iglambda in Atlantic cod, *Gadus morhua*. Immunogenetics 61:353–370

Edholm ES, Bengten E, Stafford JL, Sahoo M, Taylor EB, Miller NW, Wilson M (2010) Identification of two IgD+ B cell populations in channel catfish, *Ictalurus punctatus*. J Immunol 185:4082–4094

Edholm ES, Bengten E, Wilson M (2011a) Insights into the function of IgD. Dev Comp Immunol 35:1309–1316

Edholm ES, Wilson M, Bengten E (2011b) Immunoglobulin light (IgL) chains in ectothermic vertebrates. Dev Comp Immunol 35:906–915

Fidler JE, Clem LW, Small PA Jr (1969) Immunoglobulin synthesis in neonatal nurse sharks (Ginglymostoma cirratum). Comp Biochem Physiol 31:365–371

Fillatreau S, Six A, Magadan S, Castro R, Sunyer JO, Boudinot P (2013) The astonishing diversity of Ig classes and B cell repertoires in teleost fish. Front Immunol 4:28

Findly RC, Zhao X, Noe J, Camus AC, Dickerson HW (2013) B cell memory following infection and challenge of channel catfish with *Ichthyophthirius multifiliis*. Dev Comp Immunol 39: 302–311

Flajnik M, Du Pasquier L (2013) Evolution of the immune system. In: Paul WE (ed) Fundamental immunology. Lippincott Williams & Wilkins, Philadelphia, PA, pp 67–128

Flajnik MF, Rumfelt LL (2000) The immune system of cartilaginous fish. Curr Top Microbiol Immunol 248:249–270

Fleurant M, Changchien L, Chen CT, Flajnik MF, Hsu E (2004) Shark Ig light chain junctions are as diverse as in heavy chains. J Immunol 173:5574–5582

Fu X, Zhang H, Tan E, Watabe S, Asakawa S (2015) Characterization of the torafugu (*Takifugu rubripes*) immunoglobulin heavy chain gene locus. Immunogenetics 67:179–193

Gambon-Deza F, Sanchez-Espinel C, Magadan-Mompo S (2010) Presence of an unique IgT on the IGH locus in three-spined stickleback fish (*Gasterosteus aculeatus*) and the very recent generation of a repertoire of VH genes. Dev Comp Immunol 34:114–122

Ghaffari SH, Lobb CJ (1993) Structure and genomic organization of immunoglobulin light chain in the channel catfish. An unusual genomic organizational pattern of segmental genes. J Immunol 151:6900–6912

Ghaffari SH, Lobb CJ (1997) Structure and genomic organization of a second class of immunoglobulin light chain genes in the channel catfish. J Immunol 159:250–258

Gomez D, Sunyer JO, Salinas I (2013) The mucosal immune system of fish: the evolution of tolerating commensals while fighting pathogens. Fish Shellfish Immunol 35:1729–1739

Gonzalez R, Charlemagne J, Mahana W, Avrameas S (1988) Specificity of natural serum antibodies present in phylogenetically distinct fish species. Immunology 63:31–36

Greenberg AS, Steiner L, Kasahara M, Flajnik MF (1993) Isolation of a shark immnoglobulin light chain cDNA clone encoding a protein resembling mammalian k light chains: implications for the evolution of light chains. Proc Natl Acad Sci USA 90:10603–10607

Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374:168–173

Greenberg AS, Hughes AL, Guo J, Avila D, McKinney EC, Flajnik MF (1996) A novel "chimeric" antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin. Eur J Immunol 26:1123–1129

Hansen JD, Landis ED, Phillips RB (2005) Discovery of a unique Ig heavy chain isotype in rainbow trout: implications for a novel B-cell developmental pathway in teleost fish. Proc Natl Acad Sci USA 102:6919–6924

Harding FA, Amemiya CT, Litman RT, Cohen N, Litman GW (1990) Two distinct immunoglobulin heavy chain isotypes in a primitive, cartilaginous fish, *Raja erinacea*. Nucl Acid Res 18:6369–6376

Haynes L, Fuller L, McKinney EC (1988) Fc receptor for shark IgM. Dev Comp Immunol 12: 561–571

Hikima J, Jung TS, Aoki T (2011) Immunoglobulin genes and their transcriptional control in teleosts. Dev Comp Immunol 35:924–936

Hinds KR, Litman GW (1986) Major reorganization of immunoglobulin VH-segmental elements during vertebrate evolution. Nature 320:546–549

Hohman VS, Schluter SF, Marchalonis JJ (1992) Complete sequence of a cDNA clone specifying sanbar shark immunoglobulin light chain: gene organisation and implications for the evolution of light chains. Proc Natl Acad Sci USA 89:276–280

Hohman VS, Schuchman DB, Schluter SF, Marchalonis JJ (1993) Genomic clone for sandbar shark l light chain: Generation of diversity in the absence of gene rearrangement. Proc Natl Acad Sci USA 90:9882–9886

Hohman VS, Schluter SF, Marchalonis JJ (1995) Diversity of Ig light chain clusters in the sandbar shark (*Carcharhinus plumbeus*). J Immunol 155:3922–3928

Honda Y, Kondo H, Caipang CM, Hirono I, Aoki T (2010) cDNA cloning of the immunoglobulin heavy chain genes in banded houndshark Triakis scyllium. Fish Shellfish Immunol 29:854–861

Hsu E, Criscitiello MF (2006) Diverse immunoglobulin light chain organizations in fish retain potential to revise B cell receptor specificities. J Immunol 177:2452–2462

Hsu E, Pulham N, Rumfelt LL, Flajnik MF (2006) The plasticity of immunoglobulin gene systems in evolution. Immunol Rev 210:8–26

Jiang N, Weinstein JA, Penland L, White RA III, Fisher DS, Quake SR (2011) Determinism and stochasticity during maturation of the zebrafish antibody repertoire. Proc Natl Acad Sci USA 108:5348–5353

Jones JC, Ghaffari SH, Lobb CJ (2004) Patterns of gene divergence and VL promoter activity in immunoglobulin light chain clusters of the channel catfish. Immunogenetics 56:448–461

Kaattari SL, Zhang HL, Khor IW, Kaattari IM, Shapiro DA (2002) Affinity maturation in trout: clonal dominance of high affinity antibodies late in the immune response. Dev Comp Immunol 26:191–200

Klapper DG, Clem LW, Small PA Jr (1971) Proteolytic fragmentation of elasmobranch immunoglobulins. Biochemistry 10:645–652

Kobayashi K, Tomonaga S, Kajii T (1984) A second class of immunoglobulin other than IgM present in the serum of a cartilaginous fish, the skate *Raja kenojei*: isolation and characterization. Mol Immunol 21:397–404

Kokubu F, Hinds K, Litman R, Shamblott MJ, Litman GW (1987) Extensive families of constant region genes in a phylogenetically primitive vertebrate indicate an additional level of immunoglobulin complexity. Proc Natl Acad Sci USA 84:5868–5872

Kokubu F, Litman R, Shamblott MJ, Hinds K, Litman GW (1988) Diverse organisation of immunoglobulin VH gene loci in a primitive vertebrate. EMBO J 7:3413–3422

Kovalenko OV, Olland A, Piche-Nicholas N, Godbole A, King D, Svenson K, Calabro V, Muller MR, Barelle CJ, Somers W, Gill DS, Mosyak L, Tchistiakova L (2013) Atypical antigen recognition mode of a shark immunoglobulin new antigen receptor (IgNAR) variable domain characterized by humanization and structural analysis. J Biol Chem 288:17408–17419

Lange MD, Waldbieser GC, Lobb CJ (2009) Patterns of receptor revision in the immunoglobulin heavy chains of a teleost fish. J Immunol 182:5605–5622

Lee SS, Fitch D, Flajnik MF, Hsu E (2000a) Rearrangement of immunoglobulin genes in shark germ cells. J Exp Med 191:1637–1648

Lee SS, Greenberg A, Hsu E (2000b) Evolution and somatic diversification of immunoglobulin light chains. Curr Top Microbiol Immunol 248:285–300

Lee SS, Tranchina D, Ohta Y, Flajnik MF, Hsu E (2002) Hypermutation in shark immunoglobulin light chain genes results in contiguous substitutions. Immunity 16:571–582

Lee V, Huang JL, Lui MF, Malecek K, Ohta Y, Mooers A, Hsu E (2008) The evolution of multiple isotypic IgM heavy chain genes in the shark. J Immunol 180:7461–7470

Leslie GA, Clem LW (1970) Reactivity of normal shark immunoglobulins with nitrophenyl ligands. J Immunol 105:1547–1552

Lewis KL, Del Cid N, Traver D (2014) Perspectives on antigen presenting cells in zebrafish. Dev Comp Immunol 46:63–73

Litman GW, Berger L, Murphy K, Litman R, Hinds K, Erickson BW (1985) Immunoglobulin VH-gene structure and diversity in Heterodontus, a phylogenetically primitive shark. Proc Natl Acad Sci USA 82:2082–2086

Litman GW, Rast JP, Shamblott MJ, Haire RN, Hulst M, Roess W, Litman R, Hinds-Frey KR, Zilch A, Amemiya CT (1993) Phylogenetic diversification of immunoglobulin genes and antibody repertoire. Mol Biol Evol 10:60–72

Litman GW, Anderson MK, Rast JP (1999) Evolution of antigen binding receptors. Annu Rev Immunol 17:109–147

Lobb CJ, Clem LW (1981a) Phylogeny of immunoglobulin structure and function - X. Humoral immunoglobulins of the sheephead, *Archosargus probatocephalus*. Dev Comp Immunol 5:271–282

Lobb CJ, Clem LW (1981b) Phylogeny of immunoglobulin structure and function – XI. Secretory immunoglobulins in the cutaneous mucus of the sheephead, *Archosargus probatocephalus*. Dev Comp Immunol 5:587–596

Lobb CJ, Hayman JR (1989) Activation of complement by different immunoglobulin heavy chain isotypes of the channel catfish (*Ictalurus punctatus*). Mol Immunol 26:457–465

Lundqvist M, Bengten E, Strömberg S, Pilström L (1996) Ig light chain gene in the Siberian sturgeon (*Acipenser baeri*). J Immunol 157:2031–2038

Lutz C, Ledermann B, Kosco-Vilbois MH, Ochsenbein AF, Zinkernagel RM, Kohler G, Brombacher F (1998) IgD can largely substitute for loss of IgM function in B cells. Nature 393:797–801

Ma C, Ye J, Kaattari SL (2013) Differential compartmentalization of memory B cells versus plasma cells in salmonid fish. Eur J Immunol 43:360–370

Magadan-Mompo S, Sanchez-Espinel C, Gambon-Deza F (2011) Immunoglobulin heavy chains in medaka (*Oryzias latipes*). BMC Evol Biol 11:165

Magadan S, Sunyer OJ, Boudinot P (2015) Unique features of fish immune repertoires: particularities of the adaptive immunity within the largest group of vertebrates. In: Hsu E, Du Pasquier L (eds) Pathogen-host interactions: antigenic variation vs. somatic adaptations. Springer, New York

Magnadottir B, Gudmundsdottir S, Gudmundsdottir BK, Helgason S (2009) Natural antibodies of cod (*Gadus morhua L.*): specificity, activity and affinity. Comp Biochem Physiol B Biochem Mol Biol 154:309–316

Magor KE (2011) Immunoglobulin genetics and antibody responses to influenza in ducks. Dev Comp Immunol 35:1008–1016

Makesh M, Sudheesh PS, Cain KD (2015) Systemic and mucosal immune response of rainbow trout to immunization with an attenuated Flavobacterium psychrophilum vaccine strain by different routes. Fish Shellfish Immunol 44:156–163

Malecek K, Brandman J, Brodsky JE, Ohta Y, Flajnik MF, Hsu E (2005) Somatic hypermutation and junctional diversification at Ig heavy chain loci in the nurse shark. J Immunol 175: 8105–8115

Malecek K, Lee V, Feng W, Huang JL, Flajnik MF, Ohta Y, Hsu E (2008) Immunoglobulin heavy chain exclusion in the shark. PLoS Biol 6, e157

Marchalonis J, Edelman GM (1965) Phylogenetic origins of antibody structure. I. Multichain structure of immunoglobulins in the smooth dogfish (*Mustelus canis*). J Exp Med 122:601–618

Marchalonis J, Edelman GM (1966) Polypeptide Chains of Immunoglobulins from the Smooth Dogfish (*Mustela canis*). Science 154:1567–1568

Marianes AE, Zimmerman AM (2011) Targets of somatic hypermutation within immunoglobulin light chain genes in zebrafish. Immunology 132:240–255

Mashoof S, Pohlenz C, Chen PL, Deiss TC, Gatlin D III, Buentello A, Criscitiello MF (2014) Expressed IgH mu and tau transcripts share diversity segment in ranched *Thunnus orientalis*. Dev Comp Immunol 43:76–86

McCumber LJ, Clem LW (1976) A comparative study of J chain structure and stoichiometry in human and nurse shark IgM. Immunochemistry 13:479–484

McKinney EC, Flajnik MF (1997) IgM-mediated opsonization and cytotoxicity in the shark. J Leukoc Biol 61:141–146

Miracle AL, Anderson MK, Litman RT, Walsh CJ, Luer CA, Rothenberg EV, Litman GW (2001) Complex expression patterns of lymphocyte-specific genes during the development of cartilaginous fish implicate unique lymphoid tissues in generating an immune repertoire. Int Immunol 13:567–580

Nemazee D, Weigert M (2000) Revising B cell receptors. J Exp Med 191:1813–1817

Nimmerjahn F, Ravetch JV (2008) Anti-inflammatory actions of intravenous immunoglobulin. Annu Rev Immunol 26:513–533

Nitschke L, Kosco MH, Kohler G, Lamers MC (1993) Immunoglobulin D-deficient mice can mount normal immune responses to thymus-independent and -dependent antigens. Proc Natl Acad Sci USA 90:1887–1891

Ochsenbein AF, Fehr T, Lutz C, Suter M, Brombacher F, Hengartner H, Zinkernagel RM (1999) Control of early viral and bacterial distribution and disease by natural antibodies. Science 286:2156–2159

Ohta Y, Flajnik M (2006) IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proc Natl Acad Sci USA 103:10723–10728

Ota T, Rast JP, Litman GW, Amemiya CT (2003) Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. Proc Natl Acad Sci USA 100:2501–2506

Page DM, Wittamer V, Bertrand JY, Lewis KL, Pratt DN, Delgado N, Schale SE, McGue C, Jacobsen BH, Doty A, Pao Y, Yang H, Chi NC, Magor BG, Traver D (2013) An evolutionarily conserved program of B-cell development and activation in zebrafish. Blood 122:e1–e11

Panda S, Ding JL (2015) Natural antibodies bridge innate and adaptive immunity. J Immunol 194:13–20

Partula S, Schwager J, Timmusk S, Pilström L, Charlemagne J (1996) A second immunoglobulin light chain isotype in the rainbow trout. Immunogenetics 45:44–51

Pettinello R, Dooley H (2014) The immunoglobulins of cold-blooded vertebrates. Biomolecules 4:1045–1069

Pilstrom L, Warr GW, Stromberg S (2005) Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation. Fish Sci 71:961–971

Quiniou SM, Wilson M, Boudinot P (2011) Processing of fish Ig heavy chain transcripts: diverse splicing patterns and unusual nonsense mediated decay. Dev Comp Immunol 35:949–958

Ramirez-Gomez F, Greene W, Rego K, Hansen JD, Costa G, Kataria P, Bromage ES (2012) Discovery and characterization of secretory IgD in rainbow trout: secretory IgD is produced through a novel splicing mechanism. J Immunol 188:1341–1349

Rast JP, Anderson MK, Ota T, Litman RT, Margittai M, Shamblott MJ, Litman GW (1994) Immunoglobulin light chain class multiplicity and alternative organizational forms in early vertebrate phylogeny. Immunogenetics 40:83–99

Rast JP, Amemiya CT, Litman RT, Strong SJ, Litman GW (1998) Distinct patterns of IgH structure and organization in a divergent lineage of chrondrichthyan fishes. Immunogenetics 47:234–245

Rock EP, Sibbald PR, Davis MM, Chien YH (1994) CDR3 length in antigen-specific immune receptors. J Exp Med 179:323–328

Roes J, Rajewsky K (1993) Immunoglobulin D (IgD)-deficient mice reveal an auxiliary receptor function for IgD in antigen-mediated recruitment of B cells. J Exp Med 177:45–55

Roman T, Andersson E, Bengten E, Hansen J, Kaattari S, Pilström L, Charlemagne J, Matsunaga T (1996) Unified nomenclature of Ig VH genes in rainbow trout (*Oncorhynchus mykiss*): definition of eleven VH families. Immunogenetics 43:325–326

Rombout JH, Yang G, Kiron V (2014) Adaptive immune responses at mucosal surfaces of teleost fish. Fish Shellfish Immunol 40:634–643

Roux KH, Greenberg AS, Greene L, Strelets L, Avila D, McKinney EC, Flajnik MF (1998) Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. Proc Natl Acad Sci USA 95:11804–11809

Rumfelt LL, Avila D, Diaz M, Bartl S, McKinney EC, Flajnik MF (2001) A shark antibody heavy chain encoded by a nonsomatically rearranged VDJ is preferentially expressed in early development and is convergent with mammalian IgG. Proc Natl Acad Sci USA 98:1775–1780

Rumfelt LL, McKinney EC, Taylor E, Flajnik MF (2002) The development of primary and secondary lymphoid tissues in the nurse shark Ginglymostoma cirratum: B-cell zones precede dendritic cell immigration and T-cell zone formation during ontogeny of the spleen. Scand J Immunol 56:130–148

Rumfelt LL, Diaz M, Lohr RL, Mochon E, Flajnik MF (2004a) Unprecedented multiplicity of Ig transmembrane and secretory mRNA forms in the cartilaginous fish. J Immunol 173:1129–1139

Rumfelt LL, Lohr RL, Dooley H, Flajnik MF (2004b) Diversity and repertoire of IgW and IgM VH families in the newborn nurse shark. BMC Immunol 5:8

Saada R, Weinberger M, Shahaf G, Mehr R (2007) Models for antigen receptor gene rearrangement: CDR3 length. Immunol Cell Biol 85:323–332

Saha NR, Ota T, Litman GW, Hansen J, Parra Z, Hsu E, Buonocore F, Canapa A, Cheng JF, Amemiya CT (2014) Genome complexity in the coelacanth is reflected in its adaptive immune system. J Exp Zool B Mol Dev Evol 322:438–463

Sahoo M, Edholm ES, Stafford JL, Bengten E, Miller NW, Wilson M (2008) B cell receptor accessory molecules in the channel catfish, *Ictalurus punctatus*. Dev Comp Immunol 32: 1385–1397

Savan R, Aman A, Sato K, Yamaguchi R, Sakai M (2005) Discovery of a new class of immunoglobulin heavy chain from fugu. Eur J Immunol 35:3320–3331

Schluter SF, Hohman VS, Edmundson AB, Marchalonis JJ (1989) Evolution of immunoglobulin light chains: cDNA clones specifying sandbar shark constant regions. Proc Natl Acad Sci USA 86:9961–9965

Schorpp M, Bialecki M, Diekhoff D, Walderich B, Odenthal J, Maischein HM, Zapata AG, Boehm T (2006) Conserved functions of Ikaros in vertebrate lymphocyte development: genetic evidence for distinct larval and adult phases of T cell development and two lineages of B cells in zebrafish. J Immunol 177:2463–2476

Shamblott MJ, Litman GW (1989a) Complete nucleotide sequence of primitive vertebrate immunoglobulin light chain genes. Proc Natl Acad Sci USA 86:4684–4688

Shamblott MJ, Litman GW (1989b) Genomic organization and sequences of immunoglobulin light chain genes in a primitive vertebrate suggest coevolution of immunoglobulin gene organization. EMBO J 8:3733–3739

Shen SX, Bernstein RM, Schluter SF, Marchalonis JJ (1996) Heavy-chain variable regions in carcharhine sharks: development of a comprehensive model for the evolution of VH domains among the gnathanstomes. Immunol Cell Biol 74:357–364

Shen L, Stuge TB, Evenhuis JP, Bengten E, Wilson M, Chinchar VG, Clem LW, Miller NW (2003) Channel catfish NK-like cells are armed with IgM via a putative FcmicroR. Dev Comp Immunol 27:699–714

Sledge C, Clem LW, Hood L (1974) Antibody structure: amino terminal sequences of nurse shark light and heavy chains. J Immunol 112:941–948

Small PA, Klapper DG, Clem LW (1970) Half-lives, body distribution and lack of interconversion of serum 19S and 7S IgM of sharks. J Immunol 105:29–37

Smith SL (1998) Shark complement: an assessment. Immunol Rev 166:67–78

Smith LE, Crouch K, Cao W, Muller MR, Wu L, Steven J, Lee M, Liang M, Flajnik MF, Shih HH, Barelle CJ, Paulsen J, Gill DS, Dooley H (2012) Characterization of the immunoglobulin repertoire of the spiny dogfish (*Squalus acanthias*). Dev Comp Immunol 36:665–679

Smith SL, Sim RB, Flajnik MF (2015) Immunobiology of the shark. CRC Press Taylor & Francis Group, Boca Raton, FL

Stanfield RL, Dooley H, Flajnik MF, Wilson IA (2004) Crystal structure of a shark single-domain antibody V region in complex with lysozyme. Science 305:1770–1773

Stanfield RL, Dooley H, Verdino P, Flajnik MF, Wilson IA (2007) Maturation of shark single-domain (IgNAR) antibodies: evidence for induced-fit binding. J Mol Biol 367:358–372

Star B, Jentoft S (2012) Why does the immune system of Atlantic cod lack MHC II? Bioessays 34:648–651

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzen A, Winer R, Knight J, Vogel JH, Aken B, Andersen O, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjoen T, Kuhl H, Thorsen J, Malde K, Reinhardt R, Du L, Johansen SD, Searle S, Lien S, Nilsen F, Jonassen I, Omholt SW, Stenseth NC, Jakobsen KS (2011) The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210

Stenvik J, Jorgensen TO (2000) Immunoglobulin D (IgD) of Atlantic cod has a unique structure. Immunogenetics 51:452–461

Streltsov VA, Varghese JN, Carmichael JA, Irving RA, Hudson PJ, Nuttall SD (2004) Structural evidence for evolution of shark Ig new antigen receptor variable domain antibodies from a cell-surface receptor. Proc Natl Acad Sci USA 101:12444–12449

Streltsov VA, Carmichael JA, Nuttall SD (2005) Structure of a shark IgNAR antibody variable domain and modeling of an early-developmental isotype. Protein Sci 14:2901–2909

Sun Y, Liu Z, Li Z, Lian Z, Zhao Y (2012) Phylogenetic conservation of the 3' cryptic recombination signal sequence (3'cRSS) in the VH genes of jawed vertebrates. Front Immunol 3:392

Sunyer JO (2013) Fishing for mammalian paradigms in the teleost immune system. Nat Immunol 14:320–326

Tomonaga S, Kobayashi K (1985) A second class of immunoglobulin in the cartilaginous fishes. Dev Comp Immunol 9:797–802

Tomonaga S, Kobayashi K, Kajii T, Awaya K (1984) Two populations of immunoglobulin-forming cells in the skate, Raja kenojei: their distribution and characterization. Dev Comp Immunol 8:803–812

Timmusk S, Partula S, Pilström L (2000) Different genomic organization and expression of immunoglobulin light-chain isotypes in the rainbow trout. Immunogenetics 51:905–914

Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay BH, Tohari S, Kong KW, Ho S, Lorente-Galdos B, Quilez J, Marques-Bonet T, Raney BJ, Ingham PW, Tay A, Hillier LW, Minx P, Boehm T, Wilson RK, Brenner S, Warren WC (2014) Elephant shark genome provides unique insights into gnathostome evolution. Nature 505:174–179

Wakae K, Magor BG, Saunders H, Nagaoka H, Kawamura A, Kinoshita K, Honjo T, Muramatsu M (2006) Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. Int Immunol 18:41–47

Weinstein JA, Jiang N, White RA III, Fisher DS, Quake SR (2009) High-throughput sequencing of the zebrafish antibody repertoire. Science 324:807–810

Wilson MR, Ross DA, Miller NW, Clem LW, Middleton DL, Warr GW (1995) Alternate pre-mRNA processing pathways in the production of membrane IgM heavy chains in holostean fish. Dev Comp Immunol 19:165–177

Wilson M, Bengten E, Miller NW, Clem LW, Du Pasquier L, Warr GW (1997) A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. Proc Natl Acad Sci USA 94:4593–4597

Wilson PC, Wilson K, Liu YJ, Banchereau J, Pascual V, Capra JD (2000) Receptor revision of immunoglobulin heavy chain variable region genes in normal human B lymphocytes. J Exp Med 191:1881–1894

Xu Z, Parra D, Gomez D, Salinas I, Zhang YA, von Gersdorff Jorgensen L, Heinecke RD, Buchmann K, LaPatra S, Sunyer JO (2013) Teleost skin, an ancient mucosal surface that elicits gut-like immune responses. Proc Natl Acad Sci USA 110:13097–13102

Yang F, Waldbieser GC, Lobb CJ (2006) The nucleotide targets of somatic mutation and the role of selection in immunoglobulin heavy chains of a teleost fish. J Immunol 176:1655–1667

Yasuike M, de Boer J, von Schalburg KR, Cooper GA, McKinnel L, Mesmer A, So S, Davidson WS, Koop BF (2010) Evolution of duplicated IgH loci in Atlantic salmon, *Salmo salar*. BMC Genomics 11:486

Ye J, Bromage ES, Kaattari SL (2010) The strength of B cell interaction with antigen determines the degree of IgM polymerization. J Immunol 184:844–850

Ye J, Bromage E, Kaattari I, Kaattari S (2011a) Transduction of binding affinity by B lymphocytes: a new dimension in immunological regulation. Dev Comp Immunol 35:982–990

Ye J, Kaattari I, Kaattari S (2011b) Plasmablasts and plasma cells: reconsidering teleost immune system organization. Dev Comp Immunol 35:1273–1281

Yoder JA, Litman GW (2000) Immune-type diversity in the absence of somatic rearrangement. Curr Top Microbiol Immunol 248:271–282

Zhang YA, Salinas I, Li J, Parra D, Bjork S, Xu Z, LaPatra SE, Bartholomew J, Sunyer JO (2010) IgT, a primitive immunoglobulin class specialized in mucosal immunity. Nat Immunol 11:827–835

Zhang C, Du Pasquier L, Hsu E (2013) Shark IgW C region diversification through RNA processing and isotype switching. J Immunol 191:3410–3418

Zhang T, Tacchi L, Wei Z, Zhao Y, Salinas I (2014) Intraclass diversification of immunoglobulin heavy chain genes in the African lungfish. Immunogenetics 66:335–351

Zhao X, Findly RC, Dickerson HW (2008) Cutaneous antibody-secreting cells and B cells in a teleost fish. Dev Comp Immunol 32:500–508

Zhu C, Hsu E (2010) Error-prone DNA repair activity during somatic hypermutation in shark B lymphocytes. J Immunol 185:5336–5347

Zhu C, Feng W, Weedon J, Hua P, Stefanov D, Ohta Y, Flajnik MF, Hsu E (2011) The multiple shark Ig H chain genes rearrange and hypermutate autonomously. J Immunol 187:2492–2501

Zhu C, Lee V, Finn A, Senger K, Zarrin AA, Du Pasquier L, Hsu E (2012) Origin of immunoglobulin isotype switching. Curr Biol 22:872–880

Zhu L, Yan Z, Feng M, Peng D, Guo Y, Hu X, Ren L, Sun Y (2014) Identification of sturgeon IgD bridges the evolutionary gap between elasmobranchs and teleosts. Dev Comp Immunol 42: 138–147

Zielonka S, Empting M, Grzeschik J, Konning D, Barelle CJ, Kolmar H (2015) Structural insights and biomedical potential of IgNAR scaffolds from sharks. MAbs 7:15–25

Zimmerman AM, Yeo G, Howe K, Maddox BJ, Steiner LA (2008) Immunoglobulin light chain (IgL) genes in zebrafish: Genomic configurations and inversional rearrangements between (V(L)-J(L)-C(L)) gene clusters. Dev Comp Immunol 32:421–434

Zimmerman AM, Moustafa FM, Romanowski KE, Steiner LA (2011a) Zebrafish immunoglobulin IgD: unusual exon usage and quantitative expression profiles with IgM and IgZ/T heavy chain isotypes. Mol Immunol 48:2220–2223

Zimmerman AM, Romanowski KE, Maddox BJ (2011b) Targeted annotation of immunoglobulin light chain (IgL) genes in zebrafish from BAC clones reveals kappa-like recombining/deleting elements within IgL constant regions. Fish Shellfish Immunol 31:697–703

Zwollo P, Cole S, Bromage E, Kaattari S (2005) B cell heterogeneity in the teleost kidney: evidence for a maturation gradient from anterior to posterior kidney. J Immunol 174:6608–6616

Unique Features of Fish Immune Repertoires: Particularities of Adaptive Immunity Within the Largest Group of Vertebrates

Susana Magadan, Oriol J. Sunyer, and Pierre Boudinot

An immune system is characterized by two linked properties: a somatic learning process to make a self–nonself discrimination and a mechanism for determining the class of the response that optimally rids the target
Melvin Cohn (Cohn 1994)

Abstract Fishes (i.e., teleost fishes) are the largest group of vertebrates. Although their immune system is based on the fundamental receptors, pathways, and cell types found in all groups of vertebrates, fishes show a diversity of particular features that challenge some classical concepts of immunology. In this chapter, we discuss the particularities of fish immune repertoires from a comparative perspective. We examine how allelic exclusion can be achieved when multiple Ig loci are present, how isotypic diversity and functional specificity impact clonal complexity, how loss of the MHC class II molecules affects the cooperation between T and B cells, and how deep sequencing technologies bring new insights about somatic hypermutation in the absence of germinal centers. The unique coexistence of two distinct B-cell lineages respectively specialized in systemic and mucosal responses is also discussed. Finally, we try to show that the diverse adaptations of immune repertoires in teleosts can help in understanding how somatic adaptive mechanisms of immunity evolved in parallel in different lineages across vertebrates.

S. Magadan (✉) • P. Boudinot (✉)
Virologie et Immunologie Moléculaires, Institut National de la Recherche Agronomique, Jouy-en-Josas, France
e-mail: Susana.Magadan@jouy.inra.fr; Pierre.Boudinot@jouy.inra.fr

O.J. Sunyer
School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, USA

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_10

1 Introduction

When a pathogen confronts a host organism, it stimulates defense mechanisms leading to its elimination, reduction, or containment. Besides innate pathways based on factors encoded in the genome as "ready to use" units, systems that undergo somatic modifications offer opportunities of focused "adaptive" responses to pathogens. Specific recognition of antigens by lymphocyte receptors diversified through VDJ somatic rearrangements is the archetype of such systems of adaptive response.

During their differentiation, immunoglobulin (Ig)—respectively, T-cell receptor (TR)—loci are subjected to random genomic rearrangements of V, D, and J gene segments, leading to the expression of a unique antigen receptor by each lymphocyte. The universe of antigenic motifs is matched to a large population of lymphocytes through the specific recognition of an epitope by a given receptor unique to a lymphocyte clone. During the differentiation of lymphocytes, epitope-specific, selective processes lead to deletion of most autoreactive cells and expansion of mature T and B cells. These populations are dramatically affected during antigen-driven responses, for example, during pathogen infections, as clones specific of pathogen epitopes are expanded.

The concept of immune repertoire was created to describe the diversity of lymphocyte receptors involved in this network of interactions. Niels Jerne referred to the immune repertoire as a dual concept integrating both the potential diversity allowed by the genetic resources of the genome and the available set of receptors expressed in a given tissue at a given moment (Jerne 1971). This notion of "immune repertoires" represents a useful tool to describe lymphocyte and receptor populations, their development, and their modifications by responses to infections.

In teleost fish, three immunoglobulin classes have been described: IgM, IgD, and IgT. While IgM constitutes the main systemic immunoglobulin, IgT plays the prevalent role in mucosal surfaces. The role of IgD in fish immunity remains to be elucidated. Both IgM and IgD are co-expressed in B cells found both in systemic and mucosal lymphoid areas, whereas IgT is uniquely expressed by a B-cell subset devoid of IgM and IgD expression. The three isotypes share the same genomic repertoire of VH gene segments; IgM and IgD heavy chains are generally produced by alternative splicing of constant exons from a common long transcript, while mRNAs for IgT heavy chains are transcribed from another genomic region. Importantly, B lymphocytes express either IgM and IgD or IgT, which distinguishes two fundamental B-cell subsets. However, it has been shown that a third B-cell subset uniquely expressing IgD exists both in catfish and rainbow trout. TCR isotypes are much more conserved across vertebrates compared to immunoglobulins, and for the time being two main T-cell subsets, which, respectively, express αβ and γδ T-cell receptors, have been described in teleosts.

In this review, we examine a number of distinctive features of B- and T-cell immunity in fishes and show how repertoire studies shed light on particular somatic adaptations found in these animals. We first review the great diversity of teleost fishes and the canonical features of adaptive immunity they share with other (jaw)

vertebrates. We then discuss selected mechanisms or features that represent distinctive adaptations of the fish immune system. Finally, we consider how these adaptations can help in understanding how the somatic adaptive mechanisms of immunity evolved in parallel in different lineages.

2 Common Conserved Features of Immune Repertoires Across the Great Diversity of Teleost Fishes

2.1 The Diversity of Fishes

Following Nelson (Nelson 2006), a fish is "a poikilothermic vertebrate with gills and with limbs in the shape of fins." In this chapter we focus on bony fishes which are by far the largest group of vertebrates with more than 26,000 species (Helfman et al. 2010), while there are about 10,000 species of birds and 5000 species of mammals. The diversity of shape, size (from 8 to 10 mm gobies and *Danionella* to very large sunfishes, swordfishes, and tunas), life span, and adaptations is spectacular. Most species are marine (about 60 %), with the remainder primarily living in freshwater and about 1 % moving between salt- and freshwater in their life cycle. Fishes have colonized almost all aquatic environments and evolve special adaptations to extreme habitats such as deep sea, polar regions, strong currents, caves, and seasonal water bodies in arid regions. Some species are warm-blooded, while other species living in cold environments have antifreeze peptides in the blood. Fish physiological adaptations to physical parameters such as pressure, temperature, alkalinity and salinity, light, high-energy water zones, etc., have been extensively studied, but the impact of these adaptations on immunity remains poorly known. It is certainly significant, however, as such adaptations lead to changes at the anatomical level (e.g., deep sea fish have lost the swim bladder) as well as at cellular level and blood composition or even at molecular scale (with adaptations of proteins including enzymes to different temperature and pressure ranges). Importantly, adaptation to multiple environments brought fishes in contact to diverse types of pathogen exposure, which likely represents the most important selection pressure on the defense system.

Fishes share the basic components of their immune system with all other jaw vertebrates (Gnathostomes) (Flajnik and Du Pasquier 2013, Table 1), of which the oldest fossils have been found in Ordovician sediments. The jaw acquisition likely has been pivotal for the later evolution of vertebrates, as it made possible diverse adaptations to a great number of ecological niches and food resources (Romer 1962). This shift from the microphage diet of agnathans should have modified significantly the interactions of the fish ancestors with their pathogens, as well as with the commensal bacterial flora in their gut (Matsunaga and Rahman 1998). It is precisely at this step of vertebrate evolution—in early Gnathostomes—that a new adaptive immunity emerged, in contrast to the VLR-based specific antigen

Table 1 Characteristics of the adaptive immune system in teleosts and mammals

	Teleosts	Mammals
Primary organs	Thymus/head kidney	Thymus/bone marrow
Secondary organs	Spleen	Spleen/lymph nodes
Mucosa-associated lymphoid tissues	+ (no Peyer's Patch)	+
Germinal centers/FDCs	–[a]	+
B cells		
Immunoglobulins		
Heavy chain (locus configuration)	IgD, IgM, IgT/Z[b] (translocon)[c]	IgD, IgM, IgA, IgG, IgE (translocon)
Light chain (locus configuration)	Kappa, lambda, sigma, sigma-cart (multi-cluster)	Kappa/lambda (translocon/multi-cluster)
Somatic hypermutation	+	+
Affinity maturation	Low efficiency	High efficiency
Class switch	–	+
T cells		
CD4/CD8 subsets	+[d]	+
Th1/Th2/Th17	+	+
TCR αβ/γδ	+	+

FDCs follicular dendritic cell
[a]Putative primordial germinal centers have been characterized
[b]In medaka and catfish the IgT/Z has not been identified
[c]More than one functional IgH locus in most of studied teleost species
[d]Absence of the *cd4* and *mhc II* functional genes in some species (e.g. cod and pipefish)

recognition found in Agnathans (Herrin and Cooper 2010). The novel adaptive immune system was based on antigen receptors made of Ig domains and diversified by genomic rearrangements mediated by RAG in specialized cells, the lymphocytes. As for VLR, the expression of a unique receptor per clone allowed clonal somatic selection of lymphocytes by their cognate antigen. Fishes and other jaw vertebrates also inherited from these early ancestors a common array of innate immune pathways and receptors, which were later amplified, reduced, or lost in the different lineages.

Bony fishes appear in late Ordovician (≈440 My) with Acanthodians, but the oldest fossils of teleosts were found only in Triassic deposits (≈200 My). The major radiation of teleosts occurred during Cretaceous, leading to the main groups of modern ray-finned fishes. The number of species and the diversity of adaptations make it one of the great successes of vertebrates. While fish inherited the basic components of the jaw vertebrate immune system, an open question is how such a spectacular expansion did affect immunity and somatic selection of lymphocyte populations. In fact, the immune system has been extensively studied in only a few key fish species: mainly aquaculture fishes, like carp, catfish, trout, and salmon, and among the model species, essentially the zebrafish. Importantly, complete genome

sequences are now available for many species belonging to a number of fish families, revealing several particular features—including traits important for immunity—of the evolution of the group. As shown in Fig. 1, species in which a complete genome sequence is available cover the main fish lineages. These genomes still represent a minute fraction of the whole fish diversity, but their analysis clearly showed that a whole genome duplication occurred during the early evolution of ray-finned fishes. Pairs of duplicated genes from this early event provided a large resource for subfunctionalization and likely favored the diversity of defense mechanisms among other adaptations. Fish genome diversity was further increased by lineage-specific events of genome duplication and/or contraction (Fig. 1). Specific expansions of gene families were also frequent. A striking example in the immune system is the fish-specific *trim* family (*fintrims*) (van der Aa et al. 2009); this multigenic family shows various degrees of expansion across fishes and diversified through distinct mechanisms. For example, local duplications occurred in zebrafish, leading to multiple clusters, while retropositions were the key mechanisms in the medaka, leading to many intronless genes. With respect to the adaptive immune system, duplicated loci may create significant complications of the finely tuned clonal selection pathways. For example, the expression of multiple major histocompatibility complex (MHC) molecules in an Ag-presenting cell might perturbate lymphocyte selection or lead to low density of Ag-presenting molecules. In fact, only one or a few copies of MHC gene duplicates are generally conserved. The fish MHC is unique among vertebrates since genes encoding MHC class I (MHC I) and MHC class II (MHC II) molecules are not encoded in the same genomic region but in two distant loci, which may each correspond to old duplicated regions (Kelley et al. 2005). Another important question is the maintenance of allelic exclusion with multiple loci of Ig and/or TCR loci. Efficient mechanisms certainly have been developed for this, as several fish species have maintained multiple Ig loci; however, many duplicated Ig segments have become pseudogenes. Finally, the presence of multiple families of transposable elements in fish genomes likely has been an additional driver for the evolution of fish genomes and multigenic families.

3 Selected Distinctive Features of Fish Immune Repertoires

3.1 How Allelic Exclusion Can Be Achieved with Multiple, Complex, and Large Loci Found in Some Fish Species?

The adaptive immune system can recognize millions of different antigens/antigenic determinants in a highly specific way. This fundamental property was modeled by Jerne (1955) and Burnet (1957) in a Darwinian perspective: the world of antigens was matched to a large population of lymphocytes through a specific recognition of

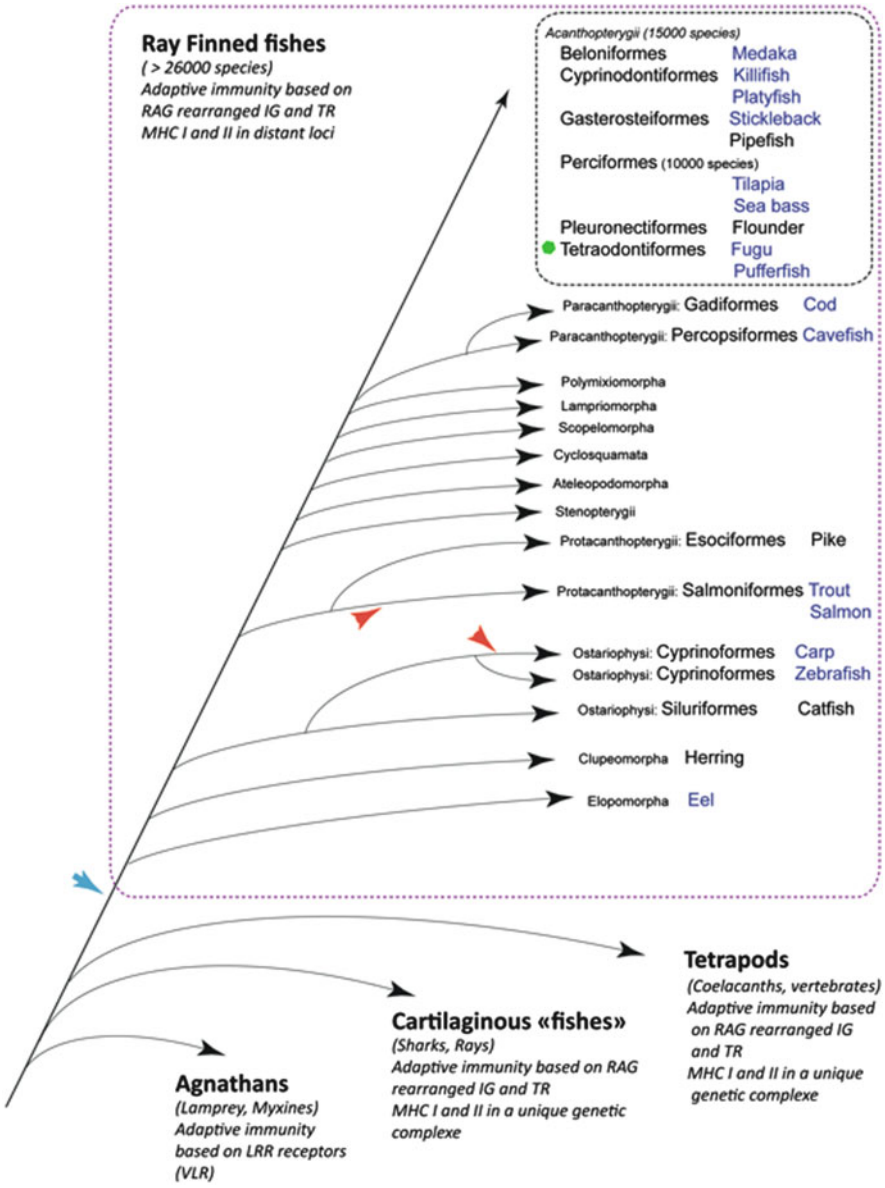


Fig. 1 Schematic phylogenetic tree of fishes. Taxonomy as in Helfman et al. (2010). The *blue arrow* indicates the whole genome duplication (WGD) which occurred during the early evolution of ray-finned fishes and led to a tetraploidization followed by a rediploidization. *Red arrows* indicate examples of posterior WGD in specific branches. The *green dot* denotes an example of genome contraction in tetraodontiformes. A genome sequence is available for all species in *blue*

an epitope by a given receptor ***unique*** to each lymphocyte clone. The genetic basis of this unique specificity of the Ag receptor expressed by each lymphocyte was not obvious even after the discovery of Ig gene rearrangements explained the generation of the molecular diversity of antibodies (Abs). The mechanisms ensuring an **allelic exclusion** to allow the expression of a unique receptor at the surface of B and T cells were progressively discovered mainly in the mouse.

In humans and mice, the IgH or TCR beta locus is activated in proB (T) cells and a D–J recombination occurs. This is followed by a locus contraction allowing the V to DJ rearrangement, which only occurs in one allele. Finally, the VDJ rearrangement is expressed on the surface of immature B or T cells as a pre-B-cell receptor (pre-BCR) or pre-T-cell receptor (pre-TCR) that allows to test its "functionality." If the expression of the pre-receptor leads to the right signal, loci of IgH (TCR beta) are closed and rearrangements of the light chain (TCR alpha) locus start. If the first VDJ joining is nonproductive or encodes a non-pairing chain, the second IgH/TCR beta allele is rearranged. Although the exact mechanisms are not fully understood, the current models suggest the presence of controls at multiple levels including the accessibility of the RAG recombinase to the different alleles within chromatin, the locus topology, and repositioning of loci to pericentromeric heterochromatin (Oltz 2001; Jhunjhunwala et al. 2009). Importantly, the allelic exclusion at the genomic level is not absolute especially for the second rearrangement (i.e., for IgL or TCR alpha/gamma). Regulation at the expression level is therefore also important to ensure that a unique receptor is mainly expressed by a given lymphocyte.

The enzymatic complex RAG is well conserved in jaw vertebrates. However, the accessibility of the antigen receptor genes to RAG, as well as the modalities of the VDJ rearrangements, has evolved with the organization of the Ig and TCR loci, which differ significantly between fishes and tetrapods (Hirano et al. 2011; Das et al. 2012). Studies in different vertebrate species including rabbit, chicken, and shark indicate that the VDJ rearrangement does not seem to always follow an ordered model (Hsu 2009). In chicken, for example, there is no sequential rearrangement of IgH and IgL genes; however, most of chicken B cells present only one functionally rearranged Ig allele, the other remaining in DJ configuration (Weill et al. 2002; Ratcliffe 2006). It was suggested that the V rearrangement happens in one allele randomly and that this is an event of such low efficiency that the probability to occur in both alleles is very low (stochastic/asynchronous model) (Weill et al. 2002). In such a model of simultaneous rearrangement of IgH and IgL genes, there would likely be no pre-BCR.

In non-tetrapod vertebrates, this question has been studied mainly in cartilaginous fishes as the presence of multiple VDJ recombination units in the genome represents an obvious issue for allelic exclusion (Hsu et al. 2006). In Chondrichthyans, the IgH locus is arranged in multiple independent clusters consisting in a few gene segments (V–D–D–J–C) and representing recombination units, as there is no evidence for intercluster recombination. These clusters can even be situated on different chromosomes (Rast and Litman 1998). In this context, a locus contraction would not have the same regulatory potential as in a translocon

configuration. In fact, the close proximity of the gene segments makes them recombine all at once and to completion (Rast et al. 1998). How different recombination units mutually exclude each other remains unknown.

Eason et al. (2004) identified different productive IgH gene transcripts in isolated single peripheral blood lymphocytes from the clearnose skate (*Raja eglanteria*), suggesting a simultaneous expression from multiple IgH rearrangements, potentially from different loci. However, most of the IgH loci are not characterized in this species, and some of them have germline-joined VHDJH, which makes these results difficult to evaluate. In the nurse shark (*Ginglymostoma cirratum*), there are no germline-joined VHDJH and the multiple IgH genes can rearrange autonomously (Zhu et al. 2011). Only one or a few IgH genes were found to be completely rearranged in each B cell, leading to only one functional transcript; the other loci remain in germline configuration. Interestingly, most IgH loci were partly rearranged in thymocytes, suggesting that a differential permissive state of chromatin in this region controls the recombination events in B-cell versus T-cell precursors (Malecek et al. 2008).

In fish, mechanisms of allelic exclusion remain largely unknown. The structure of TCR and Ig loci resembles the one found in mammals, rather than the one of Chondrichthyans: TCR beta and alpha/delta as well as IgH loci are organized as translocons while IgL and TCR gamma are made of multiple clusters. An important feature of fish Ig loci is the very large number of VH gene segments found in many species, especially in salmonids (Yasuike et al. 2010). Additionally, as bony fish underwent one or several additional genome-wide duplications (Petit et al. 2004) compared to most tetrapod species, the number of loci has increased. Thus, *Salmo salar* (Yasuike et al. 2010), *Oncorhynchus mykiss* (Hansen et al. 2005), *Oryzias latipes* (Magadán-Mompó et al. 2011), and *Gasterosteus aculeatus* (Bao et al. 2010; Gambón-Deza et al. 2010) present more than one functional IgH locus per haplotype. Thus, as Chondrichthyans, polyploid *Xenopus*, and genetically manipulated models with multiple IgH loci such as mice triallelic for IgH, the clonality of the system in bony fish requires not only allelic exclusion but also locus exclusion (Du Pasquier and Hsu 1983; Barreto et al. 2001).

Regarding IgL loci, the genomic organization of IgL was found to harbor clusters of V segments in opposite transcriptional polarity to J and C segments in trout and zebrafish. This configuration implies that primary gene rearrangements would be generated by inversion, which maximizes secondary rearrangements at IgL loci and serves as an important mechanism for receptor editing (Hsu and Criscitiello 2006). The impact of this structure on allelic/locus exclusion is not understood.

Importantly, the presence of a pre-BCR in fish has not been confirmed. In human and in the mouse, the pre-BCR consists of a homodimer of μ heavy chains associated with surrogate light chains (VpreB, homologous to a Vλ, and lambda 5, homologous to a Jλ-Cλ (Mårtensson et al. 2007)) and with the transmembrane signal molecules (Igα and Igβ). To date, homologues of the VpreB and λ5 genes have not been found in teleost genomes. In several fish species including zebrafish (Haire et al. 2000), Atlantic cod (Daggfeldt et al. 1993; Hsu and Criscitiello 2006),

and medaka (Magadán-Mompó et al. 2013), IgL transcripts without a V segment have been reported. However, these JC_L transcripts may produce functional surrogate JCκ proteins, as reported in humans for germline Vκ and JCκ transcripts encoding proteins that can functionally replace VpreB and λ5 (Francés et al. 1994; Rangel et al. 2005). Alternatively, it would be interesting to study if any of the many teleost light chain subtypes/isotypes could play the role of a surrogate light chain. Regarding the pre-TCR, of which the surrogate chain pTalpha that associates to TR beta has been described only recently in birds (Smelty et al. 2010), nothing is known in fish.

Altogether, an allelic (and interlocus) exclusion seems to be required in fish as in other vertebrates to ensure the clonal responses that have been observed, but the mechanisms remain to be discovered. In fact, a better knowledge of the early stages of differentiation of B and T lymphocytes, especially good molecular markers, would be critical to uncover such mechanisms. It is not even excluded to date that genic conversion might play a role in the diversity generation of the primary Ig repertoire in fish. Another important aspect is the diversity of fishes, as multiple solutions might have been selected in different fish groups to ensure lymphocyte clonality.

3.2 What Does Clonal Complexity of Responses to Pathogens Reveal About Roles of Fish Ig Isotypes and the Structure of T-/B-Cell Responses?

In fish, particularities of immune responses to Ag—for example, in Atlantic cod—might suggest they do not always follow the paradigm of the clonal selection theory. However, a number of features shared between the immune system of fish and mammals (or other tetrapods) advocate analogous sets of lymphocytes and response mechanisms. For example, the typical counterparts of coreceptors (CD4, CD8, CD28, etc.), as well as Th1 and Th17 cytokines, suggest the presence of the similar types of T-cell subsets in fish and mammals. With respect to fish B lymphocytes, cells expressing, respectively, IgT or IgM are specialized in mucosal versus systemic immunity, reminding the roles of IgA and IgM/IgG in mammals, respectively. The analysis of the immune complexity of fish naive repertoires, as well as of responses to pathogens, has just started to decipher to what extent fish and mammalian adaptive immunity may be similar.

Spectratyping of TCR beta CDR3 after a systemic viral infection revealed a strong polyclonal secondary response in the spleen, implicating most of the rainbow trout TCR beta V families (Boudinot et al. 2001, 2002). As observed in humans and in mice, there was a public response consisting of a number of slightly varying CDR3 sequences differing by conservative amino acid substitutions. This was observed in all tested individuals of a trout clone sharing the same genetic background. Also, the same CDR3 protein sequence could be encoded by different

nucleotide sequences produced by distinct rearrangements, which was a good evidence of Ag-driven clonal selection. These observations showed that salmonids have an available TCR beta repertoire diverse enough to allow public responses, besides a large diversity of private responses that were observed only in some individuals. After DNA vaccination by intramuscular injection of an expression plasmid for a viral G protein, the complexity of the response dropped, and only the public component could be detected by CDR3 spectratyping (Boudinot et al. 2004). This response was apparently specific for the Ag, as it was never observed after injection of an expression plasmid for another viral protein, the nucleocapsid. These analyses are not highly sensitive as they are performed in the absence of in vitro restimulation because the necessary reagents are not available in the trout model. However, they indicate that DNA vaccination elicits a TCR beta response mainly restricted to an epitope located on the target of the neutralizing Ab response (the viral G protein), which is absolutely essential for the protection in this disease model. This observation evokes the idea of a T-/B-cell cooperation for the anti-G response, although it does not provide a direct evidence for it. With the recent production of MoAbs against CD8, CD4, and potentially other T-cell markers, it would be important to determine if this public response is mediated either by $CD8^+$, presumably cytotoxic, T cells (which have been detected close to the site of DNA vaccine injection (Utke et al. 2008) or by $CD4^+$ helper T cells. Furthermore, it would be interesting to investigate if responses similar to the public component observed in double haploid trout would also be found in genetically mixed farmed or wild trout populations, as suggested by preliminary results (Boudinot et al. 2004).

Such CDR3 spectratyping using V-, C-, and J-specific primers can be used in an "immunoscope" strategy to identify and target particular strong components of the adaptive response or to provide an overview of the degree of oligoclonality of a repertoire. However, it does not allow to describe the whole clonal composition and complexity of a lymphocyte population. Such comprehensive analyses have been made possible by the development of deep sequencing technologies, which provide a complete or quasi complete description of sequence distribution within a V/C or VJ PCR product (Six et al. 2013). These technologies have been used to analyze the B-cell repertoire in two fish species: zebrafish and rainbow trout.

A first seminal study used high-throughput sequencing to perform a global analysis of the μ IgH repertoire in 14 adult zebrafish (Weinstein et al. 2009). This work showed that a large part of the potential combinations of V, D, and J genes are indeed used in the available repertoire, with VDJ frequency patterns often common to different individuals, as observed in mouse or in human. Different methods of the number of different μ IgH chains estimated that 1200–3500 are expressed per individual zebrafish. This large dataset was subjected to several in-depth analyses that revealed additional features of the zebrafish IgH repertoire. Using methods of statistical physics, Mora et al. investigated pairwise correlations between residues in sequences comprising the end of the VH gene segment and the CDR3 region (Mora et al. 2010). Their analysis confirms that each individual fish repertoire covers only a small part of the potential repertoire and that the VDJ combinations

expressed are correlated between fishes. In contrast, each fish has its own CDR3 diversity and 13 individuals are not enough to express the whole potential diversity of this region. Overall, these analyses conclude that (1) the antibody diversity is not directly limited by the number of VDJ sequences present in the genome and that (2) about half of the repertoire diversity of μ IgH CDR3 is unique to each individuals, while the other half is shared by many (or all) fishes. Importantly, this represent a large potential for public responses, still keeping each fish able to mount individual specific components of responses to pathogens. Importantly, these proportions (50 % shared/50 % individual) might be different in large species or in fishes exposed to a vast diversity of pathogens in their ecological niche: intuitively, one would expect that the "minimal shared component" selected by the pathogenic/antigenic landscape would represent a smaller proportion of the repertoire of bigger species having much higher numbers of B cells. However, this view is likely oversimplistic, and experimental studies will be necessary to understand the repertoire dynamics and statistical properties in diverse fish species. Interestingly, an independent network analysis of the same dataset identified two distinct groups of fishes: one group with a uniform use of VJ combinations and network of clonotypes and another group with some VJ combinations present at much higher frequencies and highly connected to each other. While these fishes had not been intentionally immunized, the clonal structure of the second group was interpreted as a set of clonotypes participating to an adaptive (pathogen-driven?) response (Ben-Hamo and Efroni 2011).

When studying the μ IgH repertoire of a 2-week-old zebrafish, a high stereotypy, i.e., a preferential use of a small number of VDJ combinations among the potential diversity, was observed (Jiang et al. 2011). Such similar primary μ IgH repertoires expressed in different individuals—which remind observations published in mice at the level of expression of VH families (Huetz et al. 1993)—had apparently disappeared by 1 month in young fishes: many dominant VDJ combinations are not observed anymore at this stage. To differentiate (1) the rate of rearrangement for a given V(D)J combination and (2) its promotion by clonal selection during the formation of the repertoire, the authors distinguished the total number of reads and the number of lineages (identified by the junction sequence) for a given V(D)J. Interestingly, while the number of reads per VDJ was highly correlated between different fishes only at the 2-week time point, the "VDJ lineage diversity" (i.e., the diversity of clonotypes per V(D)J combination) appeared highly stereotyped in older animals and even highly correlated between fishes of different ages.

In rainbow trout, a similar technology was used to investigate the clonal structure of the B-cell response against a fish rhabdovirus, the viral hemorrhagic septicemia virus (VHSV). The response was analyzed in spleen by a combined approach of CDR3 spectratyping and deep sequencing of the IgH transcripts (Castro et al. 2013). Double haploid fish were vaccinated using an attenuated strain of VHSV, and challenged 3 weeks later. In naive animals, bell-shaped CDR3 spectratypes suggested a polyclonal repertoire for the three isotypes, while deep sequencing revealed the presence of a few IgM- and IgT-secreting cells in the spleen. The analysis of the modifications of IgM, IgD, and IgT repertoires after the

viral challenge and secondary response revealed complex and highly diverse IgM responses involving all VH subgroups and dominated by a few large public and private clones. No public IgT response was identified, but a fair number of amplified clonotypes were detected, showing that this Ig class specialized in mucosal protection also participates to some degree to the response against a systemic viral infection in the spleen. In contrast, the IgD response to the infection appeared to be negligible in this context.

The clonal complexity of this response and its public IgM component were consistent with the general rules of B-cell response and clonal selection known in mammals, suggesting that general properties were already in place in the common ancestors of fish and tetrapods and were conserved in both lineages. However, the implication of all VH subgroups is intriguing and might be explained partly by polyclonal bystander activation during the secondary responses to an acute infection. It will be important to determine how the primary response is structured and to what extent it is stable over the following months in order to better understand the mechanisms of B-cell memory.

Remarkably, fish have been among the first models in which global repertoire studies were conducted using deep sequencing, and their repertoires are currently among the best described of all vertebrates. With new sequencing techniques coming such as IgL/IgH-coupled sequencing and molecular bar-coded reads, we anticipate that rejuvenated repertoire studies will help in understanding the structure of responses, bases of protection, and mechanisms of memory both in fish and Vertebrates in general.

3.3 Independent Loss of MHC Class II Molecules Occurred in Two Different Evolutionary Branches of Fish; Is T Cell Help Dispensable for Good Ig Response to Pathogens?

In men and mice, B-cell responses to T-dependent (TD) antigens require $CD4^+$ helper T cells to cooperate with B cells expressing an Ag-specific Ig and presenting the antigen. This cooperation leads to the synthesis of Ag-specific Abs and finally to the induction of immune memory. In contrast, T-independent (TI) antigens, which often have repeated structure such as viral capsids, can trigger production of IgM in the absence of T cells.

It has been known for a long time that responses to murine TD and TI antigens required cooperation between different cell types in fish. Using an in vitro PFC test and catfish leukocyte culture models, Miller and Clem (Miller and Clem 1984; Miller et al. 1985) studied hapten-carrier effect and primary responses to both TI (TNP-LPS adsorbed on bentonite particles) and TD (DNP-BSA; TNP_{10}-KLH; TNP_{10}-HorseSA). They could demonstrate that mixing lymphocytes expressing Ig (surf-Ig^+) and macrophages allowed responses to a TI antigen, while responses

to TD antigens required the presence of surf-Ig$^+$ lymphocytes, macrophages, and surf-Ig$^-$ lymphocytes (comprising helper T cells). This remarkable work was the first evidence that fish B cells cooperate with T cells to build responses to TD antigens, as in mammals. Many other indications confirmed this view in other fish species, such as the reduction of the secondary Ab response to human gamma globulin in young trout when fry had been thymectomized at 1 month post hatch (Tatner 1986).

Furthermore, orthologues of CD4 and CD8 coreceptors have been found in fish, and they are specific for T-cell subsets that generally have similar functions to their mammalian counterparts: while the CD8$^+$ population contains cytotoxic cells (Somamoto et al. 2009; Takizawa et al. 2011; Nakanishi et al. 2011), the CD4/CD4rel$^+$ populations appear to be able to help B cells (Toda et al. 2011; Somamoto et al. 2014). Typical polymorphic *mhc I* and *II* and the key genes for antigen processing and presentation molecules also have been found in many fish species; the structure and expression pattern of *mhc* genes suggest that they present the antigen to CD8$^+$ and CD4$^+$ T cells, respectively, as in other vertebrates. Collectively, these observations suggest that the antigen processing, presentation, and B- and T-cell antigen recognition are conserved in jaw vertebrates including fish, setting up a general mode of lymphocyte selection and cooperation during responses.

However, the antibody response of Atlantic cod (*Gadus morhua*, Gadidae) did not follow this paradigm. Despite extensive efforts to immunize cod using different antigens and protocols, it remained for a long time impossible to induce a typical specific antibody response (reviewed in Pilstrom et al. (2005)). However, the level of serum "natural" antibodies was very high in non-immunized cod when compared to other fish species (Israelsson et al. 1991; Pilström and Petersson 1991; Magnadottir et al. 1999). The search for a genetic explanation in cod Ig sequences and loci was unsuccessful, leading to the idea that a deficiency in MHC class II likely was the reason for the lack of specific antibody response in this species (Pilstrom et al. 2005). This hypothesis was confirmed when the complete genome sequence of cod was deciphered (Star et al. 2011): the *mhc II* genes were absent, as well as the invariant chain (Ii21) which is involved in the assembly of the MHC II/peptide complex. Also, the gene encoding the coreceptor CD4 was represented only by a truncated pseudogene, indicating that the Atlantic cod had lost the classical pathway of antigen presentation by *mhc II* to helper T cells.

Lack of *mhc II* sequence also has been observed in the transcriptome of another gadoid, the burbot (*Lota lota*) (Star et al. 2011), suggesting that the deficiency is probably shared within the gadoid lineage. Strikingly, the absence of *mhc II* and *cd4* genes and a nonfunctional *li21* was reported in a phylogenetically distant species, the pipefish *Syngnathus typhle* (Syngnathidae) (Table 2). This observation indicates that the loss of the key genes of the MHC II pathway occurred at least twice independently during the evolution of teleosts.

These particular situations in cod and pipefish raise several fundamental questions about somatic selection of lymphocytes and the importance of cooperation between T and B cells in the fish defense system:

Table 2 Absence of functional genes from the MHCII pathway in cod and pipefish

Gene	Cod	Pipefish
MHCI	+	+
MHCIIa	–	–
MHCIIb	–	–
Invariant chain	–	Nonfunctional
CD3e	+	+
CIITA	+	–
RFX5	+	+
RFX7	+	+
CD4	Truncated	–
CD8a	+	+
CD8b	+	–

1. *How mhc II-deficient immune systems, which would lead to a strong decrease of resistance to pathogens in mammals, could be adopted by very successful species as Atlantic cod?*
2. *Can such species mount efficient Ig responses without MHC II pathway?*
3. *Do they have compensatory mechanisms that make their defense system well adapted to their ecological niche?*

Several evolutionary pathways explaining the absence of mhc II genes in cod have been discussed by Star and Jentoft (Star and Jentoft 2012). A first scenario involves genetic drift and hypothesizes that the MHC II pathway is not crucial for defense against parasites. Other scenarios are based on directional selection: for example, the MHC II pathway could have metabolic costs that would not be compensated by its contribution to the protection of the individual against pathogens or other defense mechanisms could have compensated for the deficiency. Such scenarios may be conditioned by the environmental conditions, and it has been proposed that the cold marine environment may have favored the *mhc II* loss in the cod lineage. However, the *mhc II* deficiency of the pipefish and the conservation of *mhc II* in a large number of fishes living in cold sea do not strongly support this hypothesis.

Following Star et al. (2011), it is tempting to link the loss of the *mhc II* pathway in cod to the particularities of its repertoire of immune genes, especially to the high number of *mhc I* genes (Persson et al. 1999). Interestingly, the amphibian axolotl has a notoriously poor humoral response that was associated with the limited presentation capacity of its non-polymorphic MHC II molecules (Tournefier et al. 1998); as the Axolotl possesses multiple polymorphic *mhc I* genes (Sammut et al. 1999), *mhc I*-based compensatory mechanisms may represent an interesting case of convergence with cod. In fact, a detailed analysis of the large repertoire of cod *mhc I* sequences (Malmstrøm et al. 2013) identified a particular gene subset with a novel combination of two endosomal sorting motifs in the cytoplasmic tail: a tyrosine-based motif associated with exogenous peptide presentation by cross-presenting MHC I molecules and a dileucine-based motif associated with normal

MHC II functionality in mammals. Although these observations evoke MHC II function and a specific adaptation to antigen cross-presentation by MHC I molecules, it does not explain how this mechanisms could actually compensate for the loss of the MHC II pathway. Would this cross-presentation be restricted to cytotoxic T cells (providing Th1-like responses)? Or might specialized MHC I molecules expressed on B cells cross-present extracellular pathogen-derived peptides to $CD8^+$ T cells with TH2-like functions? While initial immunization trials did not elicit specific Ab production in cod, it was recently reported that the administration of emulsions of bacterin antigen in mineral oil leads to robust and specific antibody responses mainly targeting the LPS and A-layer components (Schrøder et al. 2009; Arnesen et al. 2010; Mikkelsen et al. 2011). The mechanism by which such immunizations succeed in eliciting good Ab responses may rely on Ag stabilization in oil deposits or on the induction of local inflammatory and co-stimulatory signals. Hence, in the absence of MHC II presentation and T-cell help, Ab responses to pathogen TI epitopes could be pivotal for the fish defense and may account for the protection afforded by oil-based vaccines against *Vibrios* in challenges performed 7 weeks post vaccination. In fact, Ab *secondary* responses to TD, *but also to TI antigens*, are significantly prolonged as compared to the primary responses in rainbow trout (Ma et al. 2013); such particular features of fish B-cell responses could explain increased protection long after vaccination *even if Ab responses are fully thymus independent*.

Thus, when a fish species is able to mount efficient anamnestic responses to TI antigens, the loss of TD response due to genomic inactivation of a key gene of the MHC II pathway may not affect significantly the fitness, at least in some ecological niches where pathogens can be controlled by TI responses. Of note, it seems that the fish MHC II pathway is less sophisticated than the one of mammals—as it lacks, for example, the DM chains (Dijkstra et al. 2013)—hence it could be less critical for the overall resistance to pathogens. The remarkable plasticity of fish genomes and the complexity of their chromosome evolution across a vast adaptive radiation certainly increase the probability of independent events of *mhc II* loss in several fish lineages. Furthermore, the location of *mhc I and II* genes in distinct regions of the genome, a feature unique to fish, may have facilitated the deletion of *mhc II* genes without impact on the *mhc I* region. The other elements of the MHC II pathway could have been lost by genetic drift after the first inactivating event, in the absence of "protective" selection pressure. It is difficult to decide if compensatory mechanisms (e.g., via MHC I-based cross-presentation) represent adaptations posterior to the initial event or a preexisting context favorable to the loss of the MHC II pathway. In this respect, it will be interesting to characterize the diversity of the *mhc I* genes in the pipefish. Another question is what factor(s) led to the success of individuals deficient in *mhc II* within their own species or population: lucky neutral evolution or selection of advantageous genomic event(s) parallel to the *mhc II* inactivation? Many other pathways remain possible.

3.4 *Are Fish Ig Sequences Diversified in the Absence of Germinal Centers? New Insights from Deep Sequencing Datasets*

The mammalian immune system can respond specifically to antigens and allows Ag-specific memory after a first encounter: the secondary response is faster and based on antibodies with higher affinity compared to the primary response. In mice and humans, this strong increase of affinity is based on the ability of immune system to modify the immunoglobulin diversity engaged in the response through mutation of Ig genes in responding B cells within germinal centers (GC), followed by clonal selection of B cells expressing high-affinity BCRs (Shlomchik and Weisel 2012). In a typical primary TD immune response, Ag-stimulated B cells rapidly move to the T-/B-cell interface of lymphoid tissues where they interact with $CD4^+$ helper T cells and cytokines. T cells then induce B-cell proliferation and differentiation into short-lived antibody-producing plasma cells, or B-cell translocation to the germinal centers where the variable region of rearranged Ig genes may undergo somatic hypermutation. B cells finally differentiate into long-lived plasma cells or memory B cells (Zielinski et al. 2011).

A number of studies showed that in fishes as in mammals, the B-cell repertoire diversity is not only due to V(D)J recombination in adult individuals but also based on the hypermutation of the variable region. In fact, fishes express the enzyme activation-induced cytidine-deaminase (AID), which is responsible for Ig somatic hypermutation and class switch recombination in mammals and other tetrapods (Barreto et al. 2005). Although fishes lack conventional class switch (Stavnezer and Amemiya 2004), fish AID has similar biochemical functions and deaminates cytosine, thus inducing point mutations; it even mediates class switch in mouse B cells (Barreto et al. 2005). The mutational pattern observed in Ig sequences from catfish (Yang et al. 2006), zebrafish (Marianes et al. 2011), and also in nurse shark (Diaz et al. 1999) reminds actually the one described in mammals, with a preference for transitions but no major bias for mutation of G:C or A:T pairs (Diaz et al. 2001).

However, the capacity to perform somatic hypermutation in Ig genes does not imply that efficient antigen-driven selection and affinity maturation of immune response occur. It is generally accepted that the affinity maturation in fishes and other ectotherms is much less efficient than in mammals (Wilson et al. 1992; Yang et al. 2006), based on two different types of data: (1) measurements of the affinity of Ag-specific serum antibodies using analytical tools and (2) computing of the dS/dN ratio of synonymous (S) and non-synonymous (N) mutations in variable regions of immunoglobulin genes.

The classical analytical tools, such as equilibrium dialysis or fluorescence quenching, only provide an estimation of the average affinity for all Ag-specific antibodies within a serum sample, which is moderately sensitive. The use of new analytical techniques, as solid-phase ELISA (Shapiro et al. 1996), has allowed the fractionation of an Ab response into affinity subpopulations (Kaattari et al. 2002) and more sensitive analyses. Such experiments revealed a consistent increase of

affinity for the Ag during the antibody immune response in teleosts (Ye et al. 2011): after immunization of rainbow trout with the TNP-KLH in Freund's complete adjuvant, a significant proportion of the lower affinity subpopulations appeared during the first 5 weeks and decline after a few months, giving way to higher affinity subpopulations with an increase of affinity reaching 100-fold, which is very low but reminds that observed in mammals in secondary responses.

A typical pattern of synonymous (S) and non-synonymous (N) mutations within Ig variable region is observed after affinity maturation, with low dS/dN ratio specifically in CDRs. In fish, very few works have addressed precisely this process. Somatic mutation was detected within VH regions of channel catfish Ig sequences. However, dS/dN was not significantly different in FR and CDR (Yang et al. 2006), suggesting that the process potentially induced a repertoire diversification that was not followed by efficient clonal selection of high-affinity B cells. In contrast, the distribution of mutations focused on CDRs in immunoglobulin kappa-like light chain in medaka suggested selection and affinity maturation; additionally, the dS/dN ratio was consistent with an antigen-driven selection in 5 out of 13 mutated sequences (Magadán-Mompó et al. 2013). In zebrafish, the expansion of VJC clonal lineages with an increased frequency of mutation might also suggest selection of B cells expressing hypermutated Ig (Marianes et al. 2011).

A comprehensive assessment of the hypermutation changes of Ig sequences has become theoretically feasible with massive parallel sequencing methods (Weinstein et al. 2009). As mentioned above, simultaneous sequencing of Ig transcripts from millions of B cells recently gave access to a whole-organism IgH repertoire of small organisms such as zebrafish (Jiang et al. 2011). To measure the effect of hypermutation and clonal selection during zebrafish development, junctional diversity and VH sequence mutations were compared as a function of sequence abundance among different age groups of fish. Importantly, it appeared that mutated sequences dominate *highly expressed VDJ combinations in older fish, but not in young animals*, which was consistent with an accumulation of B-cell clones expressing hypermutated IgH with age.

In these studies, fishes were naturally (and progressively) immunized (i.e., not intentionally), and further studies using targeted immunizations or vaccines will have to address directly the question of affinity maturation: isolation and Ig gene analysis of activated Ag-specific B cells will have to be done and the Ig affinity evaluated as in the work of Dooley et al. in nurse shark new antigen receptor (IgNAR) (Dooley et al. 2006). In nurse shark hyperimmunized with hen egg white lysozyme (HEL), HEL-binding clones expressed IgNAR with hypermutated V regions, and an increase in affinity was confirmed. Admittedly, the presence of light chains in teleostean immunoglobulins complicates such analyses. However, advances in high-throughput sequencing of VH:VL pairs (DeKosky et al. 2013; Tan et al. 2014) likely will allow a comprehensive analysis of VH:VL gene usage and somatic hypermutation in FACS-sorted Ag-specific B-cell populations. Such methods will certainly be more powerful than the current complex and expensive methods based on single sorting and RT-PCR (Tanaka et al. 2010).

It is still generally believed that the reason for the poor affinity maturation of Ab responses in ectothermic vertebrates is an inefficient selection of high-affinity clones. This is consistent with the absence of true GC in these organisms, although lymphotoxin-**β** KO mice succeed to perform good affinity maturation of Ab responses in the absence of GC (Matsumoto et al. 1996). The affinity maturation pathways have not been characterized in these mice, but other studies have identified hypermutation and/or selection outside of conventional microenvironments (Schröder et al. 1996). On the other hand, cell clusters were observed in the spleen and kidney of channel catfish and goldfish, which were put forward as putative primordial GC (Saunders et al. 2010; Barreto and Magor 2011). These cell clusters contain melanomacrophages that express CSFR1, B cells, and $CD4^+$ T cells, as well as AID^+ cells. In the model proposed, $CD4^+$ cells would play a similar role as their counterparts in mammals, providing survival and differentiation signals to B cells; melanomacrophages would act as follicular dendritic cells (FDCs) in the GC of mammals, trapping and presenting Ag or Ag-antibody complexes on their surface. Thus, Ag-specific B cells with hypermutated Ig genes would be selected on their capacity to bind the (limiting) antigen on melanomacrophages.

Another question is the contribution of long-lived plasma cells and true memory B cells to the ***maintenance*** of an effective humoral adaptive immunity in teleosts. Results obtained in several models indicate that serum Ag-specific antibodies are not detectable 1–2 years after exposition and that Ag-specific plasma cells are not any more visualized by ELISPOT in head kidney if no challenge is performed. This has been observed in channel catfish infected with the ciliate *Ichthyophthirius multifiliis* (Findly et al. 2013) and in rainbow trout immunized with an expression plasmid for the G of IHNV (Kurath et al. 2006) or with *Streptococcus iniae* bacterin (Costa et al. 2012). Taken together, these data may suggest that IgM^+ memory B cells, not long-lived plasma cells, likely are responsible for the long-term protective immunity in teleosts. In fact, a population of true memory B cell remains to be identified and characterized in fish.

Memory and long-term maintenance of Ag-specific B-cell immunity in species where *mhc II* and/or *cd4* genes have been lost, such as Atlantic cod or pipefish, is another interesting issue. Cod-specific Ab responses seem to be mainly directed to prototypic TI antigen as LPS and A-layer components (Espelid et al. 1991; Gudmundsdóttir et al. 2009). In mice and humans, such antigens that bypass T-cell help indeed induce specific memory B cells, whose secondary activation is inhibited by the presence of Ag-specific antibodies (Brodeur and Wortis 1980; Obukhanych and Nussenzweig 2006). The quality and life span of these cellular subsets likely depend on the strength of BCR-derived signals and complementary effects of toll-like receptor (TLR) stimulation and inflammatory cytokines (Alugupalli et al. 2007; Defrance et al. 2011). Their existence in cod and pipefish remains unproven.

A better understanding of mechanisms involved in induction and maintenance of long-lived and memory B cells in teleosts will be critical to develop efficient vaccines affording a durable protective immunity, especially against pathogens that evolve subversion strategies. In that sense, there is still a very long way to go.

3.5 What Are the Particular Features of the Mucosal Repertoires in Fish?

Fishes have the most extensive and complex mucosal surfaces of interaction with environment among vertebrates: not only gut and respiratory mucosa (gills) but also all the skin surface is a bona fide mucosa. Even in the best studied models, the immunity of these tissues remains poorly understood, although significant advances have recently been made in the understanding of mucosal B-cell responses in teleosts.

3.5.1 Fishes Possess a Mucosal Specialized Ig Isotype, But IgM (and IgD?) May Also Participate in the Protection of Mucosa

In 2005 a new teleost fish immunoglobulin (Ig) isotype was identified. In rainbow trout this Ig was named IgT (Hansen et al. 2005), whereas in zebrafish it was called IgZ (Danilova et al. 2005). This Ig has been identified in all analyzed teleost fish, except in medaka and in channel catfish. However, the complete genome of catfish is not available thus far, and it remains possible that the catfish orthologue for IgT has not yet been found. Rainbow trout IgT and IgM use the same VH segments, which rearrange to either to Dτ or Dμ segments. Analogous to the structure of the TCRα/δ locus, Igτ Dτ, Jτ, Cτ genes are located downstream of the VH and upstream of the IgM Dμ, Jμ, Cμ (Hansen et al. 2005), suggesting that VH–DμJμ precludes further VHDτJτ rearrangement. Combined with the lack of obvious switch regions in the IgT locus, commitment to IgT or IgM appears to involve alternative V(D)J recombination rather than isotype switching.

It was not until 2010 that IgT was biochemically characterized and its function in mucosal immunity revealed (Zhang et al. 2010). It was shown that while plasma IgT is a monomeric Ig (~180 kDa), gut mucus IgT is polymeric (4–5 monomers). Significantly a previously unknown IgT^+ B-cell lineage was identified. This B-cell subset expresses surface IgT, but not IgM or IgD, thus constituting the first vertebrate B-cell lineage devoid of surface IgD expression. IgT^+ B cells represented the predominant B-cell subset in the gut. Moreover, plasma-like cells uniquely producing IgT could also be detected. The same study identified a polymeric Ig receptor in rainbow trout (tpIgR) whose putative secretory component (sC) was found associated with gut mucus but not serum IgT and IgM. This result strongly suggested that like in mammals, pIgR in fish is involved in the transport of polymeric IgT and IgM from the mucosal epithelium into the gut lumen. The IgT/IgM ratio was found much higher in the gut mucus when compared to that in serum, thus suggesting that IgT could play a role in gut mucosal immunity. Confirming this hypothesis it was shown that trout infected with Ceratomyxa shasta (a gut parasite) induced parasite-specific IgT responses only in the gut mucus, but not in serum. Conversely, the serum of surviving fish contained significant parasite-specific IgM, but not IgT titers. Moreover, fish that survived parasite infection

contained large accumulations of IgT^+ B cells in the gut, whereas the number of IgM^+ B cells did not change with respect to control fish. Overall, this model provided the first evidence in teleost fish or any other non-tetrapod species for a compartmentalization of immunoglobulin isotypes into mucosal (IgT) and systemic (IgM) areas in response to pathogenic challenge. Supporting further the idea that IgT is involved in mucosal homeostasis, it was found that IgT, similar to mammalian IgA, was the prevalent Ig found coating the gut microbiota. This finding represented the first example of coating of microbiota by a nonmammalian mucosal Ig and pointed to a conserved role of mucosal Igs in the control of the bacterial microbiota at mucosal surfaces. Moreover, this finding provided further evidence for the specialization of IgT in mucosal homeostasis. More recently, very similar responses were reported in the skin mucosa of rainbow trout (Xu et al. 2013). Similar to what was found in the gut mucus, IgT in the skin mucus was mainly detected in polymeric form. In addition, IgT^+ B cells also constituted the main B-cell subset of the trout skin-associated lymphoid tissue (SALT). Fish that survived infection with a skin pathogen (Ichthyophthirius multifiliis) showed large accumulations of IgT^+, but not IgM^+ B cells, in the SALT. Along with these substantial increases in IgT^+ B cells, large increases of IgT protein (~10-fold) mucus were detected, while IgM levels remained unchanged. Moreover, survivor fish had significant titers of parasite-specific IgM in the serum, but not in the mucus of most fish. Conversely, significant titers of parasite-specific IgT were detected in the skin mucus but not in serum. Supporting further the prevalent role of IgT in the SALT, it was found that like in the gut, a majority of SALT microbiota were coated with IgT. Essentially very similar IgT responses to those described in the trout GALT and SALT were also found in the trout gill-associated lymphoid tissue (Dr. Sunyer, personal communication). Moreover, a mucosal-associated lymphoid tissue has recently been described in the nose of rainbow trout (NALT) (Tacchi et al. 2014). In this tissue, IgT^+ B cells are the predominant B-cell subset, thus suggesting also that IgT plays a major role in NALT immunity. Overall, these findings support further the initial discovery in the gut, indicating that IgT is an immunoglobulin specialized in mucosal immunity. Thus, it seems reasonable to suggest that IgT responses in all main fish mucosal areas operate under the guidance of primordial common principles. While it is likely that IgT is the major responder in mucosal surfaces, IgM responses have also been detected in the gut and skin mucosa upon vaccination or infection (Salinas et al. 2011). Whether fish IgD plays a role in systemic or mucosal immunity remains unknown, although IgD responses were not detected in the gill of rainbow trout upon infection with *I. multifiliis* (Dr. Sunyer, personal communication), nor in the spleen of trout infected upon systemic viral infection (Castro et al. 2013).

Future repertoire studies of mucosal B-cell populations in healthy or infected fish will provide more insights about the clonal structure of B-cell response in these tissues. It is tempting to speculate that IgT^+ B cells, like IgA^+ B cells in mammals, could have a particular repertoire with a few very large clones amplified in mucosal tissues (Stoel et al. 2005). The compartmentalization of IgT^+ B-cell repertoires in

different mucosae of the individual will also be an important issue to clarify for a better understanding of mucosal fish defenses.

Indeed, the absence of switch in fish likely introduces important differences in the selection constraints exerted on mucosal B cells between fish and mammals: IgT^+ B cells constitute a separate lineage of B cells and express IgT from their early differentiation stages, while mammalian B cells are all first selected on the basis of a membrane-bound IgM. One might imagine that a fully independent differentiation pathway of mucosal B cells in fish might bring more freedom to select specialized resident B-cell subsets for each mucosal territory.

3.5.2 Fish Anatomical Structures Represent Particular Microenvironments for B and T Cells, which Constraint Available Repertoires and Initiation of Adaptive Responses

It is generally accepted from observations made in men and mice that the adaptive immune response is initiated in special microenvironments within secondary lymphoid organs. This view is challenged by the anatomical differences between fishes and mammals: while fishes are able to mount protective immune responses, they lack lymph nodes and germinal centers. In the model proposed by Kaattari and colleagues, B-cell differentiation occurs in the fish kidney (mainly in the anterior kidney), from which mature B lymphocytes migrate to the blood and immune tissues including spleen and likely mucosal territories. It is generally believed that upon encounter with the Ag, B cells then differentiate into plasmablasts, proliferate, and migrate to the kidney where they become (long-lived) plasma cells. It is not clear where the interactions between B cells and T cells occur. Interestingly, a number of observations realized in mammals support the notion that T cells are able to respond to the Ag outside of a lymph node environment (Hofmann et al. 2010); for example, LT$\alpha-/-$ mice are devoided of lymph nodes but have a vigorous cell-mediated immunity (De Togni et al. 1994). In fish, adaptive response and T-/B-cell cooperation likely take place in the major secondary peripheral lymphoid organ, the spleen; although not clearly regionalized, this organ contains ellipsoidal blood vessels with a layer of macrophages specialized in antigen trapping as macrophages of the marginal zone in mammals and melanomacrophage aggregate.

The mechanisms of the initiation of adaptive responses in fish mucosa remain even more elusive. Like all jaw vertebrates, fish possess lymphoid tissues associated to mucosa. Macrophages, B and T lymphocytes as well as granulocytes are found in the gut and constitute the GALT, which is not made of encapsulated structures as Peyer's patches (Rombout et al. 2011). Specialized structures are also present in gills, with aggregations of lymphocytes including a large proportion of T cells in the interbranchial lymphoid tissue (ILT) (Haugarvoll et al. 2008; Koppang et al. 2010). Moreover, fish skin constitutes a true mucosa as it is not keratinized and is coated by mucus (Xu et al. 2013).

The absence of lymph nodes and delimitated structures—especially in the skin—rises the issue of the selection of a relevant available *local* repertoire, well adapted to and actually shaped by the interactions with the microbiota present at mucosal surfaces. Such a repertoire would ideally reflect the selection of lymphocyte populations by the microbe populations (which can be more or less stable depending on the tissue and on the environment) and would express a diversity of receptors sufficient *locally* to mount efficient responses to a large set of potential pathogens. Thus, the structure of the B- and T-cell repertoires in naive and infected animals would provide many insights into the spatial dynamics of adaptive responses in the absence of integrated secondary structures as lymph nodes.

Although IgM+ and IgT+ B cells are present in fish gut, gills, and skin, the corresponding repertoires remain almost unknown. In mammals, CDR3 spectratyping analysis showed that the gut IgA (and IgM) repertoire contained highly expanded peaks suggesting oligoclonal proliferations (Holtmeier et al. 2000; Stoel et al. 2005). However, a recent study using deep sequencing found that the gut IgA available repertoire comprises a large diversity of low-frequency clones in addition to the highly expanded ones (Lindner et al. 2012). In fact, this likely reflects an equilibrium between the impact of gut bacteria on the B-cell repertoire and the influence of secreted Ig on the composition of the microbiota (Wei et al. 2011). It will be interesting to determine if a similar available repertoire is found in the gut of naive fish, with a limited number of highly expanded clones expressing IgM or IgT. It is interesting to note that in mice the selection of the IgA repertoire depends on T cells, microbiota, and RORgt but not on Peyer's patches; as fishes possess a true RORgt orthologue, IgT repertoire might be determined by a similar pathway (Lindner et al. 2012).

While it remains unknown if such a structure of B-cell repertoire is shared by fish and mouse or human, CDR3 spectratyping of TCR**β** in rainbow trout has revealed important differences in T-cell diversity. In men and mice, early studies of intraepithelial T lymphocytes (IEL) found oligoclonal repertoires of dominant clones in adult individuals (Gross et al. 1994; Regnault et al. 1994). In contrast, youngster IEL repertoires were highly diverse and polyclonal, with bell-shaped CDR3 length distributions in mammals and birds (Dunon et al. 1994; Williams et al. 2004). In contrast, rainbow trout IEL repertoire did not show restricted diversity in the young adult, suggesting very different selective constraints in fish compared to chicken, rodent, and human (Bernard et al. 2006). However, more restricted repertoires were found in older fish from farms, suggesting that IEL diversity might be reduced—or at least that large clones might be positively selected—in older fishes living in natural conditions (data not shown). While it is increasingly clear that commensal bacteria of the gut microbiota play a critical role in the peripheral selection of T-cell subsets such as T_{reg} (Lathrop et al. 2011), it must be noted that these observations in trout integrated all TCR**β**-expressing T cells, which might have masked even strong effects of selection on minor cell subsets. More detailed studies on sorted T-cell populations at different ages would be necessary to clarify the mechanisms of IEL selection in fish and differences in men and mice.

Mucosal responses also are particular. As previously stated, large accumulation of IgT^+ (but not IgM^+) B cells was found in the gut of rainbow trout surviving infection with the parasite *Ceratomyxa shasta* (myxosporidian) and in skin epidermis of fish infected by the ciliated *Ichthyophthirius* (Zhang et al. 2010; Xu et al. 2013). As fish B cells express either IgM or IgT, these observations suggest that mucosal adaptive response triggers activation and proliferation of specialized B cells expressing the mucosal isotype, while IgM^+ B cells responding to the pathogens would migrate and lead to a systemic response with circulating secreted Ab. The relative importance of local proliferation at the infection site versus infiltration and concentration of Ag-specific IgT^+ B cells from other territories remains unknown. However, it seems unlikely that the IgT^+ B cells locally present in the skin would be diverse enough to match any possible pathogen; hence, IgT^+ B cells likely would have to gather at inflammatory sites and the Ag-specific cells sampled and locally amplified. In interbranchial lymphoid tissue (ILT), a transcriptome analysis found a small delayed increase in IgT transcripts after infection with infectious salmon anemia virus by immersion, suggesting an expansion of IgT-expressing B cells (Austbø et al. 2014).

Regarding mucosal T cells, comparison of repertoires of naive and VHSV-infected trout showed that IEL were responsive to the infection; the most significant alterations of CDR3 length profiles were found for the same V**β** segments in the gut and in the spleen. Amplified clonotypes were also found in both tissues of infected fish, suggesting that T-cell clonotypes selected by a systemic viral infection were shared between gut mucosa and other peripheral (non-mucosal) lymphoid tissues. However, repertoire analysis of the response to the bacterium *Yersinia ruckeri* administered per anal indicated that gut and spleen contained different compartments at least during the first phase of the response (unpublished data). A first infection was performed using a vaccinal strain and was followed by a second infection with a virulent strain 3 weeks later; the TCR**β** repertoire was characterized 15 and 35 days after the second infection. Bacterium was not detected in the gut, but modifications of the IEL TCR**β** repertoire were already very clear, while in contrast, spleen and head kidney repertoires were not significantly altered. Three weeks later, the TCR**β** repertoire has been drastically modified in both tissues: the CDR3 length profiles of IEL are overall normalized (bell shaped as in naive animals), while those in spleen were highly biased. These observations showed that there is at least some level of compartmentalization between spleen and gut T-cell populations in rainbow trout, regarding responses to local infections. However, as for B-cell responses, the respective contributions of local T cells expanded in the mucosa versus T cells recruited from other tissues (mucosal or not) remain unclear.

Understanding the impact of the environment and local microbiota on the lymphocyte repertoires of fish mucosa is critical to improve natural resistance and to validate better vaccines against many pathogens that are notoriously difficult to fight; also, it provides an interesting model for comparative study with human mucosal immunity and may identify novel general mechanisms.

4 Conclusions

Somatic diversification of fish Ag receptors generally reflects the unity of RAG-based immunity across vertebrates, with shared recombination mechanisms and common patterns of clonal selection. However, unique configurations of adaptive immune system have been discovered in fish, such as the lack of *mhc II* and $CD4^+$ T cells in cod and pipefish, which questions the flexibility of the regulation of adaptive immunity.

The fish-specific Ig isotypes reflect the evolutionary plasticity of Ig classes across vertebrates. Interestingly, fish IgT, amphibian IgX, and mammalian IgA appear as convergent-specific adaptations to protect mucosa. In contrast, the four TcR types (α, β, γ, δ) are ubiquitously conserved from Chondrichthyans to mammals. This is certainly linked to the locked structure of the MHC/Ag/TcR tri-complex; however, this may not fully explain the conservation of TcR classes, especially for TcR γδ. A better understanding of fish TcR γδ functions might shed light on these aspects.

With very different lymphoid anatomical structures and microenvironments, fish and mammals provide a very good subject for comparative approaches to distinguish fundamental conserved properties of B or T cells and convergent adaptations driven by the necessity to fight pathogens in critical tissues or at entry points. Thus, the similarity of fish pronephros and mammalian bone marrow is striking as they both host B-cell differentiation and constitute the survival niche for memory B cells. In the same line, the lack of lymph nodes in fish raises issues about the sites of encounters between lymphocytes and antigens and about modalities of T-/B-cell cooperation.

Beyond the interest of the particularities of fish adaptive immunity, recent studies of fish immune repertoires have reactivated the debate on determinism and contingency in the generation of the expressed Ag-receptor diversity. One easily conceives that the unique immunological history of each individual would allege that contingency plays a key role in selecting somatically diversified immune repertoires. In keeping with this, the high complexity of the regulation of VDJ recombination (epigenetic status of VDJ segments, RSS efficiency, etc.) has impeded a full understanding of the establishment of primary repertoires. There are clear biases in the recombination of given VDJ segments, partly due to the RSS sequences, but the choice of recombining segments appears intuitively largely contingent. Interestingly, however, recent studies of zebrafish primary repertoire provided evidences in favor of determinism in this respect (Jiang et al. 2011). In fact, it remains difficult to evaluate the respective contributions of determinism and contingency in building immune repertoires; quoting Darwin: "*I have hitherto sometimes spoken as if the variations so common and multiform in organic beings {. . .} had been due to chance. This, of course, is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation.*"[1] Comparative studies will certainly provide important new insights about the importance of determinism in somatic adaptation of immune repertoires

[1] The origin of the species, Chap. 5, Laws of Variation

in the future: to this respect, the large diversity of fishes provides a rich resource to assess the impact of size, temperature regulation, and other parameters. This issue has even direct and important practical implications in aquaculture, as the determined arm of repertoire composition is accessible to genetic selection, and the contingent arm due to the immunological history of the fish is the target of vaccinations.

Acknowledgements This article is dedicated to the memory of Steve Kaattari, who pioneered the study of fish B cells and will be missed by fish immunologists after passing away in November 2014.

This work was supported by Institut National de la Recherche Agronomique, by the European Commission under the Work Programme 2012 of the 7th Framework Programme for Research and Technological Development of the European Union (Grant Agreement 311993 TARGETFISH), and by the National Institutes of Health Grant R01GM085207 (to J.O.S.). We acknowledge S. Fillatreau, T. Mora, A. Six, and Dr. G. Wiegertjes for helpful discussions.

References

Alugupalli KR, Akira S, Lien E, Leong JM (2007) MyD88- and Bruton's tyrosine kinase-mediated signals are essential for T cell-independent pathogen-specific IgM responses. J Immunol 178:3740–3749

Arnesen KR, Mikkelsen H, Schrøder MB, Lund V (2010) Impact of reattaching various Aeromonas salmonicida A-layer proteins on vaccine efficacy in Atlantic cod (*Gadus morhua*). Vaccine 28:4703–4708

Austbø L, Bergva Aas I, König M et al (2014) Transcriptional response of immune genes in gills and the interbranchial lymphoid tissue of Atlantic salmon challenged with infectious salmon anaemia virus. Dev Comp Immunol 45:107–114

Bao Y, Wang T, Guo Y et al (2010) The immunoglobulin gene loci in the teleost *Gasterosteus aculeatus*. Fish Shellfish Immunol 28:40–48

Barreto VM, Magor BG (2011) Activation-induced cytidine deaminase structure and functions: a species comparative view. Dev Comp Immunol 35:991–1007

Barreto V, Meo T, Cumano A (2001) Mice triallelic for the Ig heavy chain locus: implications for VHDJH recombination. J Immunol 166:5638–5645

Barreto VM, Pan-Hammarstrom Q, Zhao Y et al (2005) AID from bony fish catalyses call switch recombination. J Exp Med 202:733

Ben-Hamo R, Efroni S (2011) The whole-organism heavy chain B cell repertoire from Zebrafish self-organizes into distinct network features. BMC Syst Biol 5:27

Bernard D, Six A, Rigottier-Gois L et al (2006) Phenotypic and functional similarity of gut intraepithelial and systemic T cells in a teleost fish. J Immunol 176:3942–3949

Boudinot P, Boubekeur S, Benmansour A (2001) Rhabdovirus infection induces public and private T cell responses in teleost fish. J Immunol 167:6202–6209

Boudinot P, Boubekeur S, Benmansour A (2002) Primary structure and complementarity-determining region (CDR) 3 spectratyping of rainbow trout TCRbeta transcripts identify ten Vbeta families with Vbeta6 displaying unusual CDR2 and differently spliced forms. J Immunol 169:6244–6252

Boudinot P, Bernard D, Boubekeur S et al (2004) The glycoprotein of a fish rhabdovirus profiles the virus-specific T-cell repertoire in rainbow trout. J Gen Virol 85:3099–3108

Brodeur PH, Wortis HH (1980) Regulation of thymus-independent responses: unresponsiveness to a second challenge of TNP-Ficoll is mediated by hapten-specific antibodies. J Immunol 125:1499–1505

Burnet FM (1957) A modification of Jerne's theory of antibody production using the concept of clonal selection. Aust J Sci 20:67–69

Castro R, Jouneau L, Pham H-P et al (2013) Teleost fish mount complex clonal IgM and IgT responses in spleen upon systemic viral infection. PLoS Pathog 9, e1003098

Cohn M (1994) The wisdom of hindsight. Annu Rev Immunol 12:1–62

Costa G, Danz H, Kataria P, Bromage E (2012) A holistic view of the dynamisms of teleost IgM: a case study of *Streptococcus iniae* vaccinated rainbow trout (*Oncorhynchus mykiss*). Dev Comp Immunol 36:298–305

Daggfeldt A, Bengtén E, Pilström L (1993) A cluster type organization of the loci of the immunoglobulin light chain in Atlantic cod (*Gadus morhua L.*) and rainbow trout (*Oncorhynchus mykiss* Walbaum) indicated by nucleotide sequences of cDNAs and hybridization analysis. Immunogenetics 38:199–209

Danilova N, Bussmann J, Jekosch K, Steiner LA (2005) The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat Immunol 6:295–302

Das S, Hirano M, Tako R et al (2012) Evolutionary genomics of immunoglobulin-encoding Loci in vertebrates. Curr Genomics 13:95–102

De Togni P, Goellner J, Ruddle NH et al (1994) Abnormal development of peripheral lymphoid organs in mice deficient in lymphotoxin. Science 264:703–707

Defrance T, Taillardet M, Genestier L (2011) T cell-independent B cell memory. Curr Opin Immunol 23:330–336

DeKosky BJ, Ippolito GC, Deschner RP et al (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. Nat Biotechnol 31:166–169

Diaz M, Velez J, Singh M et al (1999) Mutational pattern of the nurse shark antigen receptor gene (NAR) is similar to that of mammalian Ig genes and to spontaneous mutations in evolution: the translesion synthesis model of somatic hypermutation. Int Immunol 11:825–833

Diaz M, Flajnik MF, Klinman N (2001) Evolution and the molecular basis of somatic hypermutation of antigen receptor genes. Philos Trans R Soc Lond B Biol Sci 356:67–72

Dijkstra JM, Grimholt U, Leong J et al (2013) Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. BMC Evol Biol 13:260

Dooley H, Stanfield RL, Brady RA, Flajnik MF (2006) First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. Proc Natl Acad Sci USA 103:1846–1851

Du Pasquier L, Hsu E (1983) Immunoglobulin expression in diploid and polyploid interspecies hybrid of *Xenopus*: evidence for allelic exclusion. Eur J Immunol 13:585–590

Dunon D, Schwager J, Dangy JP et al (1994) T cell migration during development: homing is not related to TCR V beta 1 repertoire selection. EMBO J 13:808–815

Eason DD, Litman RT, Luer CA et al (2004) Expression of individual immunoglobulin genes occurs in an unusual system consisting of multiple independent loci. Eur J Immunol 34:2551–2558

Espelid S, Rødseth O, Jørgensen T et al (1991) Vaccination experiments and studies of the humoral immune responses in cod, *Gadus morhua* L., to four strains of monoclonal-defined *Vibrio anguillarum*. J Fish Dis 14:185–197

Findly RC, Zhao X, Noe J et al (2013) B cell memory following infection and challenge of channel catfish with Ichthyophthirius multifiliis. Dev Comp Immunol 39:302–311

Flajnik M, Du Pasquier L (2013) Evolution of the immune system. In: Paul W (ed) Fundamental Immunology, 7th edn. Wolters Kluwer & Lippincott Williams & Wilkins, New York, pp 67–128

Francés V, Pandrau-Garcia D, Guret C et al (1994) A surrogate 15 kDa JC kappa protein is expressed in combination with mu heavy chain by human B cell precursors. EMBO J 13:5937–5943

Gambón-Deza F, Sánchez-Espinel C, Magadán-Mompó S (2010) Presence of an unique IgT on the IGH locus in three-spined stickleback fish (*Gasterosteus aculeatus*) and the very recent generation of a repertoire of VH genes. Dev Comp Immunol 34:114–122

Gross GG, Schwartz VL, Stevens C et al (1994) Distribution of dominant T cell receptor beta chains in human intestinal mucosa. J Exp Med 180:1337–1344

Gudmundsdóttir S, Magnadóttir B, Björnsdóttir B et al (2009) Specific and natural antibody response of cod juveniles vaccinated against *Vibrio anguillarum*. Fish Shellfish Immunol 26:619–624

Haire RN, Rast JP, Litman RT, Litman GW (2000) Characterization of three isotypes of immunoglobulin light chains and T-cell antigen receptor alpha in zebrafish. Immunogenetics 51:915–923

Hansen J, Landis E, Phillips R (2005) Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: implications for a distinctive B cell developmental pathway in teleost fish. Proc Natl Acad Sci 102:6919–6924

Haugarvoll E, Bjerkås I, Nowak BF et al (2008) Identification and characterization of a novel intraepithelial lymphoid tissue in the gills of Atlantic salmon. J Anat 213(2):202–209

Helfman G, Collette B, Facey D, Bowen B (2010) The diversity of fishes, 2nd edn. Wiley Blackwell, Hoboken

Herrin BR, Cooper MD (2010) Alternative adaptive immunity in jawless vertebrates. J Immunol 185:1367–1374

Hirano M, Das S, Guo P, Cooper MD (2011) The evolution of adaptive immunity in vertebrates. Adv Immunol 109:125–157

Hofmann J, Greter M, Du Pasquier L, Becher B (2010) B-cells need a proper house, whereas T-cells are happy in a cave: the dependence of lymphocytes on secondary lymphoid tissues during evolution. Trends Immunol 31:144–153

Holtmeier W, Hennemann A, Caspary WF (2000) IgA and IgM V(H) repertoires in human colon: evidence for clonally expanded B cells that are widely disseminated. Gastroenterology 119:1253–1266

Hsu E (2009) V(D)J recombination: of mice and sharks. Adv Exp Med Biol 650:166–179

Hsu E, Criscitiello MF (2006) Diverse Immunoglobulin Light Chain Organizations in Fish Retain Potential to Revise B Cell Receptor Specificities. J Immunol 177(4):2452–2462

Hsu E, Pulham N, Rumfelt LL, Flajnik MF (2006) The plasticity of immunoglobulin gene systems in evolution. Immunol Rev 210:8–26

Huetz F, Carlsson L, Tornberg UC, Holmberg D (1993) V-region directed selection in differentiating B lymphocytes. EMBO J 12:1819–1826

Israelsson O, Petersson A, Bengtén E et al (1991) Immunoglobulin concentration in Atlantic cod, *Gadus morhua* L., serum and -reactivity between anti-cod-antibodies and immunoglobulins from other species. J Fish Biol 39:265–278

Jerne NK (1955) The natural selection theory of antibody formation. Proc Natl Acad Sci USA 41:849

Jerne N (1971) What precedes clonal selection? Ontogeny of acquired immunity. In Proceedings of CIBA Foundation Symposium, CIBA, Amsterdam, pp 1–15

Jhunjhunwala S, van Zelm MC, Peak MM, Murre C (2009) Chromatin architecture and the generation of antigen receptor diversity. Cell 138:435–448

Jiang N, Weinstein JA, Penland L et al (2011) Determinism and stochasticity during maturation of the zebrafish antibody repertoire. Proc Natl Acad Sci USA 108:5348–5353

Kaattari SL, Zhang HL, Khor IW et al (2002) Affinity maturation in trout: clonal dominance of high affinity antibodies late in the immune response. Dev Comp Immunol 26:191–200

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. Immunogenetics 56:683–695

Koppang EO, Fischer U, Moore L et al (2010) Salmonid T cells assemble in the thymus, spleen and in novel interbranchial lymphoid tissue. J Anat 217:728–739
Kurath G, Garver KA, Corbeil S et al (2006) Protective immunity and lack of histopathological damage two years after DNA vaccination against infectious hematopoietic necrosis virus in trout. Vaccine 24:345–354
Lathrop SK, Bloom SM, Rao SM et al (2011) Peripheral education of the immune system by colonic commensal microbiota. Nature 478:250–254
Lindner C, Wahl B, Föhse L et al (2012) Age, microbiota, and T cells shape diverse individual IgA repertoires in the intestine. J Exp Med 209:365–377
Ma C, Ye J, Kaattari SL (2013) Differential compartmentalization of memory B cells versus plasma cells in salmonid fish. Eur J Immunol 43:360–370
Magadán-Mompó S, Sánchez-Espinel C, Gambón-Deza F (2011) Immunoglobulin heavy chains in medaka (*Oryzias latipes*). BMC Evol Biol 11:165
Magadán-Mompó S, Zimmerman AM, Sánchez-Espinel C, Gambón-Deza F (2013) Immunoglobulin light chains in medaka (*Oryzias latipes*). Immunogenetics 65:387–396
Magnadottir B, Jonsdottir H, Helgason S et al (1999) Humoral immune parameters in Atlantic cod (*Gadus morhua* L.), II: the effects of size and gender under different environmental conditions. Comp Biochem Physiol 122:181–188
Malecek K, Lee V, Feng W et al (2008) Immunoglobulin heavy chain exclusion in the shark. PLoS Biol 6, e157
Malmstrøm M, Jentoft S, Gregers TF, Jakobsen KS (2013) Unraveling the evolution of the Atlantic cod's (*Gadus morhua* L.) alternative immune strategy. PLoS One 8, e74004
Marianes AE, Zimmerman M, Zimmerman AM (2011) Targets of somatic hypermutation within immunoglobulin light chain genes in zebrafish. Immunology 132:240–255
Mårtensson IL, Keenan RA, Licence S (2007) The pre-B-cell receptor. Curr Opin Immunol 19:137–142
Matsumoto M, Lo SFF, Carruthers CJJ et al (1996) Affinity maturation without germinal centres in lymphotoxin-alpha-deficient mice. Nature 382:462–466
Matsunaga T, Rahman A (1998) What brought the adaptive immune system to vertebrates? The jaw hypothesis and the seahorse. Immunol Rev 166:177–186
Mikkelsen H, Lund V, Larsen R, Seppola M (2011) Vibriosis vaccines based on various serosubgroups of *Vibrio anguillarum* O2 induce specific protection in Atlantic cod (*Gadus morhua* L.) juveniles. Fish Shellfish Immunol 30:330–339
Miller NW, Clem LW (1984) Microsystem for in vitro primary and secondary immunization of channel catfish (*Ictalurus punctatus*) leukocytes with hapten-carrier conjugates. J Immunol Methods 72:367–379
Miller NW, Sizemore RC, Clem LW (1985) Phylogeny of lymphocyte heterogeneity: the cellular requirements for in vitro antibody responses of channel catfish leukocytes. J Immunol 134:2884–2888
Mora T, Walczak AM, Bialek W, Callan CG (2010) Maximum entropy models for antibody diversity. Proc Natl Acad Sci USA 107:5405–5410
Nakanishi T, Toda H, Shibasaki Y, Somamoto T (2011) Cytotoxic T cells in teleost fish. Dev Comp Immunol 35:1317–1323
Nelson JS (2006) Fishes of the world, 4th edn. Wiley, Hoboken
Obukhanych TV, Nussenzweig MC (2006) T-independent type II immune responses generate memory B cells. J Exp Med 203:305–310
Oltz EM (2001) Regulation of antigen receptor gene assembly in lymphocytes. Immunol Res 23:121–133
Persson AC, Stet RJ, Pilström L (1999) Characterization of MHC class I and beta(2)-microglobulin sequences in Atlantic cod reveals an unusually high number of expressed class I genes. Immunogenetics 50:49–59
Petit J, Stange-Thomann N, Mauceli E et al (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946–957

Pilström L, Petersson A (1991) Isolation and partial characterization of immunoglobulin from cod (*Gadus morhua* L.). Dev Comp Immunol 15:143–152
Pilstrom L, Warr G, Stromberg S (2005) Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation. Fish Sci 71:961–971
Rangel R, McKeller MR, Sims-Mourtada JC et al (2005) Assembly of the kappa preB receptor requires a V kappa-like protein encoded by a germline transcript. J Biol Chem 280:17807–17814
Rast JP, Litman GW (1998) Towards understanding the evolutionary origins and early diversification of rearranging antigen receptors. Immunol Rev 166:79–86
Rast JP, Amemiya CT, Litman RT et al (1998) Distinct patterns of IgH structure and organization in a divergent lineage of chrondrichthyan fishes. Immunogenetics 47:234–245
Ratcliffe MJ (2006) Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. Dev Comp Immunol 30:101–118
Regnault A, Cumano A, Vassalli P et al (1994) Oligoclonal repertoire of the CD8 alpha alpha and the CD8 alpha beta TCR-alpha/beta murine intestinal intraepithelial T lymphocytes: evidence for the random emergence of T cells. J Exp Med 180:1345–1358
Rombout JH, Abelli L, Picchietti S et al (2011) Teleost intestinal immunology. Fish Shellfish Immunol 31:616–626
Romer A (1962) The vertebrate body, 3rd edn. Saunders WB Co., Philadelphia
Salinas I, Zhang Y-A, Sunyer JO (2011) Mucosal immunoglobulins and B cells of teleost fish. Dev Comp Immunol 35:1346–1365
Sammut B, Du Pasquier L, Ducoroy P et al (1999) Axolotl MHC architecture and polymorphism. Eur J Immunol 29:2897–2907
Saunders HL, Oko ALAL, Scott ANAN et al (2010) The cellular context of AID expressing cells in fish lymphoid tissues. Dev Comp Immunol 34:669
Schröder AE, Greiner A, Seyfert C, Berek C (1996) Differentiation of B cells in the nonlymphoid tissue of the synovial membrane of patients with rheumatoid arthritis. Proc Natl Acad Sci USA 93:221–225
Schrøder MB, Ellingsen T, Mikkelsen H et al (2009) Comparison of antibody responses in Atlantic cod (*Gadus morhua* L.) to Vibrio anguillarum, Aeromonas salmonicida and Francisella sp. Fish Shellfish Immunol 27:112–119
Shapiro D, Adkison M, Kaattari S (1996) Antibody affinity analysis using the ELISA. In: Lefkovitz I Jr (ed) Immunology methods manual: comprehensive sourcebook of techniques. Academic, New York, p 2353e65
Shlomchik MJ, Weisel F (2012) Germinal center selection and the development of memory B and plasma cells. Immunol Rev 247:52–63
Six A, Mariotti-Ferrandiz ME, Chaara W et al (2013) The past, present, and future of immune repertoire biology – the rise of next-generation repertoire analysis. Front Immunol 4:413
Smelty P, Marchal C, Renard R et al (2010) Identification of the pre-T-cell receptor alpha chain in nonmammalian vertebrates challenges the structure-function of the molecule. Proc Natl Acad Sci USA 107:19991–19996
Somamoto T, Okamoto N, Nakanishi T et al (2009) In vitro generation of viral-antigen dependent cytotoxic T-cells from ginbuna crucian carp, Carassius auratus langsdorfii. Virology 389:26–33
Somamoto T, Kondo M, Nakanishi T, Nakao M (2014) Helper function of $CD4^+$ lymphocytes in antiviral immunity in ginbuna crucian carp, *Carassius auratus langsdorfii*. Dev Comp Immunol 44:111–115
Star B, Jentoft S (2012) Why does the immune system of Atlantic cod lack MHC II? Bioessays 34:648–651
Star B, Nederbragt AJ, Jentoft S et al (2011) The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210
Stavnezer J, Amemiya CT (2004) Evolution of isotype switching. Semin Immunol 16:257–275

Stoel M, Jiang H, van Diemen CC et al (2005) Restricted IgA repertoire in both B-1 and B-2 cell-derived gut plasmablasts. J Immunol 174:1046–1054
Tacchi L, Musharrafieh R, Larragoite ET et al (2014) Nasal immunity is an ancient arm of the mucosal immune system of vertebrates. Nat Commun 5:5205
Takizawa F, Dijkstra JM, Kotterba P et al (2011) The expression of CD8α discriminates distinct T cell subsets in teleost fish. Dev Comp Immunol 35:752–763
Tan Y-C, Blum LK, Kongpachith S et al (2014) High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. Clin Immunol 151:55–65
Tanaka Y, Nakasone H, Yamazaki R et al (2010) Single-cell analysis of T-cell receptor repertoire of HTLV-1 Tax-specific cytotoxic T cells in allogeneic transplant recipients with adult T-cell leukemia/lymphoma. Cancer Res 70:6181–6192
Tatner MF (1986) The ontogeny of humoral immunity in rainbow trout, *Salmo gairdneri*. Vet Immunol Immunopathol 12:93–105
Toda H, Saito Y, Koike T et al (2011) Conservation of characteristics and functions of CD4 positive lymphocytes in a teleost fish. Dev Comp Immunol 35:650–660
Tournefier A, Laurens V, Chapusot C et al (1998) Structure of MHC class I and class II cDNAs and possible immunodeficiency linked to class II expression in the Mexican axolotl. Immunol Rev 166:259–277
Utke K, Kock H, Schuetze H et al (2008) Cell-mediated immune responses in rainbow trout after DNA immunization against the viral hemorrhagic septicemia virus. Dev Comp Immunol 32:239–252
Van Der Aa LM, Levraud J, Yahmi M et al (2009) A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. BMC Biol 23:7
Wei M, Shinkura R, Doi Y et al (2011) Mice carrying a knock-in mutation of Aicda resulting in a defect in somatic hypermutation have impaired gut homeostasis and compromised mucosal defense. Nat Immunol 12:264–270
Weill JC, Cocea L, Reynaud C-A (2002) Allelic exclusion: lesson from GALT species. Semin Immunol 14:213–215, discussion 227–228
Weinstein JA, Jiang N, White RA et al (2009) High-throughput sequencing of the zebrafish antibody repertoire. Science 324:807–811
Williams AM, Bland PW, Phillips AC et al (2004) Intestinal alpha beta T cells differentiate and rearrange antigen receptor genes in situ in the human infant. J Immunol 173:7190–7199
Wilson M, Hsu E, Marcuz A et al (1992) What limits affinity maturation of antibodies in *Xenopus*-the rate of somatic mutation or the ability to select mutants? EMBO J 11:4337–4347
Xu Z, Parra D, Gómez D et al (2013) Teleost skin, an ancient mucosal surface that elicits gut-like immune responses. Proc Natl Acad Sci USA 110:13097–13102
Yang F, Waldbieser GC, Lobb CJ (2006) The nucleotide targets of somatic mutation and the role of selection in immunoglobulin heavy chains of a teleost fish. J Immunol 176:1655
Yasuike M, De Boer J, Von Schalburg KR et al (2010) Evolution of duplicated IgH loci in Atlantic salmon, *Salmo salar*. BMC Genomics 11:486
Ye J, Kaattari IM, Kaattari SL (2011) The differential dynamics of antibody subpopulation expression during affinity maturation in a teleost. Fish Shellfish Immunol 30:372–377
Zhang Y, Salinas I, Li J et al (2010) IgT, a primitive immunoglobulin class specialized in mucosal immunity. Nat Immunol 11:827–835
Zhu C, Feng W, Weedon J et al (2011) The multiple shark Ig chain genes rearrange and hypermutate autonomously. J Immunol 187:2492
Zielinski CE, Corti D, Mele F et al (2011) Dissecting the human immunologic memory for pathogens. Immunol Rev 240:40–51

The Evolution and Structure of Atypical T Cell Receptors

Victoria L. Hansen and Robert D. Miller

Abstract The T cell receptor structure and genetic organization have been thought to have been stable in vertebrate evolution relative to the immunoglobulins. For the most part, this has been true and the content and organization of T cell receptor genes has been fairly conserved over the past 400 million years of gnathostome evolution. Analyses of TCRδ chains in a broad range of vertebrate lineages over the past decade have revealed a remarkable and previously unrealized degree of plasticity. This plasticity can generally be described in two forms. The first is broad use of antibody heavy chain variable genes in place of the conventional Vδ. The second form containing an unusual three extracellular domain structures has evolved independently in both cartilaginous fishes and mammals. Two well-studied vertebrate lineages, the eutherian mammals such as mice and humans and teleost fishes, lack any of these alternative TCR forms, contributing to why they went undiscovered for so long after the initial description of the conventional TCR chains three decades ago. This chapter describes the state of knowledge of these unusual TCR forms, both their structure and genetics, and current ideas on their function.

1 Introduction

At the center of adaptive immune responses are the receptors that specifically recognize epitopes or antigens associated with pathogens. In the jawed vertebrates, or gnathostomes, the first of these receptors to be discovered were the antibodies, later also called immunoglobulins (Ig). Antibodies were discovered as a soluble activity in the bloodstream of immune animals in the late nineteenth century (von Behring 1901). The structure of Ig and the use of Ig as an antigen receptor on the surface of antibody producing B cells had been established by the end of the 1960s (Raff 1970; Porter 1973). It would not be until the early 1980s, nearly a century

V.L. Hansen • R.D. Miller (✉)
Department of Biology, Center for Evolutionary and Theoretical Immunology, University of New Mexico, Albuquerque, NM 87131, USA
e-mail: rdmiller@unm.edu

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_11

after the discovery of antibodies, that the antigen receptors used by a second lineage of gnathostome lymphocytes, the T cells, would be discovered (Hedrick et al. 1984).

The discovery of the T cell receptors (TCRs) was the result of a series of elegant experiments by Mark Davis and colleagues. They identified T cell-specific transcripts that appeared to be from germ-line genes undergoing somatic DNA recombination and encoding proteins with homology to immunoglobulins (Ig) (Hedrick et al. 1984). We now know that those first transcripts encoded the TCRβ chain, and within the next 2 years the Davis laboratory and other investigators would identify three additional TCR chains: TCRα, TCRγ, and TCRδ (Saito et al. 1984; Waldmann et al. 1985; Brenner et al. 1986; Loh et al. 1987). The discovery of four distinct TCR chains quickly leads to the recognition of two distinct T cell subpopulations defined by the composition of their TCR heterodimers: the αβ T cell and the γδ T cell (Havran and Allison 1988). Over the next decade it was firmly established that all jawed vertebrates had αβ and γδ T cells and that the genes encoding the TCR chains were well conserved from mammals to sharks (Rast et al. 1995, 1997).

The conservation of α, β, γ, and δ TCR chains and their pairing rules across gnathostomes is in contrast to the immunoglobulins (Ig) where there has been lineage-specific evolution of novel isotypes, such as IgG and IgE being unique to mammals, IgX in the amphibians, IgA in amniotes, IgT in teleost fish, etc. (Hsu et al. 1985; Warr et al. 1995; Danilova et al. 2005; Hansen et al. 2005). This contrast between Ig and TCR may not be surprising since TCRs primarily perform antigen recognition and signaling roles, whereas the Igs are both recognition/signaling cell surface receptors and effector molecules in their secreted or soluble forms (Raff 1970). Pressure to evolve isotypes with different functions or characteristics has acted on the Ig molecules in ways that are not exerted on TCR. For example, all bony vertebrates, from teleosts to mammals, have evolved an Ig specialized for mucosal surfaces. This role is performed by IgT in teleosts, IgX in amphibians, and IgA in amniotes (Mußmann et al. 1996; Zhang et al. 2010). Some isotypes have been lost in different lineages. The IgD/IgW family, for example, appears to have been retained and lost independently many times in evolution (Ota et al. 2003; Ohta and Flajnik 2006).

In contrast to the Ig isotypes, the αβ and γδ TCR appear to have been retained throughout gnathostome evolution without much modification. Indeed, some basic "rules" to the immune system appear to have been established for T or T-like cells early in vertebrate evolution (Criscitiello and de Figueiredo 2013). All vertebrates have one lymphocyte type that can convert its antigen receptor to a secreted effector molecule called the antibody, and two other cell types that do not secrete their receptor. B cells and αβ and γδ T cells fulfill this function in gnathostomes (Reviewed in Criscitiello and de Figueiredo 2013). Agnathans, such as the lamprey and hagfish, have variable lymphocyte receptor B (VLRB) expressing cells that secrete VLRB molecules as antibody. They also have VLRA+ and VLRC+ cells that do not secrete their receptor and may have functions analogous to αβ and γδ T cells, respectively (Guo et al. 2009; Kasamatsu et al. 2010). This rule evolved independently of the protein structure used to form the receptors. Gnathostomes use

Ig domain-based receptors, whereas agnathans use leucine-rich repeat (LRR) domain-based receptors (Pancer et al. 2004).

Although Ig and TCR are independent gene families in extant vertebrates, their common evolutionary origin is unquestioned (Marchalonis et al. 1997). The conservation of the enzymatic machinery for DNA rearrangement, the recombination recognition sequences that flank the gene segments, and the conserved gene organization all speak to a single evolutionary origin of Ig and TCR.

An early theory on the nature of the TCR predicted that T cells were using a different set of constant regions but the same variable (V), diversity (D), and joining (J) gene segments as antibodies. This hypothesis predicted that there would be constant region genes uniquely expressed in T cells that would be linked to the constant regions of the Ig heavy (*IgH*) chain locus. Testing this hypothesis, Frances Owen and Roy Riblet generated monoclonal antibodies (mAb) against mouse T cells and selected for those that recognized epitopes linked to the *IgH* locus on chromosome 12. They used cross-immunizations between *IgH* congenic strains to generate allotype-specific mAb (Owen and Riblet 1984). While successful in identifying *IgH*-linked T cell epitopes, these studies did not lead to the discovery of the TCR as hoped. Nonetheless, today their ideas seem prescient considering the recent discoveries of how Ig and TCR genes have teamed up to produce novel TCR forms in some vertebrate lineages. These and other novel TCRs found in non-eutherian mammals and non-mammals are the topic of this chapter.

Rather than present the discoveries made over the past decade chronologically, this chapter will attempt to present these findings in a way that aids in understanding of how the TCRδ locus may have evolved to produce a new TCR loci. As will be discussed, the TCRδ locus appears to have demonstrated a degree of plasticity not found in other TCR genes.

2 The Link Between Ig and TCR

In all sarcopterygian vertebrates examined, from coelacanths to humans, the genes encoding the TCRδ chain are nested within, or tightly linked to, those encoding TCRα (Isobe et al. 1988; Parra et al. 2008, 2012a, b; Saha et al. 2014). In some teleost fishes, TCRα and TCRδ also appear to be tightly linked, making a case for this being an ancient relationship (Fischer et al. 2002; Nam et al. 2003). One feature of the nested arrangement is an overlapping pool of V gene segments by TCRδ and TCRα during V(D)J recombination. Some, if not all, Vδ gene segments can be recombined either to a D segment when used to encode a TCRδ chain or a J segment in the case of TCRα (Olaru et al. 2005; Pellicci et al. 2014).

A more interesting example of this plasticity was demonstrated by Criscitiello and colleagues, when they found translocus V(D)J recombination between Ig and TCR loci in sharks (Criscitiello et al. 2010). The cartilaginous fish organize their antibody genes in clusters rather than the translocon organization found in bony vertebrates (Hinds and Litman 1986). Shark Ig loci are organized as clusters of

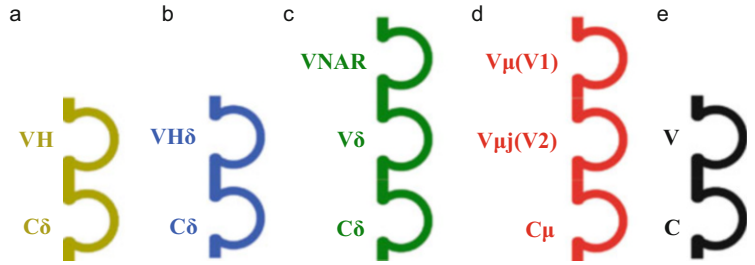


Fig. 1 Diagrams of the various novel TCR chains described in this chapter. (**a**) TCRδ chains found in sharks that are the result of VH genes from neighboring *IgH* loci recombining with Dδ and Jδ genes in the TCRδ locus. (**b**) TCRδ chains found in coelacanth, frog, birds, and prototherians that are the result of VH genes translocating into a TCRδ locus to become VHδ genes. (**c**) NAR-TCR found in cartilaginous fish. The N-terminal Ig domain is related to those found in IgNAR antibodies. The second or supporting V domain is a specialized Vδ. (**d**) TCRμ is found in the prototherian and metatherian mammals. In metatherians the second or supporting V domain, Vμj, is encoded by a germ-line joined exon. In the prototherians the two V domains have been designated V1 and V2 and both are encoded by exons that are the product of somatic V(D)J recombination. (**e**) A generic, two-domain TCR chain for comparison

genes containing V, D, J, and C gene segments necessary to encode IgH chains, and these clusters are dispersed around the shark genome. This cluster arrangement has allowed V(D)J recombination between V gene segments in an IgH locus (VH) to D and J gene segments in the TCRδ locus to encode TCRδ chains (Criscitiello et al. 2010). These are diagramed in the figures as TCRδ–VH chains and have only been described in sharks (Figs. 1a and 2).

Among the Teleostomi, and specifically the Sarcopterygii, a different gene organization has evolved to allow TCRδ chains to utilize VH genes. This appears to have involved the direct translocation of VH gene segments into the germ-line TCRα/TCRδ locus. This was first found in the amphibian *Xenopus* but has since been described in amniotes such as birds and prototherians, as well as in sarcopterygian fish, the coelacanth (Parra et al. 2010, 2012a, b; Saha et al. 2014). These gene segments have been designated VHδ genes and in all cases appear indistinguishable from VH found in *IgH* loci but are recombined and used to encode TCRδ chains (Figs. 1b and 2). In the case of the coelacanth TCRδ locus, VHδ genes are present in the germ-line TCRα/δ locus, but so far no transcripts containing this gene segment have been reported (Saha et al. 2014). This is likely due to an absence of data given that in all other species where VHδ are present, they are somatically recombined, expressed, and contribute to the expressed repertoire (Parra et al. 2010, 2012a, b).

Genomic organization information is available for the *TCRα/TCRδ* locus in the amphibian *Xenopus tropicalis*, where the genes encoding the TCRδ chains are in an inverted orientation relative to those encoding TCRα (Parra et al. 2010). The VHδ gene segments are in the same orientation as the rest of the TCRδ genes and outnumber the conventional Vδ gene segments. In *X. laevis*, the *IgH* and *TCRδ* loci are tightly linked, an arrangement that may have facilitated the translocation of

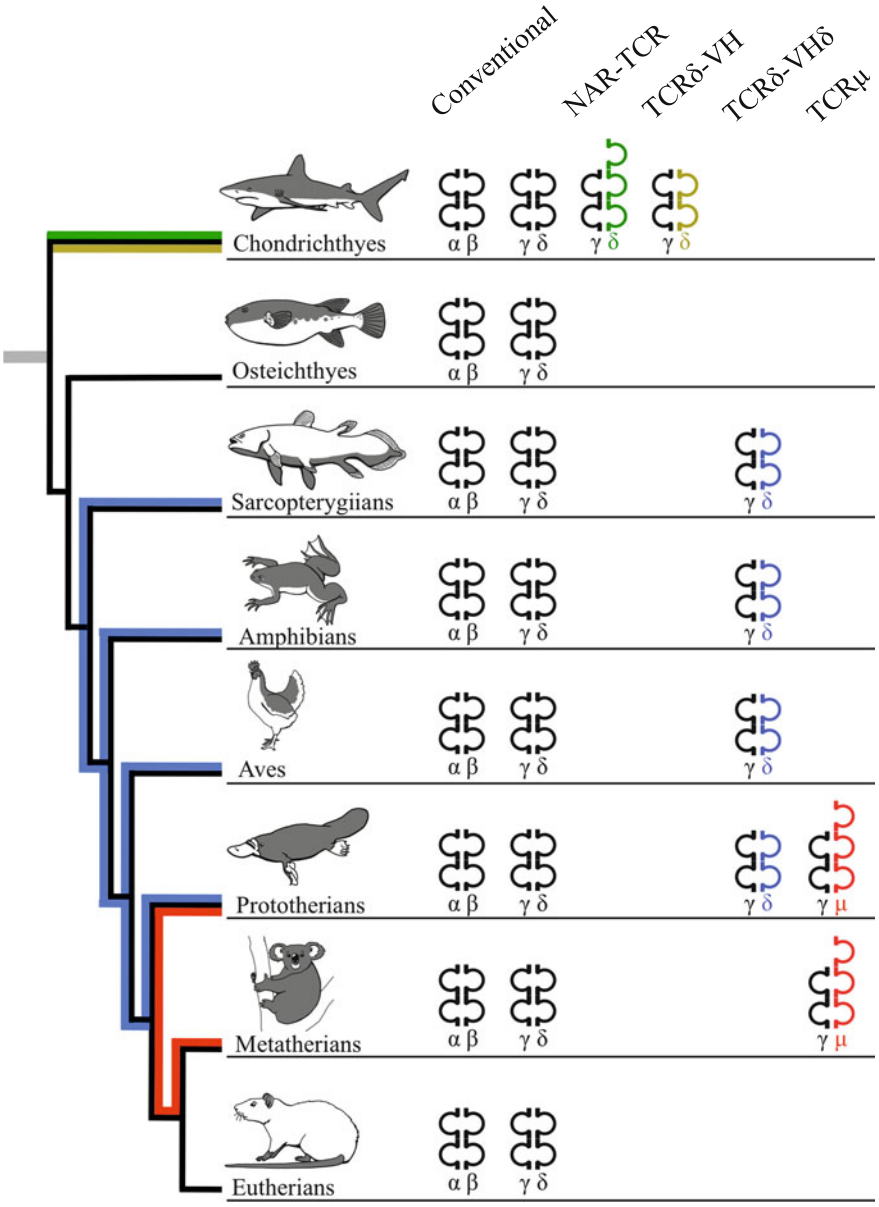


Fig. 2 Phylogenetic tree of the vertebrate lineages discussed in this chapter and their TCR repertoires. NAR-TCR and TCRδ using VH from *IgH* loci are only found so far in cartilaginous fishes, the Chondrichthyes. NAR-TCR has been described in both the Holocephali and the Elasmobranchii (for review, see Flajnik et al. 2011). TCRδ using VH have been described so far only in the Elasmobranchii. TCRδ using VHδ located within the locus encoding TCRδ chains have been described in the Sarcopterygii, from coelacanth to prototherian mammals. TCRμ is mammal specific and is found in both prototherians and metatherians but was lost in the eutherians

VH genes into the *TCRδ* locus creating the VHδ (Parra et al. 2010). There was no evidence for translocus V(D)J recombination between the frog VH genes in the *IgH* locus and genes in the *TCRδ* locus, as was reported for sharks (Parra et al. 2010; Criscitiello et al. 2010). It is noteworthy that the *X. tropicalis TCRα/TCRδ* locus contains only a single Cδ gene. As we move across the evolutionary tree to the amniotes, we find more complex evolution having occurred in the birds involving gene duplication of the TCRδ genes.

In the passerine bird, *Taeniopygia guttata* (zebra finch), the *TCRα/TCRδ* locus is much like that of *X. tropicalis* by having Vα, Vδ, and VHδ gene segments (Parra et al. 2012b). A duplication event in the birds has resulted in the presence of two Cδ genes, however, and a reduction to only a single VHδ. This single VHδ is recombined and transcribed to encode TCRδ chains (Parra et al. 2012b). In the distantly related Galliformes, such as chicken and turkey, the VHδ, along with Dδ, Jδ, and Cδ gene segments, have trans-located to create a second unlinked locus encoding TCRδ chains using only VHδ (Parra et al. 2012b). Galliform birds, therefore, have evolved two TCRδ loci. One is within the conventional *TCRα/TCRδ* locus and is fairly typical in structure relative to mouse and human standards. This locus encodes both TCRα and TCRδ chains using Vα and Vδ gene segments only (Parra and Miller 2012). The second locus is dedicated to encoding TCRδ chains using only VHδ (Parra et al. 2012b). The translocation of genes encoding TCRδ–VHδ chains to a separate locus is likely a galliform-specific event since it is not found in the passerines and has not been found outside of birds.

The duckbill platypus, *Ornithorhynchus anatinus*, is the only other species for which VHδ have been described, demonstrating that these genes persisted into the mammalian lineage (Wang et al. 2011). The platypus holds a special place in the story of TCRδ and will be discussed in more detail later in this chapter.

3 Why Have Two Ig Domains when You Can Have Three?

In 2006, Criscitiello and colleagues reported finding transcripts in sharks that would encode a form of TCRδ chain predicted to have three extracellular Ig domains (Criscitiello et al. 2006). The N-terminal domain was most related to the V domains found in an unusual form of antibody unique to sharks called Ig novel antigen receptor (IgNAR) (Greenberg et al. 1995; Dooley et al. 2003). This new TCR form was named NAR-TCR, and, although yet to be proven, it is thought to pair with TCRγ like a conventional TCRδ. This creates an unusual predicted structure where the N-terminal V-NAR domain is extended in an unpaired fashion beyond the TCRγ V domain (Fig. 1c). The shark TCRδ locus encodes conventional TCRδ chains along with NAR-TCR (Criscitiello et al. 2006). The middle or supporting V domain in a NAR-TCR chain appears to be encoded by a Vδ gene that is conventional in sequence structure. These Vδ genes, however, lack the exon encoding the leader peptide. This is an informative change consistent with evolution of a specialized subset of Vδ encoding the supporting V domain; these Vδ will not be N-terminal in the final protein and do not need a leader peptide (Criscitiello et al. 2006).

To encode a complete NAR-TCR chain, multiple V(D)J recombination events need to occur successfully. A NAR-like V gene that is upstream of the locus needs to be recombined to D and J segments to encode the complete N-terminal V-NAR domain (Fig. 1c). In addition one of the specialized Vδ genes is recombined to downstream Dδ and Jδ segments to encode the supporting or second Vδ domain. Following transcription the V-NAR exon is spliced in the mRNA to the supporting Vδ exon, presumably as splice uncomplicated by the presence of an unnecessary leader exon (Criscitiello et al. 2006).

The discovery of an atypical three-Ig domain TCR in sharks was quickly followed by an analogous TCR chain in marsupials (Figs. 1d and 2). Marsupials, or metatherians, are one of the three living branches of the mammalian tree, which also includes prototherians (egg-laying monotremes) and eutherians (the so-called placental mammals like you, me, and the mouse) (Fig. 2). The first evidence that there may be something unusual going on in metatherians came from the observation that they had a second TCRδ locus (Baker et al. 2005). Further genomic and transcriptomic analyses, primarily using the gray short-tailed opossum as the model marsupial, revealed that this was not just a second TCRδ locus, but a highly divergent TCR with atypical features shared with NAR-TCR (Parra et al. 2007). The locus and the TCR chains they encode were named TCRμ (μ or M for marsupial), and they were found to contain a number of features that were at first for a TCR. It was apparent early on that the structural similarity which TCRμ had with NAR-TCR was the product of convergent evolution. The overall genomic organization and sequence structure were too dissimilar to be due to homology by dissent (Parra et al. 2007).

The marsupial TCRμ locus is arranged in a tandem array of clusters (Parra et al. 2008). Not all the clusters are complete. Those that are complete contain one Vμ, three to five Dμ, and one Jμ gene segment that are somatically recombined to create a diverse Vμ domain that is N-terminal in the mature chain. Downstream from these V, D, and J genes is a complete exon encoding a V domain. This exon does not require V(D)J recombination to be assembled. Rather the exon encodes framework (FR) 1 through 4 in its germ-line state (Parra et al. 2007). The exon and the protein domain it encodes have been labeled Vμj due to it appearing to be an already germ-line "joined" V gene (Parra et al. 2007). Insight into the origin of the Vμj exon comes from the lack of an intron separating sequence encoding a leader peptide and the start of the exon encoding the V domain. Vμj appears to have been generated by a retro-transposition event as it is a processed gene, one that lacks introns due to reverse transcription from a mature mRNA transcript and inserted into the germ-line genome by site-specific recombination (Parra et al. 2008). Both Vμ and Vμj gene segments are more related to antibody VH genes than to conventional TCR V genes, consistent with genes with the *IgH* locus being involved in the evolution of TCRμ (Parra et al. 2007).

To generate a full-length TCRμ chain, the Vμ, Dμ, and Jμ gene segments are somatically recombined. In the RNA transcript, the rearranged Vμ exon is spliced to the Vμj exon using a splice site between the leader and FR1 sequence (Parra et al. 2006). Only 60 % of thymus TCRμ transcripts have productive

rearrangements, whereas 100 % of splenic transcripts are productive. This supports the thymus being the site of TCRμ$^+$ T cell development and there is some as-yet unknown mechanism of selecting productive rearrangements prior to those cells entering the peripheral tissues (Parra et al. 2007). The region corresponding to the complementarity determining region 3 (CDR3) of Vμj is highly conserved across marsupial species, consistent with the exon being from a single retro-transcription event early in the evolution of the metatherian lineage (Parra et al. 2007). The Vμj CDR3 is also relatively short, likely due to the absence of D segments in the original recombination event. This hypothesis is supported by the genomic organization of the platypus TCRμ locus (Wang et al. 2011).

Relative to the conventional TCR chains, the constant (C) region of TCRμ is most related to TCRδ C regions, and TCRμ, like TCRδ, is predicted to pair with TCRγ (Parra et al. 2007). Recent analysis of opossum thymocytes has revealed that probes for TCRμ and TCRγ co-stain the same cells using fluorescence in situ hybridization (unpublished). If these initial observations hold up, a third lineage of T cells, the γμT cell, is added to the repertoire along with αβ and γδ T cells. The genomic organization of the opossum TCRγ genes is fairly conventional and all TCRγ transcripts characterized would encode a typical V and C two-domain structure (Fig. 1e). It is unlikely that there is a three-domain version of TCRγ in marsupials and, like NAR-TCR, the N-terminal Vμ domain is probably unpaired (Figs. 1d and 2). This prediction is supported by the observation that when a single Vμ domain is expressed as recombinant protein, it is soluble (unpublished observation). This result is consistent with the Vμ domain being normally in an unpaired state in the extracellular protein.

The discovery of TCRμ raised the possibility of a novel T cell subset that might be a unique adaptation to metatherian reproductive strategies. Metatherians give birth to highly altricial young and most immune system development occurs postnatally (Old and Deane 2000; Parra et al 2009; Wang et al 2012). An early working hypothesis for why there would be a novel TCR chain in metatherians (that is lost in eutherian mammals) is that it provided protection to early developing young. This hypothesis predicted that γμT cells would appear early in development. Unfortunately that was not the case. Rather, γμT cells are the last of the three subsets to appear in development. αβT cells are the first to be detected in neonatal opossums, followed by γδT cells, and later γμT cells when the neonates are almost 2 weeks' old (Parra et al. 2009).

4 The Duckbill Platypus: A "Missing Link" in TCR Evolution?

The Australian duckbill platypus, *Ornithorhynchus anatinus*, was thought to be a hoax when first brought to the attention of European science (Moyal 2001). An animal with the fur of a mammal and the bill and webbed feet of a duck stretched

credulity for early nineteenth-century naturalists. Once they were accepted as reality, platypuses continued to surprise. The Scottish zoologist William Hay Caldwell, while a Cambridge University student, alerted the scientific world in the 1880s with his tersely worded telegram to the Royal Society, "*Monotremes oviparous, ovum meroblastic*" (Caldwell 1887). He was telling the scientific community that monotremes were mammals that laid eggs and the eggs were more like that of reptiles than metatherian or eutherian mammals. Even more delightful, male platypuses are venomous, although the venom glands are only active during mating season, the purpose of the venom probably being to chase off rival males and control uncooperative females. Genomic analysis of the platypus has done nothing to quell a sense that the platypus has retained features from our reptile-like ancestors that have been lost in the marsupials and eutherian mammals. Monotreme sex chromosomes, of which there are 5 Xs and 5 Ys, share gene content with both the XY system of other mammals and the WZ system of birds (Grützner et al. 2004). The platypus *IgH* locus encodes a constant region isotype, designated IgO, that is similar to both mammalian IgG and avian/reptilian IgY while also containing the genes for true orthologues of mammalian IgG and IgE (Zhao et al. 2009).

Analyses of the platypus TCR loci have yielded a wealth of information used to develop a model for the evolution of the TCRμ locus in mammals. The platypus TCRα/TCRδ locus is the only known mammalian locus to contain VHδ genes. The platypus VHδ are recombined and used in the TCRδ repertoire (Parra et al. 2012a). Hence, the platypus can generate TCRδ chains resembling those of birds, frogs, and coelacanths, something marsupials and eutherian mammal are unable to do (Fig. 2). The platypus and echidna genomes also contain clear homologues of the TCRμ loci found in metatherians (Wang et al. 2011). In the platypus, where some genomic information is available, the *TCRμ* locus appears to have a cluster organization much like that of opossum *TCRμ*. As in metatherians, the genes encoding the N-terminal most V domain undergo V(D)J recombination and are highly diversified in the expressed repertoire. Unlike metatherians, the genes encoding the second V domain also require V(D)J recombination to be assembled from the germ-line state (Wang et al. 2011). Single V and J gene segments are recombined with little junctional diversity to produce a nearly invariant second V domain (Wang et al. 2011). The V and J segments encoding the second or supporting V domain (V2 in Fig. 1d) are closely related to those encoding the N-terminal V1 domain, suggesting that the second V evolved by internal gene duplication (Wang et al. 2011). There are no D segments among the gene segments encoding V2, consistent with their deletion early in the evolution of TCRμ (Wang et al. 2011).

In both the opossum and the platypus, the N-terminal TCRμ V domains are diverse primarily due to junctional diversity from V(D)J recombination. The second or supporting V domain lacks diversity in the case of the marsupial or has limited diversity in the case of the monotreme. Presumably, the supporting V domain does not require diversity for function since its role is to extend the length of the protein chain for the terminal V domain to be sticking out where it can contact antigen. In the marsupials, this limited diversity is maintained by using a germ-line joined V domain with no recombination required. In the platypus, it is maintained by having

deleted the D segments and only recombining a V to a J to encode a CDR3 of very limited diversity (Wang et al. 2011).

The discovery that the platypus has both VHδ gene segments in its TCRα/TCRδ locus and the genes for TCRμ provided the opportunity to build a model of the evolution of these loci. It is clear that the monotreme and marsupial TCRμ loci are orthologous and are products of a single evolutionary origin (Wang et al. 2011). The platypus Vμ gene segments are most related to the platypus VHδ genes found in the *TCRα/TCRδ* locus (Parra et al. 2012a). The most parsimonious explanation for the origin of TCRμ appears to be a duplication of a cluster of genes containing VHδ–Dδ–Jδ–Cδ to a genomic location outside the *TCRα/TCRδ* locus, followed by divergence of these genes to become TCRμ (Parra et al. 2012a). This evolutionary model was less clear from analysis of metatherian genomes alone as they lacked the VHδ genes for comparison. The ever-remarkable platypus provided the "missing link" in TCRδ evolution (Parra et al. 2012a).

5 What Is the Function of γμT Cells and Why Were They Lost in Eutherians?

The current model or expectation of the function of the γμ TCR is that, like conventional TCR, it is an antigen recognition receptor. This hypothesis is supported by the N-terminal domain undergoing somatic diversification through V(D)J recombination like a conventional Ig or TCR. One current hypothesis is that the TCRμ chain binds antigen directly, much like an antibody, and does this using the N-terminal V domain as a single binding domain. This hypothesis is supported by two observations. First, the Vμ domains are more antibody-like than TCR-like. Similar to antibodies, they may bind native antigens or epitopes directly. Second, the best model for antigen binding by TCRμ may be the heavy-chain-only antibodies such as IgNAR and forms of IgG found in camelid species (Flajnik et al. 2011; Hamers-Casterman et al. 1993). It has already been speculated that shark NAR-TCR most likely binds antigen similarly to IgNAR and camel IgG, with the unpaired domain acting as a single antigen binding domain (Flajnik et al. 2011). This model seems logical for TCRμ as well.

The functional phenotype of γμT cells requires further analysis and many questions remain. Amino acid residues in the transmembrane and cytoplasmic regions that interact with CD3 chains are conserved in TCRμ consistent with much of the signaling machinery being similar to conventional T cells (Parra et al. 2007). Whether γμT cells are using CD4 or CD8 is unknown, as is their effector function.

The phylogenetic relationship among the three living mammalian lineages is no longer a matter of debate (Huxley 1880; Killian et al. 2001; Baker et al. 2004). The marsupials and eutherians sharing a more recent common ancestor over 140 million years ago and the monotremes splitting off earlier over 160 million years ago, as is shown in Fig. 2, are the broadly accepted relationship (Bininda-Emonds

et al. 2007). The presence of TCRμ in both marsupials and monotremes indicates that this locus evolved early in mammals and would have been present in the last common ancestor of marsupials and eutherians. Therefore, the eutherians lost TCRμ. The immunological consequences of having lost TCRμ will probably not be understood until there is better understanding of the role γμT cells play in the immune systems of prototherians and metatherians.

How TCRμ was lost in eutherians may have a relatively simple explanation. The genomic region where TCRμ is located in the opossum is one where there is little conserved synteny with other species (Parra et al. 2008). The disruption of syntenic blocks in this region not only exists between marsupials and eutherians but even among eutherians such as humans and mice. This is probably indicative of a region of the genome that has undergone significant rearrangement during speciation, a process that can facilitate genes loss.

6 Summary

The evolutionary story of the vertebrate T cell receptors is one of conservation. While immunoglobulins have continued to evolve new isotypes and new roles, the TCRs have remained relatively stable with all known gnathostomes having clearly homologous TCRα, TCRβ, TCRγ, and TCRδ chains (Rast et al. 1997). There is strong evidence that the genomic organization of these genes has remained relatively stable (Chen et al. 2009). This is also a story of evolutionary plasticity, at least for one of the TCR chains, TCRδ. TCRδ has learned a lot of tricks to increase its diversity over the last 400 million years. TCRδ has spun off some interesting alternatives, such as TCRμ, while also retaining its basic structure and roles. NAR-TCR and TCRμ appear to have provided T cells with an alternative mechanism of antigen recognition. What roles these T cells may be performing remains to be determined. Whether or not γδ T cells using VHδ genes are performing a distinct role, or if this is simply a way of adding diversity to the TCR repertoire, also remains to be determined.

Acknowledgments The authors were supported in part by National Institutes of Health Institutional Development Award program award P30GM110907 and National Science Foundation Award IOS-1353123.

References

Baker ML, Wares JP, Harrison GA et al (2004) Relationships among the families and orders of marsupials and the major mammalian lineages based on recombination activating gene-1. J Mamm Evol 11(1):1–16

Baker ML, Osterman AK, Brumburgh S (2005) Divergent T-cell receptor delta chains from marsupials. Immunogenetics 57(9):665–673

Bininda-Emonds OR, Cardillo M, Jones KE et al (2007) The delayed rise of present-day mammals. Nature 446(7135):507–512

Brenner MB, McLean J, Dislynas DP et al (1986) Identification of a putative second T-cell receptor. Nature 322(6075):145–149

Caldwell H (1887) The embryology of Monotremata and Marsupialia. Phil Trans R Soc Land B 178:463–486

Chen H, Kshirsagar S, Jensen I et al (2009) Characterization of arrangement and expression of the T cell receptor γ locus in the sandbar shark. PNAS 106(21):8591–8596

Criscitiello MF, de Figueiredo P (2013) Fifty shades of immune defense. PLoS Pathog 9(2), e1003110

Criscitiello MF, Saltis M, Flajnik MF (2006) An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TCR genes in sharks. PNAS 103(13):5036–5041

Criscitiello MF, Ohta Y, Saltis M et al (2010) Evolutionarily conserved TCR binding sites, identification of T cells in primary lymphoid tissues, and surprising trans-rearrangements in nurse shark. J Immunol 184(12):6950–6960

Danilova N, Bussmann J, Jekosch K et al (2005) The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat Immunol 6(3):295–302

Dooley H, Flajnik MF, Porter AJ (2003) Selection and characterization of naturally occurring single-domain (IgNAR) antibody fragments from immunized sharks by phage display. Mol Immunol 40(1):25–33

Fischer C, Bouneau L, Ozouf-Costaz C et al (2002) Conservation of the T-cell receptor α/δ linkage in the teleost fish *Tetraodon nigroviridis*. Genomics 79(2):241–248

Flajnik MF, Deschacht N, Muyldermans S et al (2011) A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? PLoS Biol 9(8), e1001120

Greenberg AS, Avila D, Hughes M et al (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374(6518):168–173

Grützner F, Rens W, Tsend-Ayush E et al (2004) In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. Nature 432 (7019):913–917

Guo P, Hirano M, Herrin BR et al (2009) Dual nature of the adaptive immune system in lampreys. Nature 459(7248):796–801

Hamers-Casterman C, Atarhouch T, Muyldermans S et al (1993) Naturally occurring antibodies devoid of light chains. Nature 363:446–448

Hansen JD, Landis ED, Phillips RB (2005) Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: implication for a distinctive B cell developmental pathway in teleost fish. PNAS 102(19):6919–6924

Havran WL, Allison JP (1988) Developmentally ordered appearance of thymocytes expressing different T-cell antigen receptors. Nature 335(6189):443–445

Hedrick SM, Cohen DI, Nielsen EA et al (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature 308(5955):149–153

Hinds KR, Litman GW (1986) Major reorganization of immunoglobulin VH segmental elements during vertebrate evolution. Nature 320:546–549

Hsu E, Flajnik MF, DuPasquier L (1985) A third immunoglobulin class in amphibians. J Immunol 135(3):1998–2004

Huxley TH (1880) On the application of laws of evolution to the arrangement of the Vertebrata, and more particularly of the Mammalia. Proc Zool Soc Lond 43:649–662

Isobe M, Russo G, Haluska FG et al (1988) Cloning of the gene encoding the δ subunit of the human T-cell receptor reveals its physical organization within the α-subunit locus and its involvement in chromosome translocations in T-cell malignancy. PNAS 85(11):3933–3937

Kasamatsu J, Sutoh Y, Fugo K et al (2010) Identification of a third variable lymphocyte receptor in the lamprey. PNAS 107(32):14304–14308

Killian JK, Buckley TR, Stewart N et al (2001) Marsupials and eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution. Mamm Genome 12(7):513–517

Loh EY, Lanier LL, Turck CW et al (1987) Identification and sequence of a fourth human T cell antigen receptor chain. Nature 330:569–572
Marchalonis JJ, Schluter SF, Edmundson AB (1997) The T-cell receptor as immunoglobulin: paradigm regained. Exp Biol Med 216(3):303–318
Moyal A (2001) Platypus: the extraordinary story of how a curious creature baffled the world. Smithsonian Books, Washington, DC
Mußmann R, DuPasquier L, Hsu E (1996) Is *Xenopus* IgX an analog of IgA? Eur J Immunol 26 (12):2823–2830
Nam B, Hirono I, Aoki T (2003) The four TCR genes of teleost fish: the cDNA and genomic DNA analysis of Japanese flounder (*Paralichthys olivaceus*) TCR α-, β-, γ-, and δ-chains. J Immunol 170(6):3081–3090
Ohta Y, Flajnik MF (2006) IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. PNAS 103(28):10723–10728
Olaru A, Petrie HT, Livák F (2005) Beyond the 12/23 rule of VDJ recombination independent of the Rag proteins. J Immunol 174(10):6220–6226
Old JM, Deane EM (2000) Development of the immune system and immunological protection in marsupial pouch young. Dev Comp Immunol 24:445–454
Ota T, Rast JP, Litman GW et al (2003) Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. PNAS 100 (5):2501–2506
Owen FL, Riblet R (1984) Genes for the mouse T cell alloantigens Tpre, Tthy, Tind, and Tsu are closely linked near Igh on chromosome 12. J Exp Med 159(1):313–317
Pancer Z, Amemiya CT, Ehrhardt GRA et al (2004) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 430(6996):174–180
Parra ZE, Miller RD (2012) Comparative analysis of the chicken TCRα/δ locus. Immunogenetics 64:641–645
Parra ZE, Arnold T, Nowak MA et al (2006) TCR gamma chain diversity in the spleen of the duckbill platypus (*Ornithorhynchus anatinus*). Dev Comp Immunol 30(8):699–710
Parra ZE, Baker ML, Schwarz R et al (2007) A unique T cell receptor discovered in marsupials. PNAS 104(23):9776–9781
Parra ZE, Baker ML, Hathaway J et al (2008) Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. BMC Genomics 9(1):111
Parra ZE, Baker ML, Lopez AM et al (2009) TCRμ recombination and transcription relative to the conventional TCR during postnatal development in opossums. J Immunol 182(1):154–163
Parra ZE, Ohta Y, Criscitiello MF et al (2010) The dynamic TCRδ: The TCRδ chains in the amphibian *Xenopus tropicalis* utilize antibody-like V genes. Eur J Immunol 40(8):2319–2329
Parra ZE, Lillie M, Miller RD (2012a) A model for the evolution of the mammalian T-cell receptor α/δ and μ loci based on evidence from the duckbill platypus. Mol Biol Evol 29(10):3205–3214
Parra ZE, Mitchell K, Dalloul RA et al (2012b) A second TCRδ locus in Galliformes uses antibody-like V domains: Insight into the evolution of TCRδ and TCRμ genes in tetrapods. J Immunol 188(8):3912–3919
Pellicci DG, Uldrich AP, LeNours J et al (2014) The molecular bases of δ/αβ T cell-mediated antigen recognition. J Exp Med 211(13):2599–2615
Porter RR (1973) Structural studies of immunoglobulins. Science 180(4087):713–716
Raff MC (1970) Two distinct populations of peripheral lymphocytes in mice distinguishable by immunofluorescence. Immunology 19(4):637–650
Rast JP, Haire RN, Litman RT et al (1995) Identification and characterization of T-cell antigen receptor-related genes in phylogenetically diverse vertebrate species. Immunogenetics 42 (3):204–212
Rast JP, Anderson MK, Strong SJ et al (1997) α, β, γ, and δ T cell antigen receptor genes arose early in vertebrate phylogeny. Immunity 6(1):1–11
Saha NR, Ota T, Litman GW (2014) Genome complexity in the coelacanth is reflected in its adaptive immune system. J Exp Zool 322(6):438–463

Saito H, Kranz DM, Takagaki Y et al (1984) Complete primary structure of a heterodimeric T-cell receptor deduced from cDNA sequences. Nature 309(5971):757–762

von Behring E (1901) Serum therapy in therapeutics and medical science. Nobel Lectures Physiology or Medicine 1901–1921. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1901/behring-lecture.html

Waldmann TA, Davis MM, Bongiovanni KF et al (1985) Rearrangements of genes for the antigen receptor on T cells as markers of lineage and clonality in human lymphoid neoplasms. N Engl J Med 313(13):776–783

Wang X, Parra ZE, Miller RD (2011) Platypus TCRμ provides insight into the origins and evolution of a uniquely mammalian TCR locus. J Immunol 187(10):5246–5254

Wang X, Sharp AR, Miller RD (2012) Early postnatal B cell ontogeny and antibody repertoire maturation in the opossum, *Monodelphis domestica*. PLoS One 7(9), e45931

Warr GW, Magor KE, Higgins DA (1995) IgY: clues to the origins of modern antibodies. Immunol Today 16(8):392–398

Zhang Y, Salinas I, Li J et al (2010) IgT, a primitive immunoglobulin class specialized in mucosal immunity. Nat Immunol 11(9):827–835

Zhao Y, Cui H, Whittington CM et al (2009) *Ornithorhynchus anatinus* (platypus) links the evolution of immunoglobulin genes in eutherian mammals and nonmammalian tetrapods. J Immunol 183(5):3285–3293

Diversification of the Primary Antibody Repertoire by AID-Mediated Gene Conversion

Dennis K. Lanning and Katherine L. Knight

Abstract Gene conversion, mediated by activation-induced cytidine deaminase (AID), has been found to contribute to generation of the primary antibody repertoire in several vertebrate species. Generation of the primary antibody repertoire by gene conversion of immunoglobulin (Ig) genes occurs primarily in gut-associated lymphoid tissues (GALT) and is best described in chicken and rabbit. Here, we discuss current knowledge of the mechanism of gene conversion as well as the contribution of the microbiota in promoting gene conversion of Ig genes. Finally, we propose that the antibody diversification strategy used in GALT species, such as chicken and rabbit, is conserved in a subset of human and mouse B cells.

1 Introduction

Activation-induced cytidine deaminase (AID) catalyzes three important immunological processes—somatic hypermutation, class-switch recombination, and somatic gene conversion (Muramatsu et al. 2000; Arakawa et al. 2002). Although somatic hypermutation and class-switch recombination are widely utilized among vertebrate species, somatic gene conversion has been observed, as an immunological mechanism, primarily in vertebrates that diversify their primary antibody (Ab) repertoires in gut-associated lymphoid tissues (GALT). These species generate their primary antibody repertoires through a fundamentally different strategy than that used by mice and humans (Fig. 1). B cells of mice and humans, for example, generate a highly diverse antibody repertoire by utilizing many different V, (D), and J gene segments during V(D)J gene rearrangement of the immunoglobulin (Ig) heavy and light chain loci. In contrast, B cells of vertebrates that diversify their primary antibody repertoires in GALT use V(D)J gene recombination

D.K. Lanning • K.L. Knight (✉)
Department of Microbiology and Immunology, Loyola University Chicago, 2160 S. First Avenue, Maywood, IL 60153, USA
e-mail: kknight@luc.edu

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_12

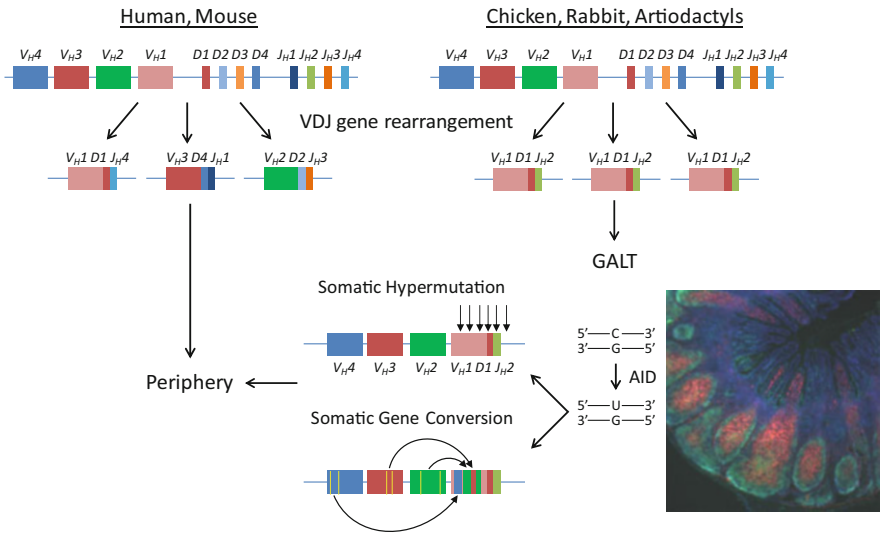


Fig. 1 Two strategies for generating a diverse primary antibody repertoire. B cells of humans and mice (*left*) utilize many different V, (D), and J gene segments during V(D)J gene rearrangement of the immunoglobulin (Ig) heavy and light chain loci (Ig heavy chain rearrangement is shown). A diverse primary antibody repertoire is thus generated directly through V(D)J gene rearrangement, and the B cells exit the bone marrow and enter the periphery. In contrast, B cells of chickens, rabbits, and artiodactyls (*right*) preferentially utilize a small subset of V (and often (D) and J) gene segments during V(D)J gene rearrangement and thus generate much less initial antibody diversity. Upon exiting the bone marrow, B cells migrate to gut-associated lymphoid tissues (GALT), where they undergo proliferation and mutate their V(D)J genes through somatic gene conversion and somatic hypermutation. A tissue section from rabbit appendix, the major site of rabbit GALT, is shown, with proliferating B cells stained green and non-proliferating B cells stained red. V(D)J gene mutation is initiated by AID-mediated deamination of cytidines, and the resulting lesions (G/U mismatches) induce somatic hypermutation or somatic gene conversion. Somatic hypermutation introduces point mutations (*arrows*) throughout the V(D)J gene and into the J–C intron. Somatic gene conversion transfers nucleotide tracts (bounded by *yellow lines*) from upstream V donors into the rearranged V gene. A highly diverse primary antibody repertoire is generated and the B cells exit GALT and enter the periphery

primarily for assembling functional Ig heavy and light chain genes, rather than for generating antibody diversity. B cells in these species preferentially utilize a small subset of V (and often (D) and J) gene segments during V(D)J gene rearrangement, and the resulting V(D)J genes serve as mutation substrates during a subsequent phase of antibody repertoire diversification in GALT. During repertoire diversification, mutations are introduced into the rearranged V(D)J genes through somatic gene conversion and somatic hypermutation. This process generates an extremely large array of different antibody specificities, collectively termed the primary antibody repertoire, which serves to anticipate and protect against a vast range of potential pathogens.

2 Prevalence of AID-Mediated Antibody Repertoire Diversification Among Vertebrates

This strategy for generating a diverse primary antibody repertoire was first recognized in the chicken, which has only one functional V gene at both the Ig heavy and light chain loci and uses gene conversion to diversify its primary antibody repertoire in a gut-associated lymphoid tissue called the bursa of Fabricius (Reynaud et al. 1985, 1987). Shortly thereafter, the Ig light chain loci of the mallard duck, turkey, quail, pigeon, hawk, and cormorant were found to preferentially utilize a single V_L gene segment during V–J recombination, suggesting that these avian species also use gene conversion to diversify their primary antibody repertoires (McCormack et al. 1989a). While this strategy was briefly thought unique to birds, it was subsequently found that rabbits (Knight and Becker 1990), swine (Sun et al. 1998), and cattle (Parng et al. 1996) also generate limited antibody diversity during V(D)J gene recombination due to preferential rearrangement of a few V gene segments. In these species, Ig gene diversification also occurs in GALT, primarily the appendix in rabbits and the ileal Peyer's patch in cattle. It is not clear where Ig gene diversification occurs in swine, as it has recently been shown not to occur in the ileal Peyer's patch (Butler et al. 2011). These tissues contain thousands of lymphoid follicles where B cells proliferate and diversify their Ig genes and are considered mammalian equivalents of the bursa of Fabricius.

3 Somatic Gene Conversion vs. Somatic Hypermutation

Although somatic hypermutation and somatic gene conversion are both mediated by AID, they are fundamentally different mutational processes. Both processes are initiated by AID-mediated deamination of cytidine residues, particularly at RGYW consensus hotspots, in V(D)J genes, but they represent two different pathways of resolving this initial DNA lesion. Somatic hypermutation introduces point mutations through error-prone repair of the resulting uracil residues (Di Noia and Neuberger 2007). While primarily known for driving affinity maturation in germinal centers during primary immune responses, somatic hypermutation is also utilized, in an antigen-independent manner, to diversify the primary antibody repertoire in some species, e.g., the rabbit. Somatic gene conversion, rather than introducing point mutations, transfers nucleotide tracts from upstream donor V gene segments into the rearranged V(D)J gene through nonreciprocal homologous recombination. The unidirectionality of the sequence exchange, leaving the donor sequence unaltered, distinguishes somatic gene conversion from double homologous recombination. During diversification of the primary antibody repertoire, a single round of gene conversion can introduce multiple nucleotide substitutions, as well as codon insertions and/or deletions, and multiple rounds of gene conversion can occur within a given V(D)J gene.

4 Gene Conversion-Mediated Antibody Repertoire Diversification in the Chicken

Somatic gene conversion was first recognized as a mechanism for antibody repertoire diversification in the chicken by Reynaud et al. (1987). The chicken Ig heavy and light chain loci each contain only one functional V gene segment (V_H1 and V_L1, respectively), as well as a single functional J gene segment (J_H and J_L) (Reynaud et al. 1985, 1989). In addition, only limited junctional diversity is introduced by V(D)J gene rearrangement, which does not continue throughout life, but solely during a brief period of early embryonic development (Weill et al. 1986; McCormack et al. 1989b). Thus, unlike humans and mice, little antibody diversity is generated during rearrangement of the chicken Ig loci. The majority of repertoire diversity is generated subsequently in the bursa of Fabricius, beginning between days 15 and 18 of embryonic development, by gene conversion-mediated replacement of nucleotide tracts in the rearranged V_H1 and V_L1 gene segments (Reynaud et al. 1987; Thompson and Neiman 1987).

The chicken Ig light chain locus contains 25 V_L pseudogene (ψV_L) segments in the 19 kb region upstream of the V_L1 and J_L gene segments (Reynaud et al. 1987, 1989). Reynaud et al. (1985, 1987) first observed that tracts of nucleotide substitutions within diversified V_L1 cDNA sequences corresponded to nucleotide sequences present in the ψV_L gene segments. The chicken Ig heavy chain locus contains around 80 ψV_H gene segments, similar to V–D joints, in a 60–80 kb region upstream of V_H1, about 15 D_H gene segments and a single J_H gene segment (Reynaud et al. 1989). Reynaud et al. (1989) similarly found tracts of nucleotide substitutions in diversified V_H1 cDNA sequences that corresponded to nucleotide sequences in ψVH gene segments (Reynaud et al. 1989). All ψV_L and ψV_H gene segments lack a promoter, leader exon, or V(D)J recombination signals. Few of them contain stop codons or frameshift mutations, but many are truncated at their 5′ or 3′ ends. Carlson et al. (1990) demonstrated that the diversifying nucleotide tracts in the chicken Ig light chain gene result from intrachromosomal gene conversion. By restriction mapping donor ψV_L gene segments and recipient V_L1 genes in a panel of v-*rel*-transformed chicken B cell lines and sequencing donor ψV_L gene segments used in V_L1 nucleotide substitutions, these authors showed that donor ψV_L gene segments were not modified during the sequence exchange. The authors further showed, by means of allelic V_L1 and ψV_L gene segment polymorphisms in the SC chicken strain (an F_1 cross between the G4 and S3 inbred strains), that gene conversion uses donor and recipient V_L gene segments from the same allele and thus from the same chromosome.

5 Gene Conversion-Mediated Antibody Repertoire Diversification in the Rabbit

The recognition of preferential V_H gene segment usage in the rabbit, first reported by Knight and Becker (1990), provided a solution to the long-standing problem of allelically inherited rabbit V_H allotypes (Knight and Becker 1990). The presence in rabbits of V_H allotypic markers, inherited in a simple Mendelian fashion, on 80–90 % of serum Ig molecules had been difficult to explain under the assumption that multiple V_H gene segments were utilized during rearrangement of the heavy chain locus. This problem was resolved with the recognition that rabbit B cells, like those of chickens, preferentially utilize the 3′-most V_H gene segment during V(D)J gene rearrangement and generate antibody repertoire diversity through subsequent mutation of the rearranged *IgH* genes (Knight and Becker 1990; Becker and Knight 1990).

While rabbits and chickens use the same strategy to generate a diverse primary antibody repertoire, there are some interesting differences between the two species. Rabbits, for example, generate a more diverse range of Ig genes during the initial V(D)J gene rearrangements. Unlike the chicken Ig heavy chain locus, many of the upstream rabbit V_H gene segments are potentially functional, and a small number of them are used in 10–20 % of VDJ gene rearrangements (Friedman et al. 1994). In contrast to the single chicken J_H gene segment, the rabbit heavy chain locus contains five functional J_H gene segments, although J_H4 is preferentially used in VDJ gene rearrangements (Becker et al. 1989; Lavinder et al. 2014). Also unlike chickens, multiple *Igκ* gene segments are utilized during rearrangement of the rabbit Ig light chain locus, thus contributing significant diversity to the initial antibody repertoire (Sehgal et al. 1999; Lavinder et al. 2014). Although the rabbit light chain locus contains three functional J_L gene segments, one of these, IGKJ1_2, is used nearly exclusively in Ig light chain gene rearrangements (Lavinder et al. 2014). Interestingly, in both chickens and rabbits, gene conversion also contributes to Ig gene mutation in germinal centers during antigen-specific immune responses, in addition to its role in diversifying the primary antibody repertoire (Winstead et al. 1999; Arakawa et al. 1996).

While chickens rely exclusively on gene conversion to diversify the primary antibody repertoire, rabbits use both gene conversion and somatic hypermutation (Becker and Knight 1990; Weinstein et al. 1994). As in chickens, rabbit V(D)J genes are diversified in GALT, primarily the appendix in rabbits (Weinstein et al. 1994; Vajdy et al. 1998). In chickens, B cells seed the bursa of Fabricius and begin diversifying their V(D)J genes shortly before hatch (Reynaud et al. 1994). Ig gene diversification in the chicken thus begins as a developmentally programmed event that occurs in the absence of exogenous antigen, as well as in experimentally manipulated B cells expressing truncated B cell receptors that lack the antigen-binding region (Sayegh et al. 1999). In contrast, rabbits begin diversifying their primary antibody repertoire shortly after birth. The differential timing of the onset of repertoire diversification in rabbits and chickens might reflect their differing

rates of embryonic development. While newly hatched chicks are well-developed precocial offspring, rabbit pups are altricial offspring requiring maternal care for the first 2–3 weeks of life.

Rabbits further differ from chickens in requiring select members of the intestinal microbiota to initiate the diversification process (Zhai and Lanning 2013; Rhee et al. 2004). Rhee et al. (2004) demonstrated that maintaining the rabbit appendix as a germ-free tissue by ligating it at birth to prevent microbial colonization prevented V(D)J gene diversification. These authors further found that V(D)J gene diversification could be induced by colonizing germ-free ligated appendix with colony-purified isolates from the rabbit intestinal microbiota. Surprisingly, only co-colonization with *Bacillus subtilis* and *Bacteroides fragilis* induced V(D)J gene diversification, despite each of several tested isolates attaining similar colonization densities (Rhee et al. 2004). Rabbit B cells thus require select members of the intestinal microbiota to initiate V(D)J gene diversification in GALT.

As in chickens, repertoire diversification in rabbits is initiated in an antigen-independent manner. The diversification patterns in appendix B cell Ig genes, for example, differ strikingly from those driven by an immunizing antigen in splenic germinal centers (Sehgal et al. 2002). Severson et al. (2010) reported evidence that bacterial superantigen-like molecules drive repertoire diversification in rabbit GALT. These authors found that spores from *Bacillus subtilis* and other *Bacillus* species bind a superantigen-like binding site on rabbit IgM and that surface molecules from *B. anthracis* stimulate B cell proliferation in rabbit GALT. Furthermore, Rhee et al. (2005) identified a putative superantigen-like binding site on rabbit V_Ha allotype Ig by comparing the amino acid sequences of the V_Ha and V_Hn allotypes. V_H1, the preferentially utilized V_H gene segment, encodes the V_Ha allotype, while the less frequently utilized V_H gene segments primarily encode the V_Hn allotype. Mutant *ali/ali* rabbits lack V_H1, and, as a result, V_Hn B cells comprise the vast majority of peripheral and GALT B cells during the first 6 weeks of life (Pospisil et al. 1995). Between 6 and 11 weeks of age, however, the small V_Ha B cell population expands rapidly and becomes the dominant B cell type. Rhee et al. (2005) demonstrated that this shift from V_Hn to V_Ha B cell dominance is driven by the intestinal microbiota in GALT, because it did not occur when the appendix was ligated to prevent microbial colonization and all other organized GALT was removed from newborn *ali/ali* pups. Taken together, these data suggest that bacterial superantigen-like molecules induce repertoire diversification in rabbit GALT by polyclonally activating B cells that express V_Ha allotype B cell receptors.

6 Determinants of Gene Conversion Donor Usage

The chicken Ig light chain locus has proven useful for determining some of the molecular requirements for gene conversion, because it is compact and the nucleotide sequences of all 25 of its ψV_L gene segments are known. Efficient intrachromosomal gene conversion in mammalian cells requires at least 200–300 bp of homologous nucleotide sequence and an overall nucleotide sequence

identity of >80 % (Liskay et al. 1987; Waldman and Liskay 1987). In the chicken IgL locus, the frequency with which ψV_L gene segments are used as gene conversion donors is similarly influenced by the extent of their homology with V_L1 (McCormack and Thompson 1990). This is dependent on both their percent nucleotide sequence identity with V_L1 and their length. Truncated ψV_L gene segments, for example, are used less frequently than full-length ψV_L gene segments (McCormack and Thompson 1990). Proximity to the rearranged V_L1–J_L gene also influences the frequency of ψV_L gene segment use, with those located most proximal utilized more frequently than those located more distally (Reynaud et al. 1987; McCormack and Thompson 1990). A third determinant of ψV_L gene segment usage is relative orientation with respect to the V_L1–J_L gene, those in inverted or antisense orientation being preferentially utilized. Gene conversion tracts are not found in the 5′ region of V_L1, despite the presence of ψV_L gene segments with homology to this region (Reynaud et al. 1987), but are found throughout the remainder of V_L1, with particularly high frequency in CDR1, the FR2/CDR2 boundary, and CDR3 (McCormack and Thompson 1990). Gene conversion tracts range in length from 8 bp to around 200 bp (McCormack and Thompson 1990). Although the 5′ ends of gene conversion tracts always begin in regions of homology between the ψV_L gene segment and V_L1, the 3′ ends can occur in nonhomologous regions and frequently contain nucleotide insertions or deletions, suggesting that the gene conversion mechanism operates with a 5′ to 3′ polarity (McCormack and Thompson 1990).

7 The Chicken DT40 Cell Line: A Model for Studying Ig Gene Conversion

The avian leukemia virus (ALV)-induced chicken B cell lymphoma line DT40 has provided additional insights into the molecular mechanism of gene conversion (Arakawa and Buerstedde 2004). DT40 appears to be arrested at the stage of bursal B cells and continues to undergo gene conversion of the Ig light chain gene during in vitro culture (Buerstedde et al. 1990; Kim et al. 1990). It also integrates transfected gene constructs highly efficiently into essentially any gene locus, regardless of transcriptional activity (Buerstedde and Takeda 1991). Although such highly efficient targeted integration is not seen in human or murine B cell lines or chicken non-B cell lines, it is seen in other chicken B cell lines that do not undergo Ig gene conversion (Buerstedde and Takeda 1991). This suggests that targeted integration in chicken B cell lines is mediated by a homologous recombination activity that supports, but is not sufficient for, gene conversion. Studies of DT40 attribute this supportive role to the RAD52 pathway.

8 The Role of the RAD52 Pathway in Gene Conversion

The RAD52 pathway mediates double-strand break (DSB) repair by homologous recombination and is required for both gene conversion and targeted integration in the yeast *Saccharomyces cerevisiae*. It is comprised of proteins that recognize double-strand breaks (RAD50, MRE11, and XRS2) and proteins that promote homology searches and strand invasion (RAD51, RAD52, RAD54, RAD55, and RAD57). Homologues of the genes encoding several of these yeast proteins have been cloned from chicken bursal cells, and their roles in targeted integration and gene conversion have been studied by disrupting them in DT40 cells (Arakawa and Buerstedde 2004). *RAD54*-deficient DT40 cells, for example, are highly sensitive to DNA damage and exhibit a 100-fold decrease in targeted integration efficiency, as well as reduced Ig light chain gene conversion activity (Bezzubova et al. 1997). DT40 cells deficient in *NBS1*, the vertebrate homologue of the yeast *XRS2* gene, exhibit a similar phenotype (Tauchi et al. 2002). Five vertebrate genes (*RAD51B*, *RAD51C*, *RAD51D*, *XRCC2*, and *XRCC3*) are considered *Rad51* paralogs on the basis of sequence similarity with the yeast *RAD51* gene. The RAD51 proteins are structural homologues of the bacterial DNA repair protein, recA (Kawabata et al. 2005). Disruption of each of these individually in DT40 cells results in reduced targeted integration efficiency and deficiencies in DSB repair, while disruption of all five loci is lethal (Takata et al. 2000, 2001; Sonoda et al. 1998). Although *S. cerevisiae RAD52* mutants exhibit severe recombination and DNA repair defects, DNA repair is normal and homologous recombination only moderately reduced in *RAD52*-deficient DT40 cells. The function of RAD52, however, appears to partially overlap that of XRCC3 in chicken B cells because, while disruption of either results in comparatively mild deficiencies, disruption of both leads to chromosome instability and cell death (Fujimori et al. 2001).

These studies suggest that gene conversion is dependent on the more general processes of homologous gene targeting and recombinational repair mediated by the RAD52 pathway. Indeed, co-localization of RAD51D and XRCC2 has been directly observed in diversifying Ig λ_L genes in DT40 cells, and ectopic expression of either accelerates the rate of gene conversion and influences the length of gene conversion tracts (Ordinario et al. 2009). RAD51 is also highly expressed in the rabbit appendix, which, like the chicken bursa, is a site of gene conversion-mediated antibody diversification (Barrington et al. 1999; Schiaffella et al. 1998). While establishing that the homologous recombination activity of the RAD52 pathway is necessary for efficient Ig gene conversion, these studies do not explain how gene conversion is initiated and specifically targeted to the Ig loci. The resolution of this question came with the surprising discovery that gene conversion of Ig genes requires activation-induced cytidine deaminase (AID), a protein previously identified as essential for both somatic hypermutation and class-switch recombination.

9 AID Is Required for Gene Conversion of Ig Genes

AID was first identified as a protein specifically upregulated in germinal center B cells that shares homology with the apolipoprotein B mRNA-editing enzyme, APOBEC-1 (Muramatsu et al. 1999). It was subsequently found to be required for both somatic hypermutation and class-switch recombination of Ig genes (Muramatsu et al. 1999, 2000; Revy et al. 2000). Shortly thereafter, disruption of the *AID* locus in chicken DT40 cells was found to cause complete loss of Ig gene conversion activity (Arakawa et al. 2002; Harris et al. 2002). Thus, remarkably, AID was identified as a master regulator of all three B cell-specific modifications of rearranged Ig genes. It is currently thought that all three modifications are initiated by AID-mediated deamination of cytidines and that the manner in which the resulting lesions (G/U mismatches) are resolved determines whether somatic hypermutation, gene conversion, or class-switch recombination follows. Inhibition of uracil-DNA glycosylase, for example, reduces gene conversion and activates somatic hypermutation in DT40 cells (Di Noia and Neuberger 2002). This observation demonstrates that gene conversion is favored by uracil excision of AID-induced G/U mismatches, rather than by their recognition by the mismatch repair complex. It further suggests that gene conversion might be facilitated by a DNA strand break generated by an apyrimidinic endonuclease acting on an abasic site (Di Noia and Neuberger 2002).

Impairment of homologous recombination in DT40 cells can also reduce gene conversion activity and activate somatic hypermutation of Ig genes. Disrupting the *RAD51* paralogs *RAD51B*, *XRCC2*, or *XRCC3*, for example, shifts the pattern of Ig V region mutations from gene conversion tracts to frequent point mutations localized preferentially at G/C base pairs, especially at RGYW motifs, known as somatic hypermutation hotspots (Sale et al. 2001; Rogozin and Kolchanov 1992). This observation supports the idea that gene conversion and somatic hypermutation represent distinct pathways for resolving a common AID-mediated lesion in the Ig V gene and suggests that recombination-mediated repair influences the choice of pathway. Furthermore, preventing Ig gene conversion by deleting the upstream ψV_L gene segments in DT40 cells induces somatic hypermutation of the *IgL* gene (Arakawa et al. 2004). As in the *RAD51* paralog mutants, point mutations are found primarily at G/C base pairs in RGYW hotspots. The mutations occur between 150 and 500 bp downstream of the Ig light chain promoter and are dependent on *AID* expression. The point mutations observed in DT40 *RAD51*- and ψV_L-deletion mutants are thus similar to those observed in human and mouse germinal center B cells in their distribution with respect to the Ig light chain promoter, dependence on AID, preference for RGYW hotspots, and restriction to the Ig locus. They do differ, however, in preferentially occurring at G/C base pairs and consisting primarily of G to C and C to G transversions, whereas somatic hypermutation in germinal center B cells targets G/C and A/T base pairs with similar frequency and exhibits a slight bias for transitions.

10 A Model for the Regulation of Gene Conversion and Somatic Hypermutation

Arakawa and Buerstedde (2009) incorporated many of the insights gained from DT40 studies into a model explaining the initiation and regulation of somatic hypermutation and gene conversion of Ig genes. In their model, both are initiated by a common DNA lesion, AID-mediated cytidine deamination, within the Ig gene. In the absence of nearby homologous donors or high homologous recombination activity, the resulting G/U mismatch is resolved by error-prone repair pathways that introduce point mutations characteristic of somatic hypermutation. The availability of homologous donors and high homologous recombination activity, on the other hand, favors resolution by gene conversion. Irreversible commitment to the gene conversion pathway, however, only occurs during strand exchange, and prior to this step, a shift to somatic hypermutation is still possible. Thus, deletion of individual RAD51 paralogs, which participate in steps preceding strand exchange, decreases Ig gene conversion and induces Ig somatic hypermutation, while deletion of RAD52 members that are active after strand exchange, such as RAD54, only decreases Ig gene conversion (Sale et al. 2001; Bezzubova et al. 1997). This model also predicts that Ig gene conversion will only occur if homologous donors and high homologous recombination activity are both available, potentially explaining why B cells in some species utilize Ig gene conversion, while those in other species do not.

11 Antibody Repertoire Diversification in GALT in Humans and Mice

Although mice and humans use combinatorial rearrangement of multiple V, (D), and J gene segments as a strategy for generating a diverse primary antibody repertoire, might the strategy employed by chickens, rabbits, and artiodactyls be conserved in some human and mouse B cell populations? A number of observations suggest that this might indeed be the case. Casola et al. (2004), for example, found that B cell receptor-deficient transgenic mice spontaneously develop germinal centers (GCs) in Peyer's patches. This antigen-independent B cell activation in mouse GALT, presumably induced by signals acquired from the intestinal microbiota, is reminiscent of the antigen-independent B cell activation stimulated by intestinal commensals in rabbit GALT that drives diversification of the primary antibody repertoire. Furthermore, Shimomura et al. (2008) identified an intestinal B cell population that develops independently of B cell receptor specificity, Lyn or Btk signaling, or T-cell help. A proportion of these B cells expressed mutated V(D)J genes, suggesting they had undergone antigen-independent diversification by somatic hypermutation.

A B cell population with similar characteristics has been identified in humans with common variable immune disorders caused by mutations in genes encoding proteins required for T-cell help and GC formation. A circulating population of IgM^+IgD^+ B cells has been identified in these patients that expresses the post-GC memory B cell marker CD27 and has mutated V(D)J genes (Weller et al. 2001). These B cells are probably not generated during T-cell-independent (TI), antigen-specific responses to bacterial capsules or cell wall components because they are also found in human infants too young to respond to TI antigens (less than 2 years of age) (Weller et al. 2008). Instead, Weill et al. (2004) suggest that human $CD27^+IgM^+IgD^+$ B cells diversify their V(D)J genes independently of antigen, in a manner similar to that seen in "GALT species" (Weill et al. 2004). Interestingly, the $CD27^+IgM^+IgD^+$ B cell population is markedly reduced in children deficient in MyD88 or IRAK-4 (interleukin-1 receptor-associated kinase 4), molecules required for MyD88-dependent Toll-like receptor (TLR) signaling (Maglione et al. 2014). These children suffer from life-threatening, often recurrent, pyogenic bacterial infections, particularly invasive pneumococcal disease (Picard et al. 2003; von Bernuth et al. 2008). They are, however, otherwise healthy and mount protective immune responses against other bacterial infections. Moreover, their clinical status improves with age, becoming essentially normal by 10 years of age. These observations suggest that human $CD27^+IgM^+IgD^+$ B cells provide important, nonredundant immune protection against pyogenic bacteria early in life. The $CD27^+IgM^+IgD^+$ phenotype is characteristic of marginal zone B cells, which provide a first line of defense because they localize in marginal sinuses where the blood first enters the spleen and produce low affinity, polyreactive "innate IgM" that effectively recognizes repetitive epitopes, such as microbial glycans and carbohydrates. Significantly, Maglione et al. (2014) found that serum IgM of children with MyD88 or IRAK-4 deficiency recognizes a greatly reduced range of microbial glycans, compared to that of normal children. Taken together, these data suggest that human $CD27^+IgM^+IgD^+$ B cells generate a diversified antibody repertoire in a TLR-dependent, antigen-independent manner and are important for protection against pyogenic bacteria early in life. These similarities to the antibody diversification strategy used by the "GALT species" suggest that this strategy might be conserved in human and mouse B cell populations serving specialized functions.

12 Concluding Statement

The generation of a diverse primary antibody repertoire in vertebrates is fundamentally important for protection against pathogens. The emergence, during vertebrate evolution, of primary antibody repertoire diversification by AID-mediated gene conversion provided a highly effective solution to this problem. This strategy is likely more widely utilized among vertebrates than is currently appreciated and might be conserved in specialized B cell populations in those that generate their

primary antibody repertoires by combinatorial rearrangement of multiple V, (D), and J gene segments.

References

Arakawa H, Buerstedde JM (2004) Immunoglobulin gene conversion: insights from bursal B cells and the DT40 cell line. Dev Dyn 229:458–464

Arakawa H, Buerstedde JM (2009) Activation-induced cytidine deaminase-mediated hypermutation in the DT40 cell line. Philos Trans R Soc Lond B Biol Sci 364:639–644

Arakawa H, Furusawa S, Ekino S, Yamagishi H (1996) Immunoglobulin gene hyperconversion ongoing in chicken splenic germinal centers. EMBO J 15(10):2540–2546

Arakawa H, Hauschild J, Buerstedde JM (2002) Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science 295:1301–1306

Arakawa H, Saribasak H, Buerstedde JM (2004) Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. PLoS Biol 2, E179

Barrington RA, Fasullo M, Knight KL (1999) A role for RAD51 in the generation of immunoglobulin gene diversity in rabbits. J Immunol 162:911–919

Becker RS, Knight KL (1990) Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. Cell 63:987–997

Becker RS, Zhai SK, Currier SJ, Knight KL (1989) Ig VH, DH and JH germ-line gene segments linked by overlapping cosmid clones of rabbit DNA. J Immunol 142:1351–1355

Bezzubova O, Silbergleit A, Yamaguchi-Iwai Y, Takeda S, Buerstedde J-M (1997) Reduced X-ray resistance and homologous recombination frequencies in a RAD54-/- mutant of the chicken DT40 cell line. Cell 89:185–193

Buerstedde JM, Takeda S (1991) Increased ratio of targeted to random integration after transfection of chicken B cell lines. Cell 67:179–188

Buerstedde JM, Reynaud CA, Humphries HE, Olson W, Ewert DL, Weill JC (1990) Light chain gene conversion continues at high rate in an ALV-induced cell line. EMBO J 9:921–927

Butler JE, Santiago-Mateo K, Sun XZ, Wertz N, Sinkora M, Francis DH (2011) Antibody repertoire development in fetal and neonatal piglets. XX. B cell lymphogenesis is absent in the ileal Peyer's patches, their repertoire development is antigen dependent, and they are not required for B cell maintenance. J Immunol 187:5141–5149

Carlson LM, Mccormack WT, Postema CE, Humphries EH, Thompson CB (1990) Templated insertions in the rearranged chicken IgL V gene segment arise by intrachromosomal gene conversion. Genes Dev 4:536–547

Casola S, Otipoby KL, Alimzhanov M, Humme S, Uyttersprot N, Kutok JL, Carroll MC, Rajewsky K (2004) B cell receptor signal strength determines B cell fate. Nat Immunol 5:317–327

Di Noia J, Neuberger MS (2002) Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature 419:43–48

Di Noia JM, Neuberger MS (2007) Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem 76:1–22

Friedman ML, Tunyaplin C, Zhai SK, Knight KL (1994) Neonatal V_H, D and JH gene usage in rabbit B-lineage cells. J Immunol 152:632–641

Fujimori A, Tachiiri S, Sonoda E, Thompson LH, Dhar PK, Hiraoka M, Takeda S, Zhang Y, Reth M, Takata M (2001) Rad52 partially substitutes for the Rad51 paralog XRCC3 in maintaining chromosomal integrity in vertebrate cells. EMBO J 20:5513–5520

Harris RS, Sale JE, Petersen-Mahrt SK, Neuberger MS (2002) AID is essential for immunoglobulin V gene conversion in a cultured B cell line. Curr Biol 12:435–438

Kawabata M, Kawabata T, Nishibori M (2005) Role of recA/RAD51 family proteins in mammals. Acta Med Okayama 59:1–9

Kim S, Humphries EH, Tjoelker L, Carlson L, Thompson CB (1990) Ongoing diversification of the rearranged immunoglobulin light-chain gene in a bursal lymphoma cell line. Mol Cell Biol 10:3224–3231

Knight KL, Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. Cell 60:963–970

Lavinder JJ, Hoi KH, Reddy ST, Wine Y, Georgiou G (2014) Systematic characterization and comparative analysis of the rabbit immunoglobulin repertoire. PLoS One 9, e101322

Liskay RM, Letsou A, Stachelek JL (1987) Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics 115:161–167

Maglione PJ, Simchoni N, Black S, Radigan L, Overbey JR, Bagiella E, Bussel JB, Bossuyt X, Casanova JL, Meyts I, Cerutti A, Picard C, Cunningham-Rundles C (2014) IRAK-4 and MyD88 deficiencies impair IgM responses against T-independent bacterial antigens. Blood 124:3561–3571

Mccormack WT, Thompson CB (1990) Chicken Ig_L variable region gene conversions display pseudogene donor preference and $5'$ and $3'$ polarity. Genes Dev 4:548–558

Mccormack WT, Carlson LM, Tjoelker LW, Thompson CB (1989a) Evolutionary comparison of the avian IgL locus: combinatorial diversity plays a role in the generation of the antibody repertoire in some avian species. Int Immunol 1:332–341

Mccormack WT, Tjoelker LW, Barth CF, Carlson LM, Petryniak B, Humphries EH, Thompson CB (1989b) Selection for B cells with productive IgL gene rearrangements occurs in the bursa of Fabricius during chicken embryonic development. Genes Dev 3:838–847

Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem 274:18470–18476

Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102:553–563

Ordinario EC, Yabuki M, Handa P, Cummings WJ, Maizels N (2009) RAD51 paralogs promote homology-directed repair at diversifying immunoglobulin V regions. BMC Mol Biol 10:98

Parng C-L, Hansal S, Goldsby RA, Osborne BA (1996) Gene conversion contributes to Ig light chain diversity in cattle. J Immunol 157:5478–5486

Picard C, Puel A, Bonnet M, Ku CL, Bustamante J, Yang K, Soudais C, Dupuis S, Feinberg J, Fieschi C, Elbim C, Hitchcock R, Lammas D, Davies G, Al-Ghonaium A, Al-Rayes H, Al-Jumaah S, Al-Hajjar S, Al-Mohsen IZ, Frayha HH, Rucker R, Hawn TR, Aderem A, Tufenkeji H, Haraguchi S, Day NK, Good RA, Gougerot-Pocidalo MA, Ozinsky A, Casanova JL (2003) Pyogenic bacterial infections in humans with IRAK-4 deficiency. Science 299:2076–2079

Pospisil R, Young-Cooper GO, Mage RG (1995) Preferential expansion and survival of B lymphocytes based on VH framework 1 and framework 3 expression: "Positive" selection in appendix of normal and VH-mutant rabbits. Proc Natl Acad Sci USA 92:6961–6965

Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Labelouse R, Gennery A, Tezcan I, Ersoy F, Kayserili H, Ugazio AG, Brousse N, Muramatsu M, Notarangelo LD, Kinoshita K, Honjo T, Fischer A, Durandy A (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell 102:565–575

Reynaud C-A, Anquez V, Dahan A, Weill J-C (1985) A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. Cell 40:283–291

Reynaud C-A, Anquez V, Grimal H, Weill J-C (1987) A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell 48:379–388

Reynaud C-A, Dahan A, Anquez V, Weill J-C (1989) Somatic hyperconversion diversifies the single V_H gene of the chicken with a high incidence in the D region. Cell 59:171–183

Reynaud C-A, Bertocci B, Dahan A, Weill J-C (1994) Formation of the chicken B-cell repertoire: Ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. Adv Immunol 57:353–378

Rhee KJ, Sethupathi P, Driks A, Lanning DK, Knight KL (2004) Role of commensal bacteria in development of gut-associated lymphoid tissues and preimmune antibody repertoire. J Immunol 172:1118–1124

Rhee KJ, Jasper PJ, Sethupathi P, Shanmugam M, Lanning D, Knight KL (2005) Positive selection of the peripheral B cell repertoire in gut-associated lymphoid tissues. J Exp Med 201:55–62

Rogozin IB, Kolchanov NA (1992) Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim Biophys Acta 1171:11–18

Sale JE, Calandrini DM, Takata M, Takeda S, Neuberger MS (2001) Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. Nature 412:921–926

Sayegh CE, Demaries SL, Iacampo S, Ratcliffe MJ (1999) Development of B cells expressing surface immunoglobulin molecules that lack V(D)J-encoded determinants in the avian embryo bursa of fabricius. Proc Natl Acad Sci USA 96:10806–10811

Schiaffella E, Fuschiotti P, Bensinger SJ, Mage RG (1998) High RAD51 mRNA levels in young rabbit appendix. A role in B-cell gene conversion? Immunogenetics 48:108–115

Sehgal D, Johnson G, Wu TT, Mage RG (1999) Generation of the primary antibody repertoire in rabbits: expression of a diverse set of Igk-V genes may compensate for limited combinatorial diversity at the heavy chain locus. Immunogenetics 50:31–42

Sehgal D, Obiakor H, Mage RG (2002) Distinct clonal Ig diversification patterns in young appendix compared to antigen-specific splenic clones. J Immunol 168:5424–5433

Severson KM, Mallozzi M, Driks A, Knight KL (2010) B cell development in GALT: role of bacterial superantigen-like molecules. J Immunol 184:6782–6789

Shimomura Y, Ogawa A, Kawada M, Sugimoto K, Mizoguchi E, Shi HN, Pillai S, Bhan AK, Mizoguchi A (2008) A unique B2 B cell subset in the intestine. J Exp Med 205:1343–1355

Sonoda E, Sasaki MS, Buerstedde JM, Bezzubova O, Shinohara A, Ogawa H, Takata M, Yamaguchi-Iwai Y, Takeda S (1998) Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. EMBO J 17:598–608

Sun J, Hayward C, Shinde R, Christenson R, Ford SP, Butler JE (1998) Antibody repertoire development in fetal and neonatal piglets. I. Four VH genes account for 80 percent of VH usage during 84 days of fetal life. J Immunol 161:5070–5078

Takata M, Sasaki MS, Sonoda E, Fukushima T, Morrison C, Albala JS, Swagemakers SM, Kanaar R, Thompson LH, Takeda S (2000) The Rad51 paralog Rad51B promotes homologous recombinational repair. Mol Cell Biol 20:6476–6482

Takata M, Sasaki MS, Tachiiri S, Fukushima T, Sonoda E, Schild D, Thompson LH, Takeda S (2001) Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. Mol Cell Biol 21:2858–2866

Tauchi H, Kobayashi J, Morishima K, Van Gent DC, Shiraishi T, Verkaik NS, Vanheems D, Ito E, Nakamura A, Sonoda E, Takata M, Takeda S, Matsuura S, Komatsu K (2002) Nbs1 is essential for DNA repair by homologous recombination in higher vertebrate cells. Nature 420:93–98

Thompson CB, Neiman PE (1987) Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. Cell 48:369–378

Vajdy M, Sethupathi P, Knight KL (1998) Dependence of antibody somatic diversification on gut-associated lymphoid tissue in rabbit. J Immunol 160:2725–2729

Von Bernuth H, Picard C, Jin Z, Pankla R, Xiao H, Ku CL, Chrabieh M, Mustapha IB, Ghandil P, Camcioglu Y, Vasconcelos J, Sirvent N, Guedes M, Vitor AB, Herrero-Mata MJ, Arostegui JI, Rodrigo C, Alsina L, Ruiz-Ortiz E, Juan M, Fortuny C, Yague J, Anton J, Pascal M, Chang HH, Janniere L, Rose Y, Garty BZ, Chapel H, Issekutz A, Marodi L, Rodriguez-Gallego C, Banchereau J, Abel L, Li X, Chaussabel D, Puel A, Casanova JL (2008) Pyogenic bacterial infections in humans with MyD88 deficiency. Science 321:691–6

Waldman AS, Liskay RM (1987) Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. Proc Natl Acad Sci USA 84:5340–5344

Weill JC, Reynaud CA, Lassila O, Pink JR (1986) Rearrangement of chicken immunoglobulin genes is not an ongoing process in the embryonic bursa of Fabricius. Proc Natl Acad Sci USA 83:3336–3340

Weill JC, Weller S, Reynaud CA (2004) A bird's eye view on human B cells. Semin Immunol 16:277–281

Weinstein PD, Anderson AO, Mage RG (1994) Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms. Immunity 1:647–659

Weller S, Faili A, Garcia C, Braun MC, Le Deist FF, de Saint Basile GG, Hermine O, Fischer A, Reynaud CA, Weill JC (2001) CD40-CD40L independent Ig gene hypermutation suggests a second B cell diversification pathway in humans. Proc Natl Acad Sci USA 98:1166–1170

Weller S, Mamani-Matsuda M, Picard C, Cordier C, Lecoeuche D, Gauthier F, Weill JC, Reynaud CA (2008) Somatic diversification in the absence of antigen-driven responses is the hallmark of the IgM+ IgD+ CD27+ B cell repertoire in infants. J Exp Med 205:1331–1342

Winstead CR, Zhai SK, Sethupathi P, Knight KL (1999) Antigen-induced somatic diversification of rabbit IgH genes: gene conversion and point mutation. J Immunol 162:6602–6612

Zhai SK, Lanning DK (2013) Diversification of the primary antibody repertoire begins during early follicle development in the rabbit appendix. Mol Immunol 54:140–147

Antibody Isotype Switching in Vertebrates

Kate Senger, Jason Hackney, Jian Payandeh, and Ali A. Zarrin

Abstract The humoral or antibody-mediated immune response in vertebrates has evolved to respond to diverse antigenic challenges in various anatomical locations. Diversification of the immunoglobulin heavy chain (IgH) constant region via isotype switching allows for remarkable plasticity in the immune response, including versatile tissue distribution, Fc receptor binding, and complement fixation. This enables antibody molecules to exert various biological functions while maintaining antigen-binding specificity. Different immunoglobulin (Ig) classes include IgM, IgD, IgG, IgE, and IgA, which exist as surface-bound and secreted forms. High-affinity autoantibodies are associated with various autoimmune diseases such as lupus and arthritis, while defects in components of isotype switching are associated with infections. A major route of infection used by a large number of pathogens is invasion of mucosal surfaces within the respiratory, digestive, or urinary tract. Most infections of this nature are initially limited by effector mechanisms such as secretory IgA antibodies. Mucosal surfaces have been proposed as a major site for the genesis of adaptive immune responses, not just in fighting infections but also in tolerating commensals and constant dietary antigens. We will discuss the evolution of isotype switching in various species and provide an overview of the function of various isotypes with a focus on IgA, which is universally important in gut homeostasis as well as pathogen clearance. Finally, we will discuss the utility of antibodies as therapeutic modalities.

K. Senger • A.A. Zarrin (✉)
Department of Immunology, Genentech Inc., South San Francisco, CA 94080, USA
e-mail: zarrin.ali@gene.com

J. Hackney
Department of Bioinformatics, Genentech Inc., South San Francisco, CA 94080, USA

J. Payandeh
Department of Structural Biology, Genentech Inc., South San Francisco, CA 94080, USA

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0_13

1 Evolution of the IgH Locus

Antibody is composed of four polypeptide chains including two heavy chain (H) and two light (L) chains joined by several disulfide bonds. Each class of Ig defines an individual set of C heavy chain (C_H) domain that corresponds to a single H chain constant region gene. The C_H gene Cμ initially displayed on developing B cells as IgM can be altered by IgH isotype switching to produce IgG, IgE, or IgA. In this review, we focus on IgH genes across evolution.

Immunoglobulin genes are found in all gnathostomes (jawed vertebrates) analyzed to date. Their earliest appearance is in jawed fish (Fig. 1). Among fish, two distinct arrangements of variable (V), diverse (D), joining (J), and heavy chain constant (C_H) regions are found (Stavnezer and Amemiya 2004; Stavnezer and Schrader 2014). Cartilaginous fish such as sharks have many (10s–100s) copies of immunoglobulin genes, each consisting of a V–D–J-like region coupled to an associated constant region (Fig. 1a). This is known as a cluster arrangement of immunoglobulin genes. Interestingly, recent data in shark suggests that V(D)J can be expressed with constant regions from different clusters, although IgH genes are spatially distant, at >120 kb (Zhu et al. 2012). The cluster arrangement gave way to the translocon arrangement seen earliest in bony fish (e.g., zebrafish), in which each constant region is paired with a varying number of V, D, and J segments (Fig. 1b). Rearranged V(D)J segments in bony and lobe-finned fish seem to be limited only to their adjacent C region exons within one cluster (Fig. 1b–d), although in bony fish there is the innovation of using alternative splicing to express IgD instead of IgM (Chen and Cerutti 2011). Extensive splicing events within expressed clusters have also been reported, which could further diversify the antibody genes in lower species (Zhang et al. 2013). Further specialization of the immunoglobulin locus is seen in amphibians such as the frog *Xenopus*, where the translocon arrangement of V–D–J is associated with a set of individual isotypes, each of which can carry out specialized functions (Fig. 1e). In this arrangement, the recombined V(D)J can make use of any of the available isotypes through class switch recombination. This proved to be a very successful arrangement, utilized in nearly all other tetrapods examined from crocodiles to humans (Fig. 1f–m).

Among the constant region sequences, Cμ (IgM) appears to be common across all immunoglobulin-bearing clades (Figs. 1 and 2, red and purple coloring). Cδ/Cω (IgD/IgW) also appears early, being present in elasmobranchs, like sharks, bony fish, and lobe-finned fish, such as coelacanths. However, there is no evidence to date of Cδ/Cω in the holocephalans, like ratfish (Rast et al. 1998) and elephant sharks (Venkatesh et al. 2014). In addition, other isotypes like IgNAR may also be present depending on the species. Cζ, which is the constant chain for IgZ ["Z" for "zebrafish" (Danilova et al. 2005), also known as IgT ("T" for "teleost") (Hansen et al. 2005)], is found in ray-finned fishes and is the earliest Ig known to play a role in mucosal biology. IgZ is highly expressed as a multimer in gut mucus and is transported across epithelial cells in a manner similar to IgA, which will be discussed later.

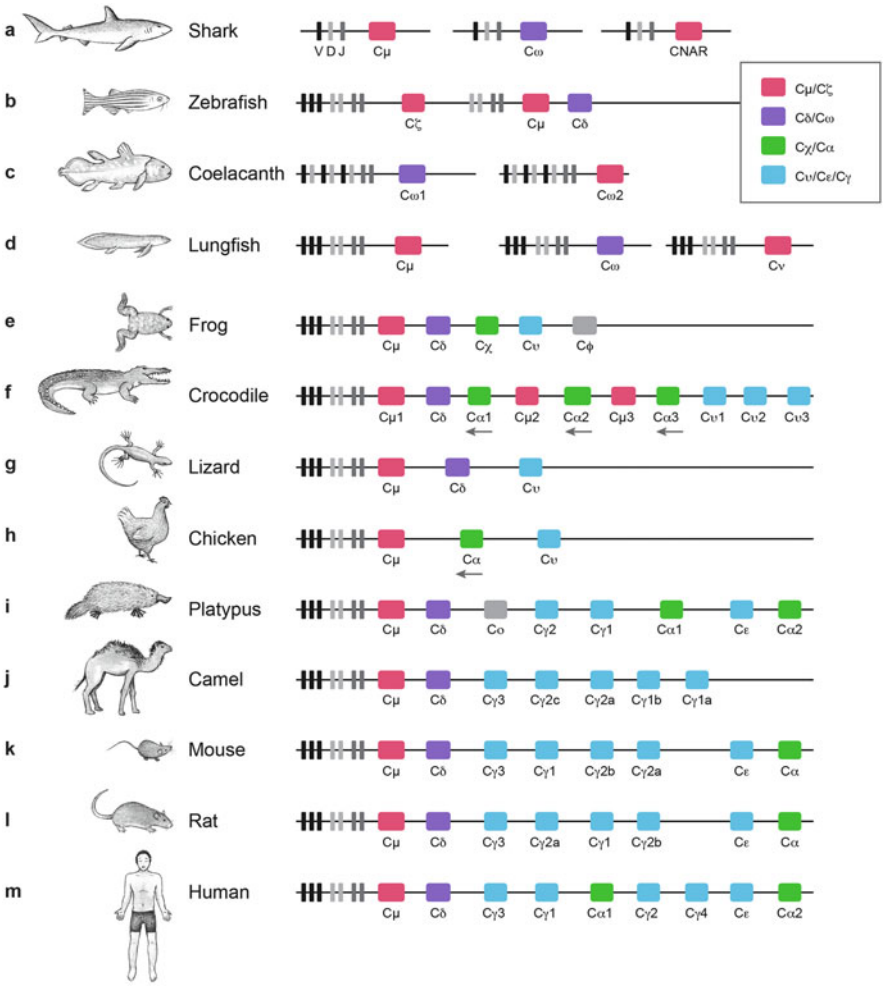


Fig. 1 Genomic arrangement of IgH loci. The arrangement of V–D–J and C regions from selected species is shown. Sharks (**a**) have a cluster arrangement comprised of individual V–D–J segments (*gray*) coupled to a constant region, with this arrangement duplicated anywhere from 10 to 100s of times. Ray-finned fish such as zebrafish (**b**) have a translocon-based arrangement of many V, D, and J segments coupled to constant region exons, but so far do not show class switch recombination. Coelacanth (**c**) and lungfish (**d**) have a similar translocon arrangement, but it is not currently known if their heavy chain loci are on the same chromosome or on different chromosomes. Salamanders, frogs, and all other vertebrates shown have a translocon arrangement of multiple V, D, and J segments coupled with several constant regions (**e–m**). These clades show class switch recombination and have S regions preceding the constant region exons. *Arrows* indicate the orientation of heavy chain coding sequences. Pseudogenes are excluded. Illustration not drawn to scale

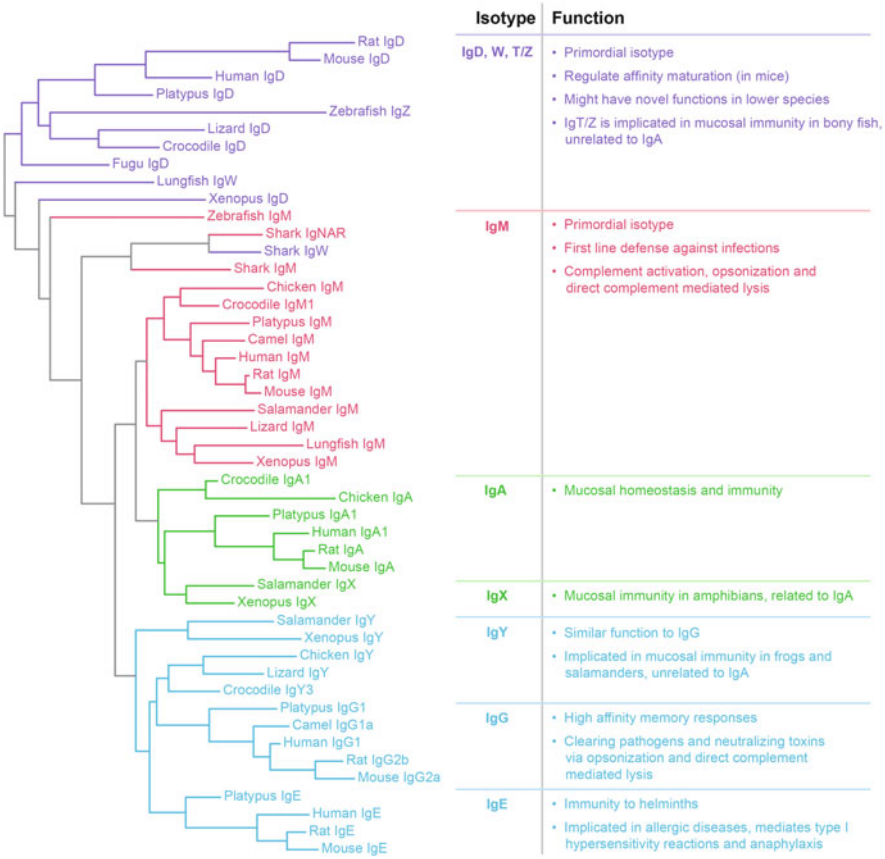


Fig. 2 Phylogenetic tree and properties of IgH constant regions. Heavy chain constant regions, from the first amino acid of the first constant region Ig domain, were retrieved from GenBank. Protein sequences were aligned using MUSCLE (Edgar 2004), and the phylogenetic tree was constructed using PhyML (Guindon et al. 2010). The resulting tree was visualized using TreeDyn (Chevenet et al. 2006). Accession numbers for sequences: shark (*Triakis scyllium*) IgM BAJ20181.1, IgW BAJ20183.1, IgNAR BAJ20188.1; zebrafish (*Danio rerio*) IgM AAK69167.1, IgZ ACH92959.1; fugu (*Takifugu rubripes*) IgD BAD34541.1; lungfish (*Protopterus dolloi*) IgM AGT03500.1, IgW AGT03500.1; salamander (*Ambystoma mexicanum*) IgX CAO82107.1, IgY CAA49247.1; *Xenopus laevis* IgM AAA49878.1, IgX AAH72981.1, IgY CAA33212.1; *Xenopus tropicalis* IgD ABC75541.1; crocodile (*Crocodylus siamensis*) IgM AFZ39206.1, IgD AFZ39209.1, IgA1 AFZ39210.1, IgY3 AFZ39221.1; lizard (*Anolis carolinensis*) IgM ABV66128.1, IgD ABV66130.1, IgY ABV66131.1; chicken (*Gallus gallus*) IgM CAA25762.1, IgA AAB22614.2, IgY CAA30161.1; platypus (*Ornithorhynchus anatinus*) IgM AAO37747.1, IgD ACD31540.1, IgG1 AAL17703.1, IgA1 AAL17700.1, IgE AAL17702.1; camel (*Camelus dromedarius*) IgM BAD00196.1, IgG1a CAD13185.1; mouse (*Mus musculus*) IgM CAC20701.1, IgD 0912262A, IgG2a CAA24179.1, IgE AAZ05128.1, IgA AAD56886.1; rat (*Rattus norvegicus*) IgM AAH92586.1, IgD AAO19643.1, IgG2b ADX94419.1, IgE CAA25439.1, IgA CAD52870.1; and human (*Homo sapiens*) IgM IGHM_HUMAN, IgD IGHD_HUMAN, IgG1 IGHG1_HUMAN, IgA1 CAC20453.1, IgE AAB59395.1

Amphibians retain Cμ and Cδ and add two new heavy chains: Cυ (IgY, Figs. 1 and 2, blue coloring), which is derived from Cμ, and Cχ (IgX, Figs. 1 and 2, green coloring), which arose from a fusion of Cμ and Cυ. Both IgY and IgX are implicated in defending the gut. In the newt *Pleurodeles waltl*, IgY-expressing B cells are mainly found in the mucosa (Schaerlinger et al. 2008), and in young axolotls, IgY localizes to the cells of the stomach and intestinal mucosa by immunofluorescence (Fellah et al. 1992a). IgX has been studied extensively in the African clawed toad (Hsu et al. 1985). It is mainly expressed as a multimer in mucosal tissues, though it is thought to lack the J chain often found linking multimeric Igs such as IgM and IgA (Amemiya et al. 1989; Fellah et al. 1992b, 1993; Haire et al. 1989; Hsu et al. 1985; Mussmann et al. 1996; Schaerlinger and Frippiat 2008; Schwager et al. 1988). The C_H1 and C_H2 domains of Cχ bear clear homology to C_H1 and C_H2 of Cυ, while the C_H3 and C_H4 domains of Cχ are most similar to Cμ.

The Cα heavy chain (IgA), found in reptiles, birds, and mammals, is highly related to Cχ in amphibians, sharing a branch in the phylogenetic tree (Figs. 1 and 2, green coloring). This isotype plays a key role in mucosal defense in all amniotes examined to date with the exception of several reptiles. For example, IgA is present in the leopard gecko *Eublepharis macularius* (Deza et al. 2007), but has not been detected in several snake species (Gambon-Deza et al. 2012), the American anole *Anolis carolinensis* (Gambon-Deza et al. 2009; Wei et al. 2009), or the Chinese soft-shelled turtle *Pelodiscus sinensis* (Xu et al. 2009). Interestingly, in birds such as chicken and ducks, the Cα gene is inverted relative to other heavy chain coding sequences (Mansikka 1992; Magor et al. 1998; Lundqvist et al. 2001). Likewise, the order Crocodilia, which houses the closest relatives to living birds, carries inverted Cα genes. The crocodile and alligator genomes have an unusual high number of Cα genes—3 plus one pseudogene (Cheng et al. 2013), and all are in a reverse orientation (Cheng et al. 2013; Magadan-Mompo et al. 2013). This genomic configuration provides a phylogenetic link between crocodiles and birds and places the inversion of the Cα gene prior to their evolutionary divergence.

The vast majority of mammals examined have a single Cα gene in the same orientation as the other C regions (Woof and Kerr 2004), with intriguing exceptions. Rabbit has 13 Cα genes (not shown), of which 11 have confirmed expression. These IgA subclasses have highly diverse C_H1 domains that are proximal to the hinge region, likely affecting their susceptibility to bacterial proteases. It has been suggested that proteases may have, in fact, been a key factor in shaping IgA diversity in lagomorphs. Another exception to the single Cα gene "rule" is the Hominoid primates (gibbons, gorillas, chimpanzees, humans), which have two IgA subclasses, IgA1 and IgA2. These subclasses arose through gene duplication and are highly similar. The major difference between them lies at the hinge between the two Fab arms and the Fc region. IgA1 has an extended hinge due to a duplicated stretch of amino acids, which may allow for higher avidity bivalent interactions with distantly spaced antigens (Boehm et al. 1999; Furtado et al. 2004). Secreted IgA1 is predominantly found in the gastrointestinal and upper respiratory tracts. Secreted IgA2 is the dominant subclass in secretions and is localized to the

lower ileum and large intestine, possibly because it has decreased susceptibility to bacterial proteases.

Cγ (IgG) and Cε (IgE) are two other heavy chains present in all mammals studied, including the most primitive egg-laying mammals such as platypus (Figs. 1 and 2, blue coloring). Both isotypes appear to have arisen from a duplication of C*υ*. IgG is required for high-affinity memory responses and clears pathogens and toxins via opsonization and complement-mediated lysis. IgE is involved in immunity to helminths and is responsible for allergies and anaphylaxis. In addition to these "classical" isotypes, there are also several species-specific isotypes that arise throughout evolution, such as Cϕ (IgF) in *Xenopus* and Co (IgO) in platypus (Fig. 1e and i, respectively). The functions of these isotypes remain undiscovered. It is important to emphasize common ancestry among isotypes does not necessarily predict functional similarity. There is much to be learned about the function of various isotypes among species.

2 Antibody Isotype Switching

Isotype switching is the process by which C_H domains encoding one isotype are exchanged for another, thereby influencing the effector function of the resulting antibody. In mammals such as mice and humans, this is regulated by T cell-dependent or T cell-independent mechanisms. T cell-dependent isotype switching involves T helper cells (Th) that are activated by antigen (Ag) and B7 co-stimulation, which in turn can activate CD40 ligand (CD40L). CD40 receptor engagement on B cells can activate the isotype switching machinery. Activated T cells also produce a variety of cytokines such as IL-13, IL-4, IL-5, or IFN-γ that can selectively impact the induction of various antibody isotypes. The net effect of Th-dependent isotype switching is production of high-affinity antibodies such as the IgG1 subclasses. Polysaccharide antigens derived from certain bacterial capsules stimulate mainly IgM antibodies, which can bind to bacteria, activate complement, and induce the phagocytosis of opsonized bacteria. During bacterial or viral infections, Th cells are activated to elicit IgG antibodies, which can block pathogen entry to tissues and facilitate their phagocytosis. Helminth infections can induce Th cells to produce Th2 cytokines to induce the IgE response. In addition, B cells localized within the gut produce IgA as a first line of defense, which is mediated by factors like TGF-β (transforming growth factor beta) that is produced by various cell types. Additional cytokines including BAFF (B cell-activating factor) and APRIL (a proliferation-inducing ligand) have been implicated as T cell-independent stimuli for IgA isotype switching via the receptor TACI (transmembrane activator and calcium modulator and cyclophilin ligand interactor). The molecular mechanism of isotype switching is known as class switch recombination (CSR).

3 Mechanism of Class Switch Recombination (CSR)

In many vertebrates, the IgH locus contains a battery of C_H encoding exons with common or distinct genomic configurations. In mice, for example, C_H regions are organized in the following order: 5′-Cμ-Cδ-Cγ3-Cγ1-Cγ2b-Cγ2a-Cε-Cα-3′ (Fig. 3). Each C_H gene is a component of a germ-line transcription (GLT) unit consisting of an intronic (I) promoter, a noncoding I exon, switch (S) region, and C_H coding exons. CSR is a genetically programmed gene recombination process in which the exons for constant region Cμ, encoding IgM, can be exchanged for downstream C_H regions, yielding B cells that express IgG, IgE, or IgA. I promoters are normally silent in naïve B cells except for Iμ, which is continuously active. Upon B cell activation, one or more downstream I promoters become activated while activation-induced cytidine deaminase (AID) is upregulated. Germ-line transcription facilitates the access of AID to S regions by creating single-stranded DNA (ssDNA), which is the optimal target for AID. AID recognizes cytosines within hotspot motifs and converts them to uracil. This DNA lesion is then processed by mismatch repair (MMR) and base excision repair (BER) pathways, resulting in point mutations and/or double-stranded DNA breaks (DSBs). Distant DSBs in two switch regions are joined mainly via nonhomologous end-joining or microhomology-dependent alternative end-joining pathways. Typically, recombination of the donor S region (Sμ) with a downstream acceptor S region is accompanied by deletion of the intervening sequences such as between Sμ and Sα, producing the large circular DNA shown in Fig. 3. In the next section, we will discuss observations on how this process evolved.

4 Evolution of CSR in Vertebrates

4.1 Evolution of Switch Regions

Switch (S) regions are 1–10 kb repetitive intronic sequences upstream of C_H regions that are the sites of DSB formation and subsequent end joining during CSR. It appears the main function of S regions is to direct AID to C_H intronic regions, causing the formation of double-stranded breaks (DSBs). Ancient *Xenopus* Sμ in place of mouse Sγ1 can support CSR, suggesting CSR substrates are interchangeable (Zarrin et al. 2004). In addition, the function of S regions and AID can be replaced by DSBs induced by the yeast endonuclease I-SceI, suggesting S regions are not required for the joining of two lesions within the IgH locus (Zarrin et al. 2007), and that any means of forming DSBs may be sufficient for isotype switching.

Mammalian S regions are G-rich on the non-template strand. Certain motifs, such as TGGGG, GGGGT, GGGCT, GAGCT, and AGCT, are common. Among these motifs, AGCT is evolutionarily conserved in amphibians (Mussmann

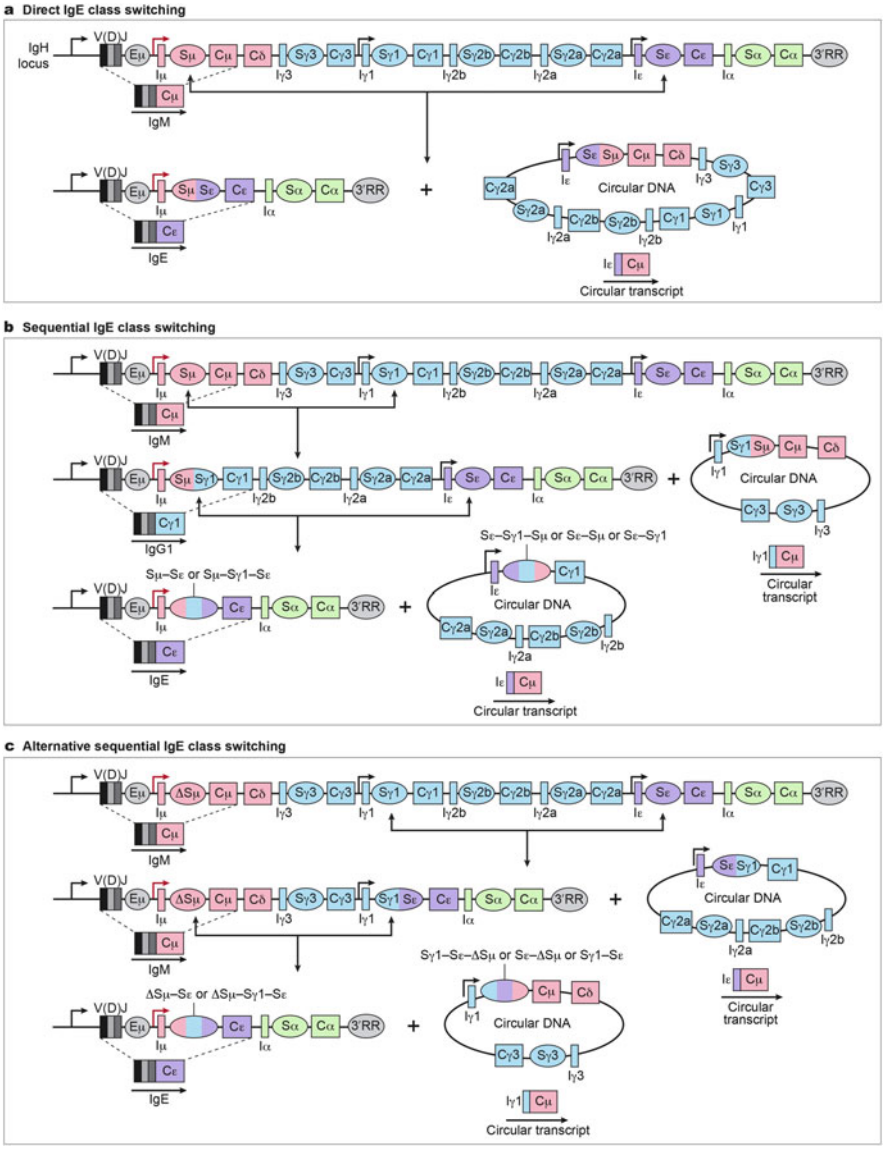


Fig. 3 Mechanism of class switch recombination (CSR). The recombination event that results in the replacement of Cμ with Cα to produce IgA is illustrated. The organization of the mouse IgH locus is diagrammed, with the *red arrow* indicating constitutive germ-line transcription through Sμ. The *black arrow* represents inducible transcription through Sα. AID targets transcribed donor Sμ and acceptor Sα to initiate double-stranded DNA breaks, which is accompanied by deletion of the intervening DNA. The recombined locus is shown *lower left*. The intervening DNA forms a short-lived circle, shown *lower right*, from which transcripts are generated

et al. 1997; Zhao et al. 2006; Ohta and Flajnik 2006) and birds (Lundqvist et al. 2001). Additional palindromic sequences such as CATG, AGCA, and TGCA are also abundant in amphibians and may be involved in CSR (Mussmann et al. 1997). The key enzyme that targets these sequences, AID, preferentially targets RGYW/WRCY motifs within S regions (where R = A/G; Y = C/T; W = A/T; hotspot = G:C) downstream of transcriptionally active start sites (Xue et al. 2006). The overlap and density of AID hotspot sequences on two DNA strands correlates with the quality of the regions and their CSR efficiency (Zarrin et al. 2004, 2005; Xue et al. 2006; Han et al. 2011). Switch regions have not yet been found in fish, although in shark the V(D)J region from one cluster has been found expressed with C_H regions from different clusters, due to recombination within the J–C intron via an unknown mechanism (Zhu et al. 2012).

Structures such as R-loops and G quartets can facilitate AID targeting and CSR. Upon transcription of GC-rich S regions, the RNA remains associated with the template DNA strand, generating an R-loop structure. These R-loops are thought to mediate AID function at the S region (Huang et al. 2007; Roy et al. 2008, 2010; Roy and Lieber 2009; Shinkura et al. 2003; Yu et al. 2003). Experiments using a mouse B cell line have shown that the propensity of a region to form R-loops is an important variable for determining the efficiency of switching to an S region (Zhang et al. 2014). The more G-rich motifs (e.g., TGGGG) seem to have been acquired recently in mammals (Tian and Alt 2000; Shinkura et al. 2003; Yu et al. 2003; Zarrin et al. 2004; Huang et al. 2007; Roy et al. 2008) suggesting that these sequences represent a more efficient target for AID, which itself has evolved over time.

4.2 Evolution of AID

AID was initially proposed to function as an RNA-editing enzyme based on sequence homology to known RNA-editing enzymes, such as the bacterial RNA cytidine deaminase and mammalian APOBEC (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like) family of enzymes, which contains 11 members (Navaratnam et al. 1998; Conticello et al. 2005). It has since been established that AID targets single-stranded DNA (ssDNA) and ssDNA–RNA hybrid structures within transcription bubbles (Petersen-Mahrt et al. 2002; Dickerson et al. 2003; Bransteitter et al. 2003, 2004; Martin and Scharff 2002; Ramiro et al. 2003; Chaudhuri et al. 2003; Yu et al. 2003). Nonimmunoglobulin loci are also susceptible to AID activity, and this has been associated with highly transcribed, topologically complex super-enhancers and regulatory clusters, demonstrating the involvement of these genomic components in AID recruitment (Qian et al. 2014; Meng et al. 2014).

The first appearance of AID is in sharks (Conticello et al. 2005; Conticello 2008), the earliest vertebrate known to have all pivotal components of the modern adaptive immune system, including RAG genes, V(D)J recombination, and major

histocompatibility complex (MHC) proteins. Putative cytidine deaminases have also been identified in the lamprey, a non-jawed vertebrate with a primordial form of adaptive immune receptors, suggesting they predate the appearance of AID (Rogozin et al. 2007; Tasumi et al. 2009; Deng et al. 2010). AID and SHM are found in bony fish, while CSR does not make an appearance until amphibians, suggesting that SHM is more ancient. It follows that the extreme C-terminal portion of AID interacts with CSR-specific factors (Barreto et al. 2003; Ta et al. 2003) and the amino acid residues within this region are divergent in bony fish.

In humans, AID is a small protein of 198 amino acids which folds into a single globular domain containing a central β-sheet that is flanked by α-helices and loops (Fig. 4a). AID contains an N-terminal nuclear localization sequence (NLS) and a C-terminal nuclear export signal (NES), and the heart of the enzyme is formed by conserved HXE and SWSPCXXC motifs which together bind the catalytic Zn^{2+} ion in the active site (Fig. 4a, b). Either directly or indirectly through other proteins, AID interacts with a number of cellular factors to regulate its localization and activity in executing SHM, CSR, and gene conversion (Reynaud et al. 2003), but the molecular basis underpinning the interactions which assemble the immunoglobulin mutasome remains an important area of discovery. Multiple crystal and solution NMR structures have been determined for the APOBEC family of enzymes in recent years, providing a framework to begin to understand the function and regulation of AID.

As seen through structural modeling, AID contains a conserved electropositive surface patch surrounding the active site cleft, as expected for interaction with ssDNA substrates (Fig. 4c). Conservation analysis reveals two additional surface patches on AID of potential biological significance, one engulfing the proximal C-terminal α-helix and surrounding β-sheet (Fig. 4d). The phosphorylation of AID at Ser38 by protein kinase A (PKA) has been implicated in regulating CSR (Basu et al. 2005, 2009) and in some instances SHM. This PKA phosphorylation site is absent in fish AID, but a neighboring acidic residue, unique to the fish, may represent a constitutive phosphomimetic. In a structural model of human AID, Ser38 is found on an exposed loop distal to the active site, where six independent experimental coordinates available for the related APOBEC3G enzyme (~45 % sequence identity) reveal an incredible structural plasticity (Fig. 4b). In APOBEC3G, the conformation of the "Ser38 loop" appears critical for positioning the HXE and SWSPCXXC motifs within the active site. It is therefore possible that phosphorylation of Ser38 by PKA serves as a direct allosteric modulator of AID activity. As with any posttranslational modification that impacts protein conformational dynamics, these may also directly impact other protein–protein interactions across the surface of AID. The field is clearly awaiting an in-depth structural and biophysical characterization of AID, particularly in complex with immunologically relevant targeting and modulatory factors.

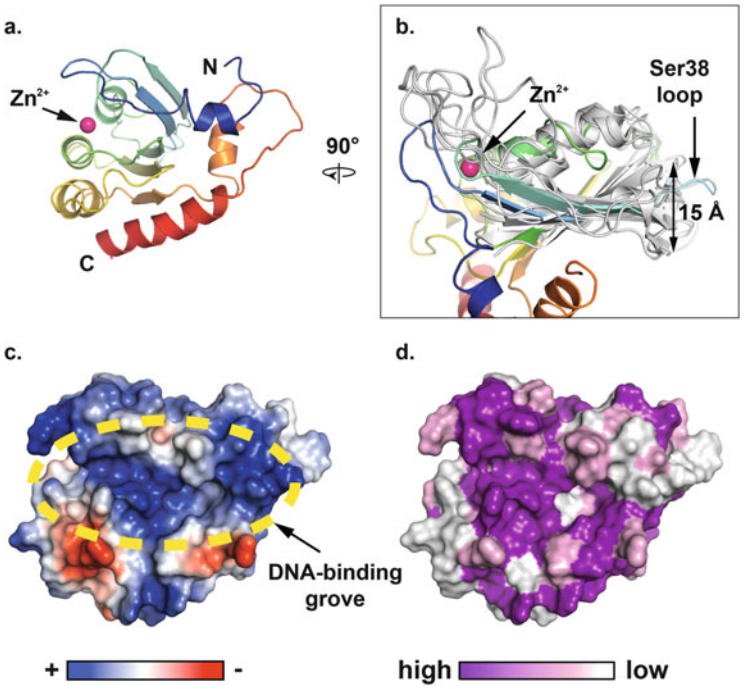


Fig. 4 Structural model of human AID based on its homology to ancient APOBEC3. (**a**) Cartoon rendering highlighting the localization of the essential Zn^{2+} ion in the catalytic site of a homology model of human AID built using the APOBEC2 structure as a template (PDB ID: 2NYT). (**b**) Available X-ray and NMR structures of APOBEC3G superimposed onto the human AID model (PDB IDs: 2JYW, 2KBO, 2KEM, 3E1U, 3IR2). In APOBEC3G, structural transitions observed in the Ser38 "phosphorylation" loop appear to be coupled to structural changes within the active site that could directly regulate enzyme activity. (**c**) Electrostatic surface potential of human AID homology model reveals a large electropositive cleft primed to bind ssDNA substrates. Coloring is as follows: *blue*, basic regions; *red*, acidic regions; *white*, hydrophobic regions. Orientation is the same as in part (**a**). (**d**) Conservation of AID across species shows a region of high sequence conservation mapped onto the active site and presumed substrate binding cleft. The surrounding regions of sequence divergence could represent functionally relevant targeting or modulatory sites on the enzyme. Coloring is as follows: *purple*, high conservation; *pink*, good conservation; *white*, low conservation. Orientation is the same as in parts (**a**) and (**c**)

5 Conserved or Diversified Biological Functions of Antibody Isotypes in Mammals

5.1 IgM

IgM is the primordial Ig from which mammalian immunoglobulins other than IgD are descended. IgM is the first isotype expressed during B cell development, forming a pool of "natural" antibodies that constitute the first line of defense during

infections. Surface-bound IgM cross-linked by its cognate antigen serves as the B cell antigen receptor. Secreted IgM exists as pentamers or, less often, hexamers that are connected via a J chain. The major function of IgM antibodies is to activate the classical pathway of complement and effectively opsonize various bacterial pathogens, as well as cause direct lysis through the assembly of the membrane attack complex.

5.2 IgD

IgD is primarily coexpressed with IgM on the surface of mature B cells before antigenic stimulation and regulated by alternative splicing (Chen and Cerutti 2011). Although typical S regions are not found upstream of Cδ, direct CSR to CδIgD has been reported (Kluin et al. 1995; White et al. 1990). Secreted IgD is detected in blood, in mucosal secretions, and on the surface of innate immune effector cells such as basophils. IgD is produced both as a surface-bound antigen receptor and in secreted form. Unlike IgM, which has been stable during evolution, the structural differences among IgDs across species are quite noticeable, suggesting it might have diverse functions in vertebrates. Ancient IgW and IgD heavy chains are orthologues, and thus both IgD and IgM were present at the inception of the adaptive immune system (Ohta and Flajnik 2006). Unlike other isotypes, the function of IgD remains controversial. In mice, IgD deficiency delays affinity maturation and impacts the localization of B cells during antigen-driven immune responses (Roes and Rajewsky 1993). Recent evidence suggests that secreted IgD enhances immune surveillance by activating antimicrobial, proinflammatory, and B cell-stimulating programs in basophils (Chen et al. 2009).

5.3 IgG

IgG is an abundant isotype in the blood and extravascular compartments. It is highly stable in circulation and has a half-life of about 3 weeks, making it suitable for a variety of bioengineering and therapeutic applications (see below). There are various subclasses of IgG in humans and other species. Common functions of IgG include crossing the placenta and the establishment of the memory humoral immune response. IgG antibodies play an important role in clearing viral and bacterial pathogens and neutralizing toxins. This isotype is also important in direct complement-mediated lysis (e.g., *Neisseria meningitis*) and can also mediate opsonization and destruction within phagosomes (e.g., *Streptococcus pneumoniae*).

There are two general classes of IgG receptors, known as FcγRs: the activating receptors (ITAM-containing) and the inhibitory receptors (ITIM-containing), which usually function in concert. The dual inhibitory and activating function of FcγRs sets the threshold for the magnitude of effector cell responses. The FcγRs

serve as a bridge between adaptive and innate immune responses. FcγRs are expressed on monocytes, macrophages, DCs, neutrophils, NK, NKT, and B cells. Upon engagement, these receptors initiate phagocytosis, antibody-dependent cell-mediated cytotoxicity (ADCC), and cytokine release. FcγRIIb is an inhibitory receptor, and its co-engagement with other FcγRs results in inhibition of FcγR-mediated signaling. In the absence of FcγRIIb, mice are prone to autoimmunity due to a higher frequency of autoreactive B cells. In addition, polymorphisms associated with a lower expression of FcγRIIb are risk alleles for developing autoimmunity (Nimmerjahn and Ravetch 2008; Tsuchiya and Kyogoku 2005).

5.4 IgE

IgE is the least abundant isotype and is involved in immunity to helminthes and allergic diseases such as asthma, allergic rhinitis, and atopic dermatitis. In addition, IgE mediates type I hypersensitivity reactions, which include both systemic and localized anaphylaxis. IgE has two major receptors, "high-affinity" FcεRI and FcεRII (CD23). FcεRI is predominantly expressed on mast cells and basophils, and its activation mediates cellular degradation and cytokine production. Upon ligation with multivalent antigens, FcεRI transduces signals that lead to the production of proinflammatory mediators such as histamine, which is associated with IgE-mediated pathology as manifested in asthma, allergic rhinitis, and anaphylaxis. FcεRII regulates IgE production and facilitates antigen processing and presentation on B cells. FcεRII on macrophages or epithelial cells facilitates the uptake of IgE-antigen complexes.

5.5 IgA

IgA maintains mucosal homeostasis by controlling host interactions with the microbiota as well as protecting the host from pathogens and their products. The human digestive tract can be considered the largest immune organ in the body and is continuously challenged by microbiota and dietary antigens (Xu and Gordon 2003). IgA is a predominant class in mucosal secretions such as intestine, breast milk, saliva, tears, colostrum, and the mucus of the bronchial, genitourinary, and digestive tracts. Unlike other isotypes of the immune system, IgA is predominantly considered noninflammatory. IgA binds a wide array of pathogenic bacteria, viruses, and toxins and tethers them to mucus, a process known as immune exclusion (Strugnell and Wijburg 2010). IgA can also be transported from the lumen to the basolateral side of intestinal epithelial cells (IECs), bringing antigens along with it for presentation to the headquarters of the mucosal immune system, the gut-associated lymphoid tissue (GALT). IgA is thought to have evolved from

amphibian IgX based on phylogeny, genomic synteny, and function (Mashoof et al. 2013) (Figs. 1 and 2).

In most mammals, IgA has three immunoglobulin domains (Cα1, Cα2, Cα3) connected by a hinge region to the variable domain. In serum, it exists as a monomer, although polymeric (dimer, trimer, tetramer) forms can occur at lower levels. Mucosal IgA, in contrast, is predominantly a dimer, formed when two antibodies are covalently linked end to end by a J-chain polypeptide with variable regions facing outward (Bastian et al. 1992; Krugmann et al. 1997). Dimeric IgA binds to a receptor known as polymeric immunoglobulin receptor (pIgR) present on the basolateral membrane of IECs. This dimeric IgA–pIgR complex is internalized by receptor-mediated endocytosis and transported to the luminal surface. The shuttling pIgR is then enzymatically cleaved, leaving a portion of itself complexed to the now soluble IgA (SIgA) dimeric antibody. This pIgR fragment is termed secretory component (SC). SC protects and extends the half-life of SIgA in the mucosa. SC also confers hydrophilic properties to SIgA, helping to retain it to the mucus environment. In effect, SC keeps SIgA in a location where it interacts with microbiota and pathogens and prevents its diffusion into the lumen (Deplancke and Gaskins 2001; Johansson et al. 2011).

6 Regulation of IgA Production

It has long been appreciated that IgA production is tightly linked to bacterial colonization of the gut. Mice raised in germ-free conditions produce little to no mucosal IgA (Benveniste et al. 1971a, b; Moreau et al. 1978), and inoculation of germ-free mice with certain strains of bacteria is capable of inducing high-titer IgA responses (Hapfelmeier et al. 2010; Geuking et al. 2012). Research into (1) how bacteria communicate with the gut, (2) where, anatomically, does CSR to IgA occur, (3) whether there are T cell-dependent (TD) and T cell-independent (TI) pathways, and (4) what the supporting cell types are has made great strides in recent years.

GALT is organized into distinct structures known as Peyer's patches (PPs), isolated lymphoid follicles (ILFs), and the lamina propria (LP) (Fig. 5) (Suzuki et al. 2010). Peyer's patches are prominent structures that form during embryogenesis. They contain multiple (>5) B cell follicles and comprise the major sources of T cell-dependent IgA synthesis. ILFs, on the other hand, have only one B cell follicle and can generate IgA in a T cell-independent manner (Tsuji et al. 2008). Unlike PPs, ILFs develop postnatally. This was originally believed to be in concert with bacterial colonization (Hamada et al. 2002; Pabst et al. 2006), but new evidence suggests diet plays a role (Kiss et al. 2011; Lee et al. 2012). Together, PPs and ILFs are considered "inductive sites," where most IgA immune responses are initiated through interactions with commensal bacteria. The lamina propria (LP) is a connective tissue underlying the mucosal epithelium. Plasmablasts that

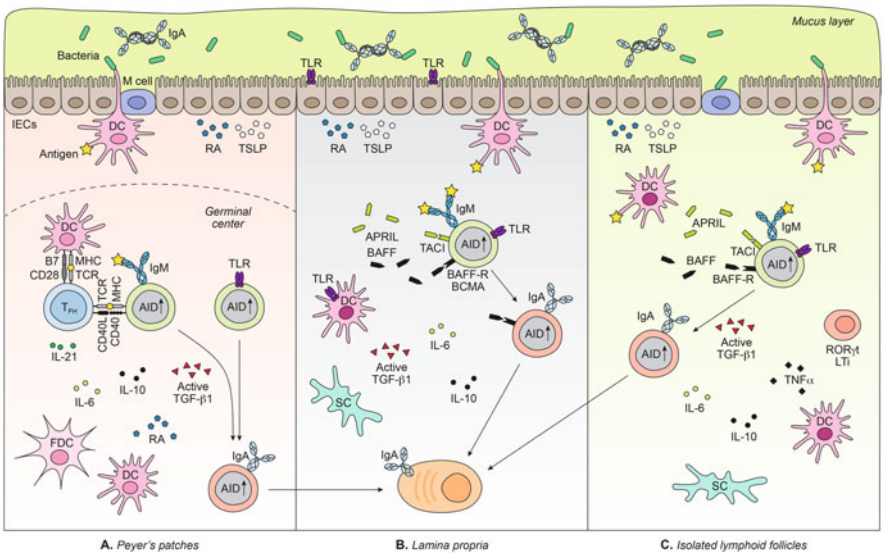


Fig. 5 IgA regulation within gut-associated lymphoid tissues (GALT). (**a**) Peyer's patches: intestinal epithelial cells (IECs) condition underlying dendritic cells (DCs) by releasing thymic stromal lymphopoietin (TSLP) and retinoic acid (RA). M cells and CX3CR1^{+} DCs transport microbial products from the lumen, which are conveyed to germinal center T cells by CD103^{+} DCs. Additional subsets of DCs contribute IL-6, IL-10, TGF-β, and RA, which polarize T cells toward Th2 responses and promote IgA CSR. Inside the germinal center, stromal cells release BAFF, APRIL, and TGF-β1, enhancing IgA CSR and promoting T_{FH} function. Once activated, B cells upregulate AID and switch to Cα. Activated B cells differentiate into plasmablasts within the germinal centers, moving into the lamina propria for terminal differentiation into IgA-expressing plasma cells. (**b**) Lamina propria: DCs respond to bacterial products through TLR signaling and release IL-6, RA, BAFF, and APRIL, which support B cell activation and differentiation. IECs also release BAFF and APRIL in a TLR-dependent manner. An IgA-secreting plasma cell is denoted in *light orange*. (**c**) Isolated lymphoid follicles: Similar to the LP, several DC types and stromal cells (SC) are thought to contribute signals conducive to IgA CSR. Subepithelial CX3CR1^{+} DCs capture and present TI antigens to local follicular B cells. CD11c^{+} DCs produce TNF, inducing matrix metalloproteinases that process precursor TGF-β1 into its mature form. These DCs can also produce BAFF and APRIL. Stromal cells can also produce BAFF in addition to molecules that attract B cells and DCs to the gut

have moved out into the lamina propria and matured into plasma cells are considered the "effector sites" of the GALT (Macpherson et al. 2008).

6.1 *T Cell-Dependent and T Cell-Independent Locations of IgA CSR*

Determining where actual switching to IgA occurs (PP, ILF, or LP) and whether it requires T cell help has been a subject of intense study. Years ago, it was proposed

that PPs were an enriched source of precursors of IgA-producing cells (Craig and Cebra 1971). These structures were hypothesized to "shed" lymphoid cells into circulation, which would "seed" the gut, proliferate, and differentiate into IgA-producing cells. While IgA CSR can take place within PPs, recent evidence suggests these organized lymphoid structures are unnecessary for the gut to be populated with IgA^+ plasma cells.

Peyer's patches are perhaps the best-studied component of the mucosal immune response (Fig. 5a). It is here that T cell-dependent reactions occur to produce high-affinity IgA. Under steady-state conditions, IECs release thymic stromal lymphopoietin (TSLP) and retinoic acid (RA) in response to TLR signals, factors that "condition" underlying DCs to be noninflammatory. The road to IgA synthesis in one of these structures begins when "snorkeling" $CX3CR1^+$ DCs capture bacterial antigens directly from the lumen (Bogunovic et al. 2009; Varol et al. 2009), or M cells convey them via transcytosis (Mabbott et al. 2013). Neither M cells nor $CX3CR1^+$ DCs can migrate to interfollicular areas to present antigens to T cells themselves; instead, a migratory $CD103^+$ DC subtype may act as a courier (Schulz et al. 2009; Bogunovic et al. 2009; Varol et al. 2009; Coombes et al. 2007). Additional DC subsets are found in and around GCs (Rescigno and Di Sabatino 2009; Salazar-Gonzalez et al. 2006; Sun et al. 2007) that together secrete a cocktail of factors (TGF-β, IL-10, IL-6, IL-21, and RA) that promote IgA responses.

Inside the GC, follicular dendritic cells (FDCs) form a scaffold for B and T cell interactions. These stromal cells can release BAFF, APRIL, and large amounts of TGF-β1 that act to enhance IgA CSR and promote T_{FH} function (Suzuki et al. 2010). Tregs in the GC make IL-10, TGF-β1, and CD40L, which drive CSR to IgA and repress inflammatory Th1 responses (Cong et al. 2009). Tregs can also respond to mucosal GCs by downregulating Foxp3 and turning into T_{FH} cells (Tsuji et al. 2009). Once activated, B cells make Iα transcripts, upregulate AID, and switch to Cα, leaving behind transient byproducts of the reaction including Sα–Sμ switch circles and Iα–Cμ circle transcripts. Activated B cells differentiate into plasmablasts within the germinal centers, subsequently moving into the lamina propria for terminal differentiation into IgA-expressing plasma cells (Tarlinton et al. 2008; Yoshida et al. 2010), or, as will be discussed next, bypass GC reactions and switch to IgA within the lamina propria itself.

The first indication that IgA switching could happen outside of follicular structures such as PPs and ILFs came when hallmarks of CSR to Cα, such as AID expression, Sα–Sμ switch circles, and Iα–Cμ circle transcripts, were detected in LP tissue (Fig. 5b) (Fagarasan et al. 2001). These findings were corroborated by others (He et al. 2007; Crouch et al. 2007). However, not every group was able to detect these products in the LP (Barone et al. 2009; Bergqvist et al. 2006, 2010), raising the question of whether the results were simply due to ILF contamination. The LP switching hypothesis was shored up by studies in genetic backgrounds that lack PPs and ILFs, yet still show intestinal IgA production and signs of CSR in the LP (Kang et al. 2002; He et al. 2007; Crouch et al. 2007). Taken together, these findings demonstrate that a visit to a germinal center is not necessary for a B cell in the gut to

switch to IgA. In fact, new evidence suggests the LP may even harbor developing B cells undergoing V(D)J recombination. This would likely be a mechanism to specifically control the gut Ig repertoire and suggests that early B cell development may not be limited to the bone marrow (Wesemann et al. 2013).

Multiple lines of evidence suggest LP switching happens in a T cell-independent manner. The key ingredients for T-independent switching likely involve the activities of several types of dendritic cells with the IECs (He et al. 2007; Tsuji et al. 2008; Tezuka et al. 2007; Shang et al. 2008; Uematsu et al. 2008; Xu et al. 2007, 2008; Uematsu and Akira 2009). DCs respond to bacterial products through TLR signaling and release factors supportive of B cell activation and differentiation, such as RA, IL-6, and B cell factors BAFF and APRIL. IECs are likewise capable of releasing BAFF and APRIL via TLR recognition of bacteria (He et al. 2007; Xu et al. 2007). APRIL is particularly effective at inducing IgA2, the dominant IgA isotype in the distal intestine, when coupled with TLR ligands like flagellin, and can do so in a CD40L-independent manner (He et al. 2007). Also, IgA+ plasma cells secrete iNOS and TNF-α, additional drivers of de novo IgA production (Fritz et al. 2012). Interestingly, evolutionary studies suggest these T cell-independent mechanisms may be ancient ways for regulating mucosal defense. In *Xenopus*, for example, mucosal IgX levels do not change when T cells are ablated via larval thymectomy (Turner and Manning 1974; Mashoof et al. 2013).

ILFs are also believed to induce IgA CSR without T cell help (Fig. 5c). In $Tcrb^{-/-}Tcrd^{-/-}$ mice, which lack T cells, ILFs were found to contain numerous IgA^+ B cells and plasmablasts (Tsuji et al. 2008). Additionally, in a RORγt KO model where lymphoid tissue-inducer (LTi)-like cells were injected, ILFs, but not PPs, developed. In these mice, $CD4^+$ T cells in the LP were greatly reduced, yet B cells were able to populate the ILFs and express AID, and gut IgA levels approached that of WT mice. Similar to the LP, in the absence of T cells, several DC types and stromal cells are thought to contribute signals conducive to IgA CSR. Subepithelial $CX3CR1^+$ DCs have been observed in ILFs (Hamada et al. 2002) that may capture and present T cell-independent antigens to local follicular B cells. $CD11c^+$ DCs present in ILFs are capable of producing TNF, which is a potent inducer of matrix metalloproteinases (MMPs)-9 and -13, which process precursor TGF-β1 into its mature form. These DCs can also produce BAFF and APRIL in a manner that is enhanced by TLR ligands (Tsuji et al. 2008; Suzuki et al. 2010). Coculture of stromal cells (SCs) with LTi cells caused BAFF upregulation as well as key molecules for B cell migration to gut and chemokines associated with DC migration to gut, and expression was enhanced by bacterial products (Tsuji et al. 2008).

7 Utility and Applications

The practical applications gained from our understanding of Ig regulation, structure, and functions are numerous, ranging from research tools to diagnostic and therapeutic applications (Fig. 6). Most prominent is the development of monoclonal

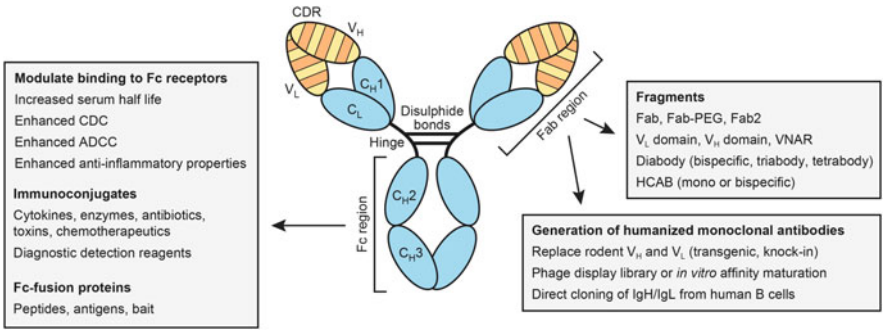


Fig. 6 Utility of antibodies and their domains. Shown is a subset of modifications to the traditional antibody to enhance functionality. *Left panel*: Fc regions can be mutated to enhance or reduce their binding affinities for Fc receptors, conjugated to various proteins or chemical moieties in order to deliver them to target cells, or isolated from the Fab and fused to various proteins. *Right upper panel*: Fab domains can be used in isolation with polyethylene glycol (PEG) added for stability, separated into V_L and V_H domains, or joined together to form bi-, tri-, and tetrabodies with multiple specificities. HCAB refers to heavy chain-only antibody (Janssens et al. 2006). *Right lower panel*: The antigen-binding pocket can be modified to have higher affinity and specificity by replacing the natural sequences with those derived from directly cloned human sequences or phage display libraries

antibodies (mAbs) for the treatment of disease that function to block or activate signaling pathways, deplete pathogenic cell types, or clear infectious agents. Several mAbs are currently used in the clinic for the treatment of various conditions, ranging from cancer to infectious and inflammatory diseases, and hundreds more are at various stages of development. Bispecific antibodies are also emerging as therapeutic modalities offering better therapeutic efficacy. The various properties of Ig molecules have been prodded and tweaked to create therapeutics with increased or decreased half-lives, muted or enhanced effector functions, multivalent antigen interactions, and smaller sizes to penetrate tumors. Other aspects of Ig regulation, such as the mutagenic properties of AID, have been used to enhance commercial antibody affinities, and switch region swapping has been used to create a hyper-IgE mouse model for the study of asthma and allergies. It is important to mention that in the coming years, the analysis of V(D)J and SHM/CSR repertoires, combined with human genome sequencing across various disease indications, will rapidly advance, helping to generate therapeutic and diagnostic hypotheses for a variety of human diseases.

7.1 Generation of Monoclonal Antibodies

Since the discovery of the Nobel prize-winning hybridoma technology for isolating mAbs (Kohler and Milstein 1975), methodologies to generate mAbs have advanced substantially and are still rapidly evolving. Various strategies have been adopted to generate mAbs in vivo by taking advantage of the de novo immune responses using

animal models amenable to genetic engineering such as mouse or rat. After immunization, various strategies such as hybridoma fusion technology or single cell cloning are utilized to obtain a panel of antigen-specific mAbs. Mouse mAbs have been widely used in research, but unfortunately have a low success rate in the clinic (Reichert et al. 2005). This is because they are highly immunogenic, resulting in rapid clearance, reduced efficacy, and an increased risk of adverse reactions ranging from rash to anaphylactic-like symptoms (Lonberg 2005). Furthermore, mouse mAbs react only weakly with human FcγRs, resulting in inefficient effector function, and do not bind the human salvage receptor FcRn (Ober et al. 2001), resulting in a less than 20 h half-life (Carter 2001; Presta 2002). These major hurdles were overcome by transgenic rodent models in which human IgH or IgL V regions are stitched together with mouse or rat constant regions with its known regulatory elements (Fig. 5) (Morrison et al. 1984; Boulianne et al. 1984; Green et al. 1994; Lonberg et al. 1994; Lonberg 2005). Various combinations of V and C regions are utilized to optimize these models. To avoid competition with endogenous loci, these transgenic models are often generated in an IgH/IgL knockout background. One drawback, however, of transgenic humanized rodents seems to be that components of B cell development and/or AID-dependent diversifications processes are suboptimal (Green and Jakobovits 1998; Pruzina et al. 2011). This could be due to reduced B cell numbers, defects in B cell development, a limited V_H repertoire, mis-regulation of the transgene, or a lack of other human compatible components to invoke a robust immune response. Aside from transgenic models, the generation of mouse models with the replacement of endogenous mouse V_H IgH or V_L loci with their human counterparts appears to be more efficient for producing mAbs.

mAbs can also be generated by technology in which human Abs are displayed on the surface of a simple organism such as phage, bacteria, or yeast. This approach has proved successful; however, it is labor intensive and repertoires are limited by the capacity of the library. mAbs have also been directly isolated from humans. This strategy has been effectively utilized to identify mAbs against various pathogens.

7.2 Variables Affecting Heavy Chain Choice

The specific heavy chain used in therapeutics is selected based on desired properties such as half-life, effector function, and valency. IgG1 is the isotype used in the majority of therapeutics (Salfeld 2007). IgG1 has a high affinity for FcγRs, which allows it to activate processes termed antibody-dependent cell-mediated cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC). ADCC results when an antibody binds to an undesirable target cell and is recognized by FcγIIIRA on the surface of NK cells, monocytes, and macrophages (Congy-Jolivet et al. 2007; Strohl 2009; Biburger et al. 2011), which subsequently destroy the cell. By introducing mutations to the IgG1 Fc domain, its affinity for Fc receptors can be

increased, thereby enhancing target cell clearance. This is particularly useful in tumor therapy. The best evidence showing clinical efficacy mediated by an Fc domain comes from studies of rituximab in patients with non-Hodgkin's lymphoma. It was found that patients carrying an FcγRIII polymorphism responded to the drug, strongly implicating ADCC as its mechanism of action (Cartron et al. 2002; Weng and Levy 2003).

CDC occurs when antibodies directly bound to cell surface antigens are recognized by complement C1q, triggering cell death via the complement cascade (Idusogie et al. 2001; Meyer et al. 2014). In some cases, this can be counteracted by tumors, which secrete complement regulatory proteins (Treon et al. 2001). Outside tumor therapy, activating complement is not considered ideal for many applications, such as neutralizing cytokines. Inappropriate activation of FcR-expressing cells can actually cause cytokine release and associated toxicities. In these cases, mutations can be introduced to IgG1 to lower its propensity to activate complement. IgG2 and IgG4 Fc domains can also be substituted for IgG1 since they have lower affinities for complement, a feature that has made them appealing for drug development recently (Bruhns et al. 2009).

Isotypes other than IgG1/2/4 are less common but have their own selling points. IgM, for example, is a natural adjuvant in vaccines and forms polymers that may mimic the adjuvant "depot effect" (Czajkowsky et al. 2010; Harte et al. 1983; Stager et al. 2003). IgM also binds C1q, which may make it suitable for complement-dependent cytotoxicity. The IgA isotype, with its affinity for mucosal surfaces, may be useful in preventing infections via the nasal cavity and gastrointestinal tract (Corthesy 2003). In these cases, IgA antibody would be applied directly to vulnerable mucosal surfaces, which is known as passive immunization. Many studies in humans and animals using passive immunization have demonstrated that administering antibodies orally, intranasally, via the intrauterine route, or by lung instillation can prevent, reduce, or even cure viral and bacterial infections. Due to the difficulty in generating SIgA with a J chain and SC, many of these studies use monomeric IgG, IgY, or a combination of IgG and IgA. Thus, there is room for improving these reagents, and greater efficacy may result from using complete SIgA.

7.3 Antibody Domains and Fragments

Worth mentioning is the notion that a heavy chain may not be desirable, or even necessary, in drug development. Traditional antibodies are quite bulky (~150–180 kDa) and have difficulty penetrating tumors and tissues, or targeting antigens in recessed spaces such as enzyme active sites. Researchers are now dismembering antibodies into their component parts, forming a toolkit from which to assemble therapeutics with exciting properties (Fig. 6). These "domain antibodies" come in numerous forms: fragment antigen binding (Fab, ~55 kDa), Fab_2 bispecific (~110 kDa), Fab_3 trispecific (~165 kDa), single chain variable fragment (scFv,

~28 kDa), scFv in diabody (~50 kDa), triabody (~75 kDa), and tetrabody (~100 kDa) formats and even the most minimal V_H and V_L domains (~15 kDa each) (Weir et al. 2002; Holliger and Hudson 2005). One issue with isolated V domains, however, is they can have problems with aggregation, poor solubility, and reduced half-life in circulation. The covalent attachment of polymers such as polyethylene glycol (PEG) to Fab regions is one way to alleviate some of these issues. Another problem is that these fragments rarely display affinities close to their parent antibodies. An interesting solution to this problem was provided by comparative biology. Certain antibodies from camelids (camels, llamas) and cartilaginous fish (wobbegong, nurse sharks) do not have light chains, so one would predict their single heavy chain V domain (termed VhH in camelids and V-NAR in sharks) is structured in such a way as to confer high antigen specificity and affinity (De Genst et al. 2004; Dooley and Flajnik 2005; Streltsov et al. 2004). Indeed, the CDR3 region of these V domains, where most of the domain's diversity is concentrated, is elongated and stabilized by disulfide bonds. These small camelid VhH domains have the potential to recognize uncommon epitopes such as clefts on antigen surfaces. Future therapies using minimal V domains will probably be based on these unique species.

Humanized heavy chain-only antibodies, known as Fc-fusion proteins, are also emerging as platforms to generate mAbs with unique properties and come in a variety of formats (bivalent monospecific, tetravalent bispecific). These proteins have the potential for broad clinical and nonclinical applications since their first description as CD4-Fc antagonists (Czajkowsky et al. 2012). Fc-fusion proteins serve as well-behaved (stably and independently folding) modules to which many different proteins (e.g., ligands, antigens, baits) can be attached and remain functional.

7.4 Other Applications

Information gleaned about the mechanistic basis for CSR has been applied in other ways. For example, AID, a potent mutagen, has been used by several groups to enhance the affinity of commercial antibodies (King et al. 2014). In this method, an antibody library is expressed in mammalian cells (B cell or non-B cell origin), and antigen-coupled magnetic beads are used to capture those bearing Ag-specific antibodies. AID is then transfected into the selected cells, recapitulating SHM in vitro, and clones bearing enhanced affinity Abs are screened for and retained. This process has been used to evolve non-Ig proteins as well. Expression of AID with a fluorescent protein in Ramos cells endowed it with increased photostability and a far-red emission.

Even switch regions are finding an application. Recently, these highly repetitive regions of DNA were utilized to create a mouse that expresses high levels of IgE (Misaghi et al. 2013). IgE is a major driver of allergic reactions and asthma because it activates mast cells and basophils via FcεRI. As a result, switching to Cε has been

regulated over the course of evolution, in part by limiting the size of Sε to a relatively small region (~0.8 kb) that contains relatively few AID hotspots. Replacement of Sε with Sμ, which is much larger in size and very rich in deamination motifs, resulted in greatly enhanced switching to the epsilon locus. These Sμ knock-in (Sμ-KI) mice displayed greatly elevated IgE levels upon infection, which could make them useful in the study of allergic responses, asthma, and FcεRI/II function. Humanized IgH constant regions will also allow for efficient generation of selected and humanized isotypes in an efficient manner (Duchez et al. 2010).

References

Amemiya CT, Haire RN, Litman GW (1989) Nucleotide sequence of a cDNA encoding a third distinct *Xenopus* immunoglobulin heavy chain isotype. Nucleic Acids Res 17:5388

Barone F, Patel P, Sanderson JD, Spencer J (2009) Gut-associated lymphoid tissue contains the molecular machinery to support T-cell-dependent and T-cell-independent class switch recombination. Mucosal Immunol 2:495–503

Barreto V, Reina-San-Martin B, Ramiro AR, Mcbride KM, Nussenzweig MC (2003) C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Mol Cell 12:501–508

Bastian A, Kratzin H, Eckart K, Hilschmann N (1992) Intra- and interchain disulfide bridges of the human J chain in secretory immunoglobulin A. Biol Chem Hoppe Seyler 373:1255–1263

Basu U, Chaudhuri J, Alpert C, Dutt S, Ranganath S, Li G, Schrum JP, Manis JP, Alt FW (2005) The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature 438:508–511

Basu U, Franklin A, Alt FW (2009) Post-translational regulation of activation-induced cytidine deaminase. Philos Trans R Soc Lond B Biol Sci 364:667–673

Benveniste J, Lespinats G, Adam C, Salomon JC (1971a) Immunoglobulins in intact, immunized, and contaminated axenic mice: study of serum IgA. J Immunol 107:1647–1655

Benveniste J, Lespinats G, Salomon J (1971b) Serum and secretory IgA in axenic and holoxenic mice. J Immunol 107:1656–1662

Bergqvist P, Gardby E, Stensson A, Bemark M, Lycke NY (2006) Gut IgA class switch recombination in the absence of CD40 does not occur in the lamina propria and is independent of germinal centers. J Immunol 177:7772–7783

Bergqvist P, Stensson A, Lycke NY, Bemark M (2010) T cell-independent IgA class switch recombination is restricted to the GALT and occurs prior to manifest germinal center formation. J Immunol 184:3545–3553

Biburger M, Aschermann S, Schwab I, Lux A, Albert H, Danzer H, Woigk M, Dudziak D, Nimmerjahn F (2011) Monocyte subsets responsible for immunoglobulin G-dependent effector functions in vivo. Immunity 35:932–944

Boehm MK, Woof JM, Kerr MA, Perkins SJ (1999) The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. J Mol Biol 286:1421–1447

Bogunovic M, Ginhoux F, Helft J, Shang L, Hashimoto D, Greter M, Liu K, Jakubzick C, Ingersoll MA, Leboeuf M, Stanley ER, Nussenzweig M, Lira SA, Randolph GJ, Merad M (2009) Origin of the lamina propria dendritic cell network. Immunity 31:513–525

Boulianne GL, Hozumi N, Shulman MJ (1984) Production of functional chimaeric mouse/human antibody. Nature 312:643–646

Bransteitter R, Pham P, Calabrese P, Goodman MF (2004) Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J Biol Chem 279:51612–51621

Bransteitter R, Pham P, Scharff MD, Goodman MF (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci USA 100:4102–4107

Bruhns P, Iannascoli B, England P, Mancardi DA, Fernandez N, Jorieux S, Daeron M (2009) Specificity and affinity of human Fcgamma receptors and their polymorphic variants for human IgG subclasses. Blood 113:3716–3725

Carter P (2001) Improving the efficacy of antibody-based cancer therapies. Nat Rev Cancer 1:118–129

Cartron G, Dacheux L, Salles G, Solal-Celigny P, Bardos P, Colombat P, Watier H (2002) Therapeutic activity of humanized anti-CD20 monoclonal antibody and polymorphism in IgG Fc receptor FcgammaRIIIa gene. Blood 99:754–758

Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature 422:726–730

Chen K, Cerutti A (2011) The function and regulation of immunoglobulin D. Curr Opin Immunol 23:345–352

Chen K, Xu W, Wilson M, He B, Miller NW, Bengten E, Edholm ES, Santini PA, Rath P, Chiu A, Cattalini M, Litzman JJ, Huang B, Meini A, Riesbeck K, Cunningham-Rundles C, Plebani A, Cerutti A (2009) Immunoglobulin D enhances immune surveillance by activating antimicrobial, proinflammatory and B cell-stimulating programs in basophils. Nat Immunol 10:889–898

Cheng G, Gao Y, Wang T, Sun Y, Wei Z, Li L, Ren L, Guo Y, Hu X, Lu Y, Wang X, Liu G, Zhang C, Yu J, Pan-Hammarstrom Q, Hammarstrom L, Wu X, Li N, Zhao Y (2013) Extensive diversification of IgH subclass-encoding genes and IgM subclass switching in crocodilians. Nat Commun 4:1337

Chevenet F, Brun C, Banuls AL, Jacq B, Christen R (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7:439

Cong Y, Feng T, Fujihashi K, Schoeb TR, Elson CO (2009) A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. Proc Natl Acad Sci USA 106:19256–19261

Congy-Jolivet N, Probst A, Watier H, Thibault G (2007) Recombinant therapeutic monoclonal antibodies: mechanisms of action in relation to structural and functional duality. Crit Rev Oncol Hematol 64:226–233

Conticello SG (2008) The AID/APOBEC family of nucleic acid mutators. Genome Biol 9:229

Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol 22:367–377

Coombes JL, Siddiqui KR, Arancibia-Carcamo CV, Hall J, Sun CM, Belkaid Y, Powrie F (2007) A functionally specialized population of mucosal CD103+ DCs induces Foxp3+ regulatory T cells via a TGF-beta and retinoic acid-dependent mechanism. J Exp Med 204:1757–1764

Corthesy B (2003) Recombinant secretory immunoglobulin A in passive immunotherapy: linking immunology and biotechnology. Curr Pharm Biotechnol 4:51–67

Craig SW, Cebra JJ (1971) Peyer's patches: an enriched source of precursors for IgA-producing immunocytes in the rabbit. J Exp Med 134:188–200

Crouch EE, Li Z, Takizawa M, Fichtner-Feigl S, Gourzi P, Montano C, Feigenbaum L, Wilson P, Janz S, Papavasiliou FN, Casellas R (2007) Regulation of AID expression in the immune response. J Exp Med 204:1145–1156

Czajkowsky DM, Hu J, Shao Z, Pleass RJ (2012) Fc-fusion proteins: new developments and future perspectives. EMBO Mol Med 4:1015–1028

Czajkowsky DM, Salanti A, Ditlev SB, Shao Z, Ghumra A, Rowe JA, Pleass RJ (2010) IgM, Fc mu Rs, and malarial immune evasion. J Immunol 184:4597–4603

Danilova N, Bussmann J, Jekosch K, Steiner LA (2005) The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat Immunol 6(3):295

De Genst E, Handelberg F, Van Meirhaeghe A, Vynck S, Loris R, Wyns L, Muyldermans S (2004) Chemical basis for the affinity maturation of a camel single domain antibody. J Biol Chem 279:53593–53601

Deng L, Velikovsky CA, Xu G, Iyer LM, Tasumi S, Kerzic MC, Flajnik MF, Aravind L, Pancer Z, Mariuzza RA (2010) A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey. Proc Natl Acad Sci USA 107:13408–13413

Deplancke B, Gaskins HR (2001) Microbial modulation of innate defense: goblet cells and the intestinal mucus layer. Am J Clin Nutr 73:1131S–1141S

Deza FG, Espinel CS, Beneitez JV (2007) A novel IgA-like immunoglobulin in the reptile *Eublepharis macularius*. Dev Comp Immunol 31:596–605

Dickerson SK, Market E, Besmer E, Papavasiliou FN (2003) AID mediates hypermutation by deaminating single stranded DNA. J Exp Med 197:1291–1296

Dooley H, Flajnik MF (2005) Shark immunity bites back: affinity maturation and memory response in the nurse shark, *Ginglymostoma cirratum*. Eur J Immunol 35:936–945

Duchez S, Amin R, Cogné N, Delpy L, Sirac C, Pascal V, Corthésy B, Cognéa M (2010) Premature replacement of μ with α immunoglobulin chains impairs lymphopoiesis and mucosal homing but promotes plasma cell maturation. Proc Natl Acad Sci U S A 107(7):3064–3069

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113

Fagarasan S, Kinoshita K, Muramatsu M, Ikuta K, Honjo T (2001) In situ class switching and differentiation to IgA-producing cells in the gut lamina propria. Nature 413:639–643

Fellah JS, Iscaki S, Vaerman JP, Charlemagne J (1992a) Transient developmental expression of IgY and secretory component like protein in the gut of the axolotl (*Ambystoma mexicanum*). Dev Immunol 2:181–190

Fellah JS, Kerfourn F, Wiles MV, Schwager J, Charlemagne J (1993) Phylogeny of immunoglobulin heavy chain isotypes: structure of the constant region of *Ambystoma mexicanum* upsilon chain deduced from cDNA sequence. Immunogenetics 38:311–317

Fellah JS, Wiles MV, Charlemagne J, Schwager J (1992b) Evolution of vertebrate IgM: complete amino acid sequence of the constant region of *Ambystoma mexicanum* mu chain deduced from cDNA sequence. Eur J Immunol 22:2595–2601

Fritz JH, Rojas OL, Simard N, Mccarthy DD, Hapfelmeier S, Rubino S, Robertson SJ, Larijani M, Gosselin J, Ivanov II, Martin A, Casellas R, Philpott DJ, Girardin SE, Mccoy KD, Macpherson AJ, Paige CJ, Gommerman JL (2012) Acquisition of a multifunctional IgA+ plasma cell phenotype in the gut. Nature 481:199–203

Furtado PB, Whitty PW, Robertson A, Eaton JT, Almogren A, Kerr MA, Woof JM, Perkins SJ (2004) Solution structure determination of monomeric human IgA2 by X-ray and neutron scattering, analytical ultracentrifugation and constrained modelling: a comparison with monomeric human IgA1. J Mol Biol 338:921–941

Gambon Deza F, Sanchez Espinel C, Magadan Mompo S (2009) The immunoglobulin heavy chain locus in the reptile *Anolis carolinensis*. Mol Immunol 46:1679–1687

Gambon-Deza F, Sanchez-Espinel C, Mirete-Bachiller S, Magadan-Mompo S (2012) Snakes antibodies. Dev Comp Immunol 38:1–9

Geuking MB, Mccoy KD, Macpherson AJ (2012) The function of secretory IgA in the context of the intestinal continuum of adaptive immune responses in host-microbial mutualism. Semin Immunol 24:36–42

Green LL, Hardy MC, Maynard-Currie CE, Tsuda H, Louie DM, Mendez MJ, Abderrahim H, Noguchi M, Smith DH, Zeng Y, David NE, Sasai H, Garza D, Brenner DG, Hales JF, Mcguinness RP, Capon DJ, Klapholz S, Jakobovits A (1994) Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. Nat Genet 7:13–21

Green LL, Jakobovits A (1998) Regulation of B cell development by variable gene complexity in mice reconstituted with human immunoglobulin yeast artificial chromosomes. J Exp Med 188:483–495

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Haire RN, Shamblott MJ, Amemiya CT, Litman GW (1989) A second *Xenopus* immunoglobulin heavy chain constant region isotype gene. Nucleic Acids Res 17(4):following 2886

Hamada H, Hiroi T, Nishiyama Y, Takahashi H, Masunaga Y, Hachimura S, Kaminogawa S, Takahashi-Iwanaga H, Iwanaga T, Kiyono H, Yamamoto H, Ishikawa H (2002) Identification of multiple isolated lymphoid follicles on the antimesenteric wall of the mouse small intestine. J Immunol 168:57–64

Han L, Masani S, Yu K (2011) Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. Proc Natl Acad Sci USA 108:11584–11589

Hansen JD, Landis ED, Phillips RB (2005) Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: implications for a distinctive B cell developmental pathway in teleost fish. Proc Natl Acad Sci 102(19):6919

Hapfelmeier S, Lawson MA, Slack E, Kirundi JK, Stoel M, Heikenwalder M, Cahenzli J, Velykoredko Y, Balmer ML, Endt K, Geuking MB, Curtiss R 3rd, Mccoy KD, Macpherson AJ (2010) Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. Science 328:1705–1709

Harte PG, Cooke A, Playfair JH (1983) Specific monoclonal IgM is a potent adjuvant in murine malaria vaccination. Nature 302:256–258

He B, Xu W, Santini PA, Polydorides AD, Chiu A, Estrella J, Shan M, Chadburn A, Villanacci V, Plebani A, Knowles DM, Rescigno M, Cerutti A (2007) Intestinal bacteria trigger T cell-independent immunoglobulin A(2) class switching by inducing epithelial-cell secretion of the cytokine APRIL. Immunity 26:812–826

Holliger P, Hudson PJ (2005) Engineered antibody fragments and the rise of single domains. Nat Biotechnol 23:1126–1136

Hsu E, Flajnik MF, Du Pasquier L (1985) A third immunoglobulin class in amphibians. J Immunol 135:1998–2004

Huang FT, Yu K, Balter BB, Selsing E, Oruc Z, Khamlichi AA, Hsieh CL, Lieber MR (2007) Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. Mol Cell Biol 27:5921–5932

Idusogie EE, Wong PY, Presta LG, Gazzano-Santoro H, Totpal K, Ultsch M, Mulkerrin MG (2001) Engineered antibodies with increased activity to recruit complement. J Immunol 166:2571–2575

Janssens R, Dekker S, Hendriks RW, Panayotou G, Van Remoortere A, San JK, Grosveld F, Drabek D (2006) Generation of heavy-chain-only antibodies in mice. Proc Natl Acad Sci USA 103:15130–15135

Johansson ME, Larsson JM, Hansson GC (2011) The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc Natl Acad Sci USA 108(Suppl 1):4659–4665

Kang HS, Chin RK, Wang Y, Yu P, Wang J, Newell KA, Fu YX (2002) Signaling via LTbetaR on the lamina propria stromal cells of the gut is required for IgA production. Nat Immunol 3:576–582

King DJ, Bowers PM, Kehry MR, Horlick RA (2014) Mammalian cell display and somatic hypermutation in vitro for human antibody discovery. Curr Drug Discov Technol 11:56–64

Kiss EA, Vonarbourg C, Kopfmann S, Hobeika E, Finke D, Esser C, Diefenbach A (2011) Natural aryl hydrocarbon receptor ligands control organogenesis of intestinal lymphoid follicles. Science 334:1561–1565

Kluin PM, Kayano H, Zani VJ, Kluin-Nelemans HC, Tucker PW, Satterwhite E, Dyer MJ (1995) IgD class switching: identification of a novel recombination site in neoplastic and normal B cells. Eur J Immunol 25:3504–3508

Kohler G, Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256:495–497

Krugmann S, Pleass RJ, Atkin JD, Woof JM (1997) Structural requirements for assembly of dimeric IgA probed by site-directed mutagenesis of J chain and a cysteine residue of the alpha-chain CH2 domain. J Immunol 159:244–249

Lee JS, Cella M, Mcdonald KG, Garlanda C, Kennedy GD, Nukaya M, Mantovani A, Kopan R, Bradfield CA, Newberry RD, Colonna M (2012) AHR drives the development of gut ILC22 cells and postnatal lymphoid tissues via pathways dependent on and independent of Notch. Nat Immunol 13:144–151

Lonberg N (2005) Human antibodies from transgenic animals. Nat Biotechnol 23:1117–1125

Lonberg N, Taylor LD, Harding FA, Trounstine M, Higgins KM, Schramm SR, Kuo CC, Mashayekh R, Wymore K, Mccabe JG et al (1994) Antigen-specific human antibodies from mice comprising four distinct genetic modifications. Nature 368:856–859

Lundqvist ML, Middleton DL, Hazard S, Warr GW (2001) The immunoglobulin heavy chain locus of the duck. Genomic organization and expression of D, J, and C region genes. J Biol Chem 276:46729–46736

Mabbott NA, Donaldson DS, Ohno H, Williams IR, Mahajan A (2013) Microfold (M) cells: important immunosurveillance posts in the intestinal epithelium. Mucosal Immunol 6:666–677

Macpherson AJ, Mccoy KD, Johansen FE, Brandtzaeg P (2008) The immune geography of IgA induction and function. Mucosal Immunol 1:11–22

Magadan-Mompo S, Sanchez-Espinel C, Gambon-Deza F (2013) IgH loci of American alligator and saltwater crocodile shed light on IgA evolution. Immunogenetics 65:531–541

Magor KE, Warr GW, Bando Y, Middleton DL, Higgins DA (1998) Secretory immune system of the duck (Anas platyrhynchos). Identification and expression of the genes encoding IgA and IgM heavy chains. Eur J Immunol 28:1063–1068

Mansikka A (1992) Chicken IgA H chains. Implications concerning the evolution of H chain genes. J Immunol 149:855–861

Martin A, Scharff MD (2002) Aid and mismatch repair in antibody diversification. Nat Rev Immunol 2:605–614

Mashoof S, Goodroe A, Du CC, Eubanks JO, Jacobs N, Steiner JM, Tizard I, Suchodolski JS, Criscitiello MF (2013) Ancient T-independence of mucosal IgX/A: gut microbiota unaffected by larval thymectomy in Xenopus laevis. Mucosal Immunol 6:358–368

Meng FL, Du Z, Federation A, Hu J, Wang Q, Kieffer-Kwon KR, Meyers RM, Amor C, Wasserman CR, Neuberg D, Casellas R, Nussenzweig MC, Bradner JE, Liu XS, Alt FW (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. Cell 159:1538–1548

Meyer S, Leusen JH, Boross P (2014) Regulation of complement and modulation of its activity in monoclonal antibody therapy of cancer. MAbs 6

Misaghi S, Senger K, Sai T, Qu Y, Sun Y, Hamidzadeh K, Nguyen A, Jin Z, Zhou M, Yan D, Lin WY, Lin Z, Lorenzo MN, Sebrell A, Ding J, Xu M, Caplazi P, Austin CD, Balazs M, Roose-Girma M, Deforge L, Warming S, Lee WP, Dixit VM, Zarrin AA (2013) Polyclonal hyper-IgE mouse model reveals mechanistic insights into antibody class switch recombination. Proc Natl Acad Sci USA 110:15770–15775

Moreau MC, Ducluzeau R, Guy-Grand D, Muller MC (1978) Increase in the population of duodenal immunoglobulin A plasmocytes in axenic mice associated with different living or dead bacterial strains of intestinal origin. Infect Immun 21:532–539

Morrison SL, Johnson MJ, Herzenberg LA, Oi VT (1984) Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. Proc Natl Acad Sci USA 81:6851–6855

Mussmann R, Courtet M, Schwager J, Du Pasquier L (1997) Microsites for immunoglobulin switch recombination breakpoints from *Xenopus* to mammals. Eur J Immunol 27:2610–2619

Mussmann R, Du Pasquier L, Hsu E (1996) Is *Xenopus* IgX an analog of IgA? Eur J Immunol 26:2823–2830

Navaratnam N, Fujino T, Bayliss J, Jarmuz A, How A, Richardson N, Somasekaram A, Bhattacharya S, Carter C, Scott J (1998) Escherichia coli cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition. J Mol Biol 275:695–714

Nimmerjahn F, Ravetch JV (2008) Fcgamma receptors as regulators of immune responses. Nat Rev Immunol 8:34–47

Ober RJ, Radu CG, Ghetie V, Ward ES (2001) Differences in promiscuity for antibody-FcRn interactions across species: implications for therapeutic antibodies. Int Immunol 13:1551–1559

Ohta Y, Flajnik M (2006) IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proc Natl Acad Sci USA 103:10723–10728

Pabst O, Herbrand H, Friedrichsen M, Velaga S, Dorsch M, Berhardt G, Worbs T, Macpherson AJ, Forster R (2006) Adaptation of solitary intestinal lymphoid tissue in response to microbiota and chemokine receptor CCR7 signaling. J Immunol 177:6824–6832

Petersen-Mahrt SK, Harris RS, Neuberger MS (2002) AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature 418:99–103

Presta LG (2002) Engineering antibodies for therapy. Curr Pharm Biotechnol 3:237–256

Pruzina S, Williams GT, Kaneva G, Davies SL, Martin-Lopez A, Bruggemann M, Vieira SM, Jeffs SA, Sattentau QJ, Neuberger MS (2011) Human monoclonal antibodies to HIV-1 gp140 from mice bearing YAC-based human immunoglobulin transloci. Protein Eng Des Sel 24: 791–799

Qian J, Wang Q, Dose M, Pruett N, Kieffer-Kwon KR, Resch W, Liang G, Tang Z, Mathe E, Benner C, Dubois W, Nelson S, Vian L, Oliveira TY, Jankovic M, Hakim O, Gazumyan A, Pavri R, Awasthi P, Song B, Liu G, Chen L, Zhu S, Feigenbaum L, Staudt L, Murre C, Ruan Y, Robbiani DF, Pan-Hammarstrom Q, Nussenzweig MC, Casellas R (2014) B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. Cell 159:1524–1537

Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC (2003) Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. Nat Immunol 4:452–456

Rast JP, Amemiya CT, Litman RT, Strong SJ, Litman GW (1998) Distinct patterns of IgH structure and organization in a divergent lineage of chrondrichthyan fishes. Immunogenetics 47:234–245

Reichert JM, Rosensweig CJ, Faden LB, Dewitz MC (2005) Monoclonal antibody successes in the clinic. Nat Biotechnol 23:1073–1078

Rescigno M, Di Sabatino A (2009) Dendritic cells in intestinal homeostasis and disease. J Clin Invest 119:2441–2450

Reynaud CA, Aoufouchi S, Faili A, Weill JC (2003) What role for AID: mutator, or assembler of the immunoglobulin mutasome? Nat Immunol 4:631–638

Roes J, Rajewsky K (1993) Immunoglobulin D (IgD)-deficient mice reveal an auxiliary receptor function for IgD in antigen-mediated recruitment of B cells. J Exp Med 177:45–55

Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, Aravind L, Pancer Z (2007) Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat Immunol 8:647–656

Roy D, Lieber MR (2009) G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. Mol Cell Biol 29:3124–3133

Roy D, Yu K, Lieber MR (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. Mol Cell Biol 28:50–60

Roy D, Zhang Z, Lu Z, Hsieh CL, Lieber MR (2010) Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Mol Cell Biol 30:146–159

Salazar-Gonzalez RM, Niess JH, Zammit DJ, Ravindran R, Srinivasan A, Maxwell JR, Stoklasek T, Yadav R, Williams IR, Gu X, Mccormick BA, Pazos MA, Vella AT, Lefrancois L, Reinecker HC, Mcsorley SJ (2006) CCR6-mediated dendritic cell activation of pathogen-specific T cells in Peyer's patches. Immunity 24:623–632

Salfeld JG (2007) Isotype selection in antibody engineering. Nat Biotechnol 25:1369–1372

Schaerlinger B, Bascove M, Frippiat JP (2008) A new isotype of immunoglobulin heavy chain in the urodele amphibian *Pleurodeles waltl* predominantly expressed in larvae. Mol Immunol 45: 776–786

Schaerlinger B, Frippiat JP (2008) IgX antibodies in the urodele amphibian *Ambystoma mexicanum*. Dev Comp Immunol 32:908–915

Schulz O, Jaensson E, Persson EK, Liu X, Worbs T, Agace WW, Pabst O (2009) Intestinal CD103+, but not CX3CR1+, antigen sampling cells migrate in lymph and serve classical dendritic cell functions. J Exp Med 206:3101–3114

Schwager J, Mikoryak CA, Steiner LA (1988) Amino acid sequence of heavy chain from *Xenopus laevis* IgM deduced from cDNA sequence: implications for evolution of immunoglobulin domains. Proc Natl Acad Sci USA 85:2245–2249

Shang L, Fukata M, Thirunarayanan N, Martin AP, Arnaboldi P, Maussang D, Berin C, Unkeless JC, Mayer L, Abreu MT, Lira SA (2008) Toll-like receptor signaling in small intestinal epithelium promotes B-cell recruitment and IgA production in lamina propria. Gastroenterology 135:529–538

Shinkura R, Tian M, Smith M, Chua K, Fujiwara Y, Alt FW (2003) The influence of transcriptional orientation on endogenous switch region function. Nat Immunol 4:435–441

Stager S, Alexander J, Kirby AC, Botto M, Rooijen NV, Smith DF, Brombacher F, Kaye PM (2003) Natural antibodies and complement are endogenous adjuvants for vaccine-induced CD8+ T-cell responses. Nat Med 9:1287–1292

Stavnezer J, Amemiya CT (2004) Evolution of isotype switching. Semin Immunol 16:257–275

Stavnezer J, Schrader CE (2014) IgH chain class switch recombination: mechanism and regulation. J Immunol 193:5370–5378

Streltsov VA, Varghese JN, Carmichael JA, Irving RA, Hudson PJ, Nuttall SD (2004) Structural evidence for evolution of shark Ig new antigen receptor variable domain antibodies from a cell-surface receptor. Proc Natl Acad Sci USA 101:12444–12449

Strohl WR (2009) Optimization of Fc-mediated effector functions of monoclonal antibodies. Curr Opin Biotechnol 20:685–691

Strugnell RA, Wijburg OL (2010) The role of secretory antibodies in infection immunity. Nat Rev Microbiol 8:656–667

Sun CM, Hall JA, Blank RB, Bouladoux N, Oukka M, Mora JR, Belkaid Y (2007) Small intestine lamina propria dendritic cells promote de novo generation of Foxp3 T reg cells via retinoic acid. J Exp Med 204:1775–1785

Suzuki K, Kawamoto S, Maruya M, Fagarasan S (2010) GALT: organization and dynamics leading to IgA synthesis. Adv Immunol 107:153–185

Ta VT, Nagaoka H, Catalan N, Durandy A, Fischer A, Imai K, Nonoyama S, Tashiro J, Ikegawa M, Ito S, Kinoshita K, Muramatsu M, Honjo T (2003) AID mutant analyses indicate requirement for class-switch-specific cofactors. Nat Immunol 4:843–848

Tarlinton D, Radbruch A, Hiepe F, Dorner T (2008) Plasma cell differentiation and survival. Curr Opin Immunol 20:162–169

Tasumi S, Velikovsky CA, Xu G, Gai SA, Wittrup KD, Flajnik MF, Mariuzza RA, Pancer Z (2009) High-affinity lamprey VLRA and VLRB monoclonal antibodies. Proc Natl Acad Sci USA 106:12891–12896

Tezuka H, Abe Y, Iwata M, Takeuchi H, Ishikawa H, Matsushita M, Shiohara T, Akira S, Ohteki T (2007) Regulation of IgA production by naturally occurring TNF/iNOS-producing dendritic cells. Nature 448:929–933

Tian M, Alt FW (2000) Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. J Biol Chem 275:24163–24172

Treon SP, Mitsiades C, Mitsiades N, Young G, Doss D, Schlossman R, Anderson KC (2001) Tumor cell expression of CD59 is associated with resistance to CD20 serotherapy in patients with B-cell malignancies. J Immunother 24:263–271

Tsuchiya N, Kyogoku C (2005) Role of Fc gamma receptor IIb polymorphism in the genetic background of systemic lupus erythematosus: insights from Asia. Autoimmunity 38:347–352

Tsuji M, Komatsu N, Kawamoto S, Suzuki K, Kanagawa O, Honjo T, Hori S, Fagarasan S (2009) Preferential generation of follicular B helper T cells from Foxp3+ T cells in gut Peyer's patches. Science 323:1488–1492

Tsuji M, Suzuki K, Kitamura H, Maruya M, Kinoshita K, Ivanov II, Itoh K, Littman DR, Fagarasan S (2008) Requirement for lymphoid tissue-inducer cells in isolated follicle formation and T cell-independent immunoglobulin A generation in the gut. Immunity 29:261–271

Turner RJ, Manning MJ (1974) Thymic dependence of amphibian antibody responses. Eur J Immunol 4:343–346

Uematsu S, Akira S (2009) Immune responses of TLR5(+) lamina propria dendritic cells in enterobacterial infection. J Gastroenterol 44:803–811

Uematsu S, Fujimoto K, Jang MH, Yang BG, Jung YJ, Nishiyama M, Sato S, Tsujimura T, Yamamoto M, Yokota Y, Kiyono H, Miyasaka M, Ishii KJ, Akira S (2008) Regulation of humoral and cellular gut immunity by lamina propria dendritic cells expressing Toll-like receptor 5. Nat Immunol 9:769–776

Varol C, Vallon-Eberhard A, Elinav E, Aychek T, Shapira Y, Luche H, Fehling HJ, Hardt WD, Shakhar G, Jung S (2009) Intestinal lamina propria dendritic cell subsets have different origin and functions. Immunity 31:502–512

Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay BH, Tohari S, Kong KW, Ho S, Lorente-Galdos B, Quilez J, Marques-Bonet T, Raney BJ, Ingham PW, Tay A, Hillier LW, Minx P, Boehm T, Wilson RK, Brenner S, Warren WC (2014) Elephant shark genome provides unique insights into gnathostome evolution. Nature 505:174–179

Wei Z, Wu Q, Ren L, Hu X, Guo Y, Warr GW, Hammarstrom L, Li N, Zhao Y (2009) Expression of IgM, IgD, and IgY in a reptile, *Anolis carolinensis*. J Immunol 183:3858–3864

Weir AN, Nesbitt A, Chapman AP, Popplewell AG, Antoniw P, Lawson AD (2002) Formatting antibody fragments to mediate specific therapeutic functions. Biochem Soc Trans 30:512–516

Weng WK, Levy R (2003) Two immunoglobulin G fragment C receptor polymorphisms independently predict response to rituximab in patients with follicular lymphoma. J Clin Oncol 21:3940–3947

Wesemann DR, Portuguese AJ, Meyers RM, Gallagher MP, Cluff-Jones K, Magee JM, Panchakshari RA, Rodig SJ, Kepler TB, Alt FW (2013) Microbial colonization influences early B-lineage development in the gut lamina propria. Nature 501:112–115

White MB, Word CJ, Humphries CG, Blattner FR, Tucker PW (1990) Immunoglobulin D switching can occur through homologous recombination in human B cells. Mol Cell Biol 10:3690–3699

Woof JM, Kerr MA (2004) IgA function–variations on a theme. Immunology 113:175–177

Xu J, Gordon JI (2003) Honor thy symbionts. Proc Natl Acad Sci USA 100:10452–10459

Xu W, He B, Chiu A, Chadburn A, Shan M, Buldys M, Ding A, Knowles DM, Santini PA, Cerutti A (2007) Epithelial cells trigger frontline immunoglobulin class switching through a pathway regulated by the inhibitor SLPI. Nat Immunol 8:294–303

Xu W, Santini PA, Matthews AJ, Chiu A, Plebani A, He B, Chen K, Cerutti A (2008) Viral double-stranded RNA triggers Ig class switching by activating upper respiratory mucosa B cells through an innate TLR3 pathway involving BAFF. J Immunol 181:276–287

Xu Z, Wang GL, Nie P (2009) IgM, IgD and IgY and their expression pattern in the Chinese soft-shelled turtle *Pelodiscus sinensis*. Mol Immunol 46:2124–2132

Xue K, Rada C, Neuberger MS (2006) The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. J Exp Med 203:2085–2094

Yoshida T, Mei H, Dorner T, Hiepe F, Radbruch A, Fillatreau S, Hoyer BF (2010) Memory B and memory plasma cells. Immunol Rev 237:117–139

Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol 4:442–451

Zarrin AA, Alt FW, Chaudhuri J, Stokes N, Kaushal D, Du Pasquier L, Tian M (2004) An evolutionarily conserved target motif for immunoglobulin class-switch recombination. Nat Immunol 5:1275–1281

Zarrin AA, Del Vecchio C, Tseng E, Gleason M, Zarin P, Tian M, Alt FW (2007) Antibody class switching mediated by yeast endonuclease-generated DNA breaks. Science 315:377–381

Zarrin AA, Tian M, Wang J, Borjeson T, Alt FW (2005) Influence of switch region length on immunoglobulin class switch recombination. Proc Natl Acad Sci USA 102:2466–2470

Zhang C, Du Pasquier L, Hsu E (2013) Shark IgW C region diversification through RNA processing and isotype switching. J Immunol 191:3410–3418

Zhang ZZ, Pannunzio NR, Han L, Hsieh CL, Yu K, Lieber MR (2014) The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. Cell Rep 8:557–569

Zhao Y, Pan-Hammarstrom Q, Yu S, Wertz N, Zhang X, Li N, Butler JE, Hammarstrom L (2006) Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*. Proc Natl Acad Sci USA 103:12087–12092

Zhu C, Lee V, Finn A, Senger K, Zarrin AA, Du Pasquier L, Hsu E (2012) Origin of immunoglobulin isotype switching. Curr Biol 22:872–880

Index

E. Hsu, L. Du Pasquier (eds.), *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, Results and Problems in Cell Differentiation 57,
DOI 10.1007/978-3-319-20819-0

R