Immunogenetics (2002) 54:87–95
DOI 10.1007/s00251-002-0436-x

ORIGINAL PAPER

**Sergei V. Guselnikov · Svetlana A. Ershova
Ludmila V. Mechetina · Alexander M. Najakshin
Olga Y. Volkova · Boris Y. Alabyev
Alexander V. Taranin**

A family of highly diverse human and mouse genes structurally links leukocyte FcR, gp42 and PECAM-1

Received: 29 August 2001 / Revised: 21 January 2002 / Published online: 14 March 2002

Abstract A group of genes encoding proteins structurally related to the leukocyte Fc receptors (FcRs) and termed the *IFGP* family was identified in human and mouse. Sequences of four human and two mouse cDNAs predict proteins differing by domain composition. One of the mouse cDNAs encodes a secreted protein, which, in addition to four immunoglobulin (Ig)-like domains, contains a scavenger receptor superfamily-related domain at the C-terminus. The other cDNAs code for the type I transmembrane proteins with the extracellular parts comprised of one to six Ig-like domains. Five homologous types of the Ig-like domains were defined and each protein was found to have a unique combination of the domain types. The cytoplasmic tails of the transmembrane proteins show different patterns of the tyrosine-based signal motifs. While the human *IFGP* members appear to be B-cell antigens, the mouse genes have a broader tissue distribution with predominant expression in brain. Sequence comparisons revealed that the *IFGP* family may be regarded as a phylogenetic link joining the leukocyte FcRs with the rat NK cell-specific gp42 antigen and platelet endothelial cell adhesion molecule-1 (PECAM-1), two mammalian leukocyte receptors whose close relatives were not found previously. It is suggested that FcRs, the IFGP proteins and gp42 have arisen by a series of duplications from a common ancestor receptor comprised of five Ig-like domains. The organization of the human genes shows that the *IFGP* family evolved through differential gain and loss of exons due to recombination and/or mutation accumulation in the duplicated copies.

Keywords Fc receptors · Evolution · B lymphocytes

S.V. Guselnikov · S.A. Ershova · L.V. Mechetina · A.M. Najakshin
O.Y. Volkova · B.Y. Alabyev · A.V. Taranin (✉)
Laboratory of Immunogenetics,
Institute of Cytology and Genetics,
Lavrentjev st. 10, Novosibirsk 630090, Russia
e-mail: taranin@bionet.nsc.ru
Tel.: +7-3832-333806, Fax: +7-3832-331278

Introduction

Comparative analysis of the human, fly, and worm genomes demonstrates that gene duplication has been the major driving force in the evolving immune system (Lander et al. 2001; Venter et al. 2001). Expansion of the immunoglobulin superfamily (IgSF) is particularly relevant in this regard. The superfamily contains about tenfold more members in mammals than invertebrates (Lander et al. 2001; Venter et al. 2001). Over the past two decades, a large number of the IgSF proteins involved in immune response has been characterized (Barclay et al. 1997). Most may be assigned to distinct families, whereas the precise relationships among many are unclear. As a rule, members of the same family serve similar functions. In some cases, however, structurally related proteins may be functionally non-overlapping and vice versa. How duplicated copies might have acquired novel functions remains elusive. Reconstruction of the molecular events, which underlie the emergence of the sophisticated network of functions fulfilled by the IgSF members, is still far from completion.

The family of the leukocyte Fc receptors is a subset of IgSF and the largest group among the structurally diverse cell-surface molecules able to bind the Ig constant regions (Ravetch and Kinet 1991; Daeron 1997; Shibuya et al. 2000; Ravetch and Bolland 2001; van Egmond et al. 2001). Four major classes of FcRs have been described and investigated: FcγRI (CD64), FcγRII (CD32), FcγRIII (CD16) and FcεRI (Ravetch and Kinet 1991). In mouse, each class is encoded by a single gene; in human, there are two to three separately encoded forms for each class except FcεRI. The classes and forms differ by their ligand specificity or signal properties. Broad functional versatility of the family is due to the expression of alternative isoforms and interaction between classes with different signaling capabilities (Ravetch and Bolland 2001). FcR homologs in non-mammalian vertebrates have not been found so far and the evolutionary history of the family has not been elucidated. From sequence comparisons and genomic analysis it is clear that the mammalian

FcR genes are derived from a common ancestor by a series of duplications and recombinations, some of which were species-specific (Ravetch and Kinet 1991). The high-affinity FcγRI differs from the other receptors in the presence of the third unique extracellular domain. Using this feature for the expression sequence tag (EST) database search, we identified a group of human and mouse genes encoding unknown FcR-related proteins. One of the proteins, termed FCRL, has been characterized in detail (Mechetina et al. 2002). In this report, we describe a family of closely related human and mouse *FcR*-like genes and discuss the evolutionary implications of this finding.

Materials and methods

cDNA

The human cDNA clones *aa63g02* (I.M.A.G.E. clone ID: 825650) and *tz31b04* (ID: 2290159), as well as the mouse cDNA clones *vf84e06* (ID: 850498), *uc35g02* (ID: 1400018), were obtained from the I.M.A.G.E. Consortium (Lennon et al. 1996) through ATCC (USA) or Research Genetics (USA). The cDNA NCI_CGAP_GCB1 library constructed from the human tonsil germinal center B cells was a kind gift of Dr. M.B. Soares, University of Iowa. The tonsil S2T cDNA library (Wong et al. 1985) was purchased from ATCC. The RACE (rapid amplificaton of cDNA ends)-ready tonsil cDNA was prepared using the SMART RACE kit (Clontech, USA) and following the manufacturer's instructions.

PCR amplification and cloning

The following pairs of primers were used to amplify human cDNAs: forward primer corresponding to the hIFGP1 leader peptide (5′-TAGGTACCATCCCTGACCTG-3′) and reverse primer (the hIFGP1 3′-untranslated region, UTR) (5′-AGGAGCCGGCAGGAATCTGGTTC-3′); forward primer (the hIFGP2 leader peptide) (5′-TTCCATGCTGCTGTGGGCGTCCTTGCTG-3′)and reverse primer (the hIFGP2 cytoplasmic tail including stop codon) (5′-TTAACTTTCTTCATCCTTAGAGCTGATC-3′; forward primer (the hIFGP3 leader peptide) (5′-GGCCGGTGCCCATGCTTCTGTGG-3′)and reverse primer (the hIFGP3 3′-UTR) (5′-AATGGTGCAGGCTGTTTCCTGTG-3′); λgt11 forward primer (Promega, USA) (5′-GGTGGCGACGACTCCTGGAGCCCG-3′) and reverse primer (the hIFGP4 3′-UTR) (5′-CATCCTTGCTGTTGATCTTCCCT-3′). The polymerase chain reaction (PCR) conditions were as follows: 45 cycles for hIFGP1 (30 s at 94°C, 30 s at 62°C, 75 s at 72°C), hIFGP3 (30 s at 94°C, 30 s at 69°C, 105 s at 72°C), and hIFGP4 (30 s at 94°C, 30 s at 68°C, 60 s at 72°C). Amplification of hIFGP2 was performed using Touchdown PCR: 30 cycles at 94°C (30 s), 65°C–50°C (–0.5°C/cycle) (30 s), and 72°C (90 s), followed by 15 cycles at 94°C (30 s), 55°C (30 s), and 72°C (90 s). The PCR fragments were cloned into the pBluescript KS vector (Stratagene, USA) and sequenced using an automated fluorescent sequencer. The mouse *vf84e06* and *c35g02* cDNAs were sequenced manually using the Fmol DNA Sequencing System (Promega, USA) and the Sequenase Version 2.0 DNA Sequencing kit (USB, USA).

Southern blot analysis

Genomic DNA from human and mouse blood cells was isolated (Sambrook et al. 1989) and digested to completion with restriction endonucleases *Bam*HI, *Hin*dIII or *Pvu*II. The digested DNA (10 µg/lane) was separated on 1% agarose gels and transferred onto Zeta-probe nylon membranes (BioRad Laboratories, USA) by the vacuum blotting technique in 0.4 M NaOH. Hybridizations with ^{32}P-labeled probes were performed following the membrane manufacturer's recommendations. The probes were the fragment of the *uc35g02* cDNA coding for a part of the second domain of mIFGP1 (151 bp), the *Xho*I/*Bam*HI fragment of the *vf84e06* insert encoding two N-terminal domains of mIFGP2 (511 bp), and PCR amplified cDNA fragments for a part of the third (177 bp) domain of hIFGP1. These probes were used for both Southern and Northern blot analyses.

Northern blot analysis

Poly(A)$^+$ RNA from tissues of BALB/c mice was isolated from total RNA on oligo-dT cellulose (Sigma, USA), fractionated (5 µg/line) on 1% agarose gels and vacuum blotted onto a Zeta-Probe nylon membrane (Bio-Rad Laboratories, USA). The blot was hybridized with either ^{32}P-labeled *mIFGP1*, *mIFGP2* or *β-actin* probes at high stringency conditions following the Bio-Rad recommendations.

Bioinformatics tools

Nucleotide and amino acid sequences were analyzed using utilities at the BCM Search Launcher (Smith et al. 1996) web site (http://www.hgsc.bcm.tmc.edu/) and at the NCBI web site (http://www3.ncbi.nlm.nih.gov/). Computer alignments were generated by CLUSTAL W 1.7 (Thompson et al. 1994). For signal peptide prediction, the SignalP program was used at http://www.cbs.dtu.dk/services/SignalP/ (Nielsen et al. 1997). Homology searches were performed using TBLASTN, BLASTP, BLASTx (Altschul et al. 1997) and FASTA2 (Pearson 1990) programs. The sequences were retrieved from GenBank using ENTREZ at http://www3.ncbi.nlm.nih.gov/. The human genomic sequences were analyzed at and retrieved from the web site of the Sanger Centre (http://www.sanger.ac.uk/HGP). The GenScan program (Burge and Karlin 1997) at http://genes.mit.edu/GENSCAN.html was used for gene predictions. The predictions were further manually inspected and corrected.

Results

Identification of the human and mouse cDNAs encoding the gp42/FcR-related proteins

The BLAST2 analysis of the protein databases showed that the amino acid sequence of the third domain of mammalian FcγRI shares 35–40% identical residues with the first domain of gp42, a specific marker of rat NK-cells with unknown function (Seaman et al. 1991). Gp42 is a GPI-anchored membrane protein consisting of two Ig-like domains. The second gp42 domain proved to be unique in that it did not belong to any one of the known IgSF families. No gp42 homologs have been found in other mammals either. However, screening of the EST database identified three sets of mouse and human cDNAs, which coded for proteins showing moderate structural similarity to rat gp42. Several of the deduced amino acid sequences contained also FcR-related domains. Based on the EST analysis, it was suggested that mammals possess a so far unknown protein family we designated IFGP (from IgSF, FcR, gp42). To characterize the family in more detail, two mouse cDNAs and one human cDNA were obtained from the I.M.A.G.E

Table 1 Degree of identity between the Ig-like domains of the IFGP proteins, rat gp42 and FcγRI calculated as percent of identical amino acid residues in the aligned sequences

	hIFGP1 -[a] |- |D3|D4|D5	hIFGP2 D1|D2|D3|D4|-	hIFGP3 D1|D2|D3|D4|D5	mIFGP1 - |- |- |D4|D5	mIFGP2 - |D2|D3|D4|D5	rat gp42 - | - |D3| - |D5
hIFGP2	- | - |38|54|-					
hIFGP3	- | - |56|52|75	38|31|52|58|-				
mIFGP1	- | - | - |60|68	- | - | - |40| -	- | - | - |44|68			
mIFGP2	- | - |44|45|59	- |24|42|63| -	- |36|47|46|56	- | - | - |34|53		
Rat gp42	- | - |20|- |40	- | - |30|- | -	- | - |32| - |43	- | - | - | - |43	- | - |34| - |34	
hFcγRI D1|D2|D3	- | - |35| - | -	37|30|46|- | -	29|31|45| - | -	- | - | - | - | -	- |19|46| - | -	- | - |36| - | -

[a] Hyphen indicates absence of the corresponding domain type

consortium (Lennon et al. 1996) and their complete nucleotide sequences were determined.

Both mouse cDNAs were full-size inserts containing an open reading frame (ORF), as well as 5′ and 3′ untranslated sequences. The *uc35g02* cDNA encoded a type I transmembrane protein of 343 residues spanning the predicted leader peptide, two Ig-like extracellular domains, a transmembrane domain, and a cytoplasmic tail (Fig. 1). The first domain showed only relatively distant similarity to the domains of the known IgSF members, while the second domain was similar to that of rat gp42. This protein will be subsequently referred to as mIFGP1. The structure of the protein encoded by the *vf84e06* cDNA was more unusual. Its predicted amino acid sequence of 507 residues was comprised of the leader peptide and five domains. The protein appeared to be secreted, as judged by the absence of a transmembrane region. Four domains were Ig-like, whereas the fifth belonged to the scavenger receptor cysteine-rich (SRCR) superfamily. The first domain exhibited weak, but statistically significant, similarity to the second domain of the FcεRI alpha chain. The second domain was homologous to FcγRId3, the third and the fourth were most similar to the mIFGP1 domains. The SRCR-like domain showed the highest degree of similarity to the domains of the secreted protein Spα/AIM, identified recently (Gebe et al. 1997; Miyazaki et al. 1999). That the *vf84e06* cDNA was not chimeric was confirmed by the finding of several similar cDNAs in the EST database. The predicted protein will be termed mIFGP2.

The human cDNA *aa63g02* had an ORF truncated at the 3′ end. Its sequence was further used for analysis of the human genomic sequences produced by the Sanger Centre. It was revealed that the human genome contains at least six IFGP-like genes organized in a tight cluster on human Chromosome 1q21. The specific primers corresponding to the predicted 5′ and 3′ untranslated sequences of these genes were constructed and used for PCR of two tonsil-derived cDNA libraries. In addition, forward and reverse primers were used for 5′- and 3′-RACE of tonsil mRNA. As a result, the cDNA products of the four genes were isolated and characterized. The encoded proteins (designated hIFGP1, hIFGP2, hIFGP3 and hIFGP4) were highly related to each other

mIFGP1
MLPWLLLLLICALPCEPAGISDVSLKTRPPGGWVMEGDKLVLICSVDRVTG
NITYFWYRGALGFQLETKTQPSLTAEFEISDMKQSDADQYYCAANDGHDP
IASELVSIHVRVPVSRPVLTFGDSGTQAVLGDLVELHCKALRGSPPIFYQ
FYHESIILGNSSAPSGGGASFNFSLTAEHSGNFSCEASNGQGAQRSEVVA
LNLTGLSLVPTENGISHLSLGLTGWLLGCLSPITMALIFCYWLKRKIGRQ
SEDPVRSPPQTVLQGSTYPKSPDSRQPEPLYENVNVVSGNEVYSLVYHTP
QVLEPAAAQHVRTHGVSESFQVSSGLYSKPRINIAHMDYEDAM

mIFGP2
MPLCLLLLVFAPVGVQSDWLSISLPHRSYEGDQVVISCTGKNNGDIKRLK
YFKDGYHIETYSSASSYTIRNARRGDSGSYSCKADRKFFLFIDTTEETGS
KWLNVQELFPAPGLTASPLQPVEGSSVTLSCNTWLPSDRATTQLRYSFFK
DGHTLQSGWTSSKFTISAISKEDSGNYWCEAMTASRSVSKQSHRSYIDVE
RIPVSQVTMEIQPSRGWGVEGEPLVVEGEPLVLACSVAKGTGLITFSWHR
QDTKESVGKKSQRSQRVELEIPTIRESHAGGYYCTADNNYGLIQSAIVNI
TVKIPVLNPLLSISVPGVLPFIGDGGVHCEDKRASPPVLYWFYHENITLA
NTSAPFGGKASFKLSLTAGHSGNYSCEAENAWGTKRSEEVTLNVTEPPPK
VRLVNGPHHCEGRVEVEQEGRWGTVCDDGWDMRDVAVVCRELGCGAANTP
IAMLYPPAVDEALPVLIQVALCNGTEKTLAECDQVEAFDCGHDEDAGAVC
EVLPSTF

Fig. 1 Deduced amino acid sequences of the mouse IFGP proteins. The leader peptides and the transmembrane regions are *underlined*. The SRCR-like domain in the mIFGP2 sequence is indicated by a *dotted line*. The nucleotide sequences of the mouse and human IFGPs are deposited in EMBL/GenBank under the accession numbers: AF329485, AF329486, AF329488, AF329490, AF390037

and to the mouse proteins, yet each contained a unique combination of the Ig-like domains. During the preparation of this manuscript three other groups reported the human sequences under the names SPAP1 (Xu et al. 2001), IRTA (Hatzivassiliou et al. 2001) and FcRH (Davis at al. 2001). IFGP1 is identical to FcRH1, IFGP2 is identical to IRTA1, IFGP3 is identical to FcRH3, and IFGP4 corresponds to SPAP1, presumably an alternative isoform of FcRH2.

Comparative analysis of the extracellular and intracellular domains

The results of comparisons of deduced amino acid sequences of the human and mouse IFGP proteins with each other and with their known homologs are summarized in Fig. 2 and Table 1. The Ig-like domains of five

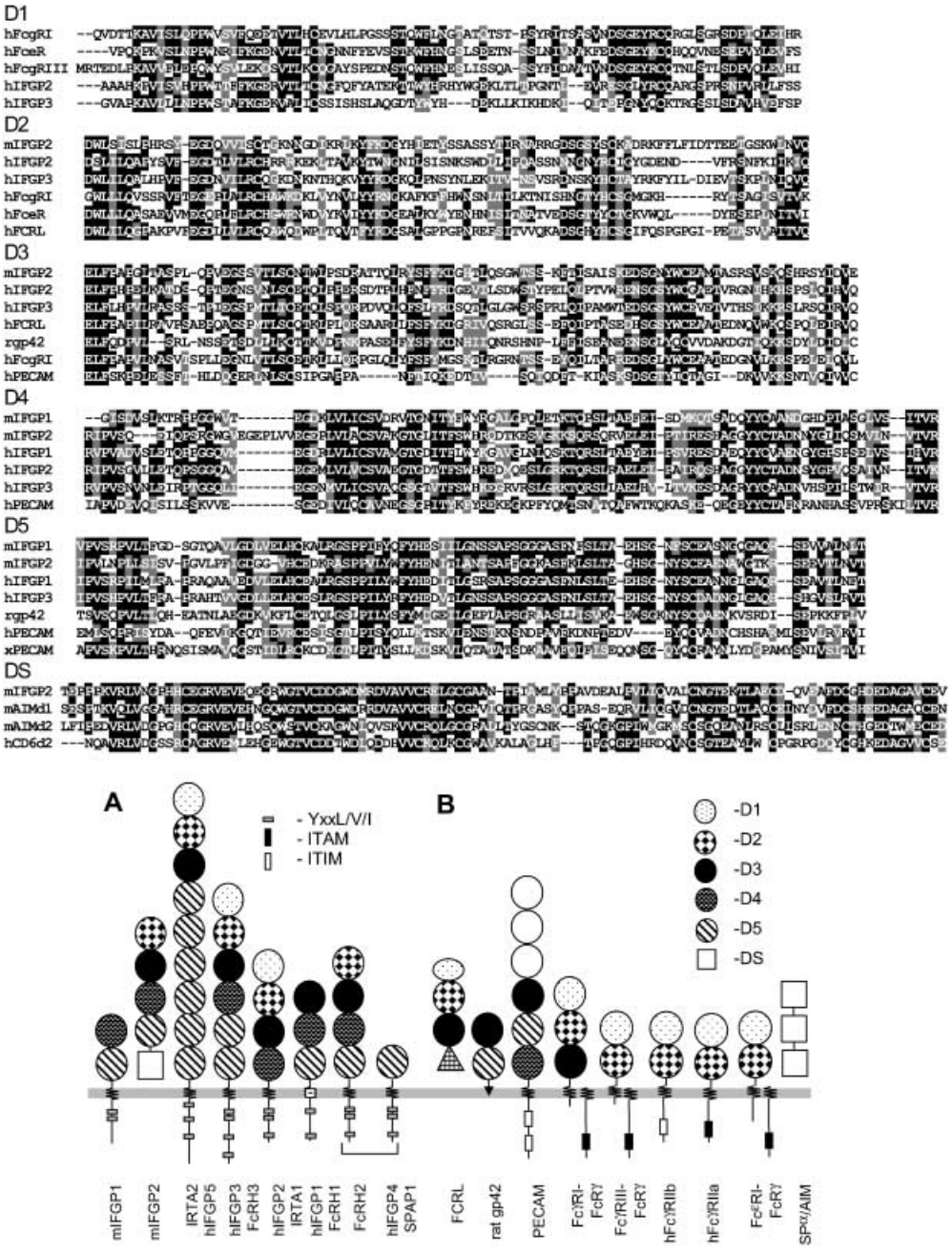

Fig. 2 Sequence comparison of the IFGP, FcR, gp42 and PECAM domains and a schematic representation of their domain composition. Identical residues are shown by *white letters* on *black background* and similar residues are *white* on *gray background* (generated by the BOXSHADE 3.21 program). The threshold for residue shading is their presence in at least a half of the sequences. *Dashes* indicate gaps introduced to maximize homology. *D1* through *D5* designate the Ig-like domain homology types. *DS* is the SRCR-like domain. The sequences of human (h) FcγRI, FcγRIII, FcεRI, FCRL, CD6, rat gp42, mouse AIM/Spα, and human (h) PECAM were taken from the EMBL/GenBank databases (X14356, X16863, X06948, AF329489, X60992, X56448, AF018268, M28526, respectively). The complete sequence of *Xenopus laevis* (x) PECAM is unpublished. In the scheme, *panel A* shows the IFGP family and *panel B*, their known homologs. The Ig-like domains and SRCR-like domains are depicted by *circles* and *squares*, respectively. Domain homology is shown by a *pattern of filling*. FcRH2 and hIFGP4/SPAP1 are the products of the same gene. The PECAM-1 N-terminal domains (*open circles*) are not related to the FcR and IFGP domains

homologous types (designated D1 through D5) were distinguished. The criteria for reference of a domain to a particular type were homology percent (expressed as the number of identical and similar amino acids), the presence of characteristic motifs, and the alignment scores in BLAST2 and FASTA2. With this classification, the domain structures of FcRs and the IFGP molecules were as follows: D1D2 for FcγRII, FcγRIII, and FcεRI, D1D2D3 for FcγRI, D3D5 for gp42, D4D5 for mIFGP1, D2D3D4D5DS for mIFGP2 (where DS is the SRCR-like domain), D3D4D5 for hIFGP1, D1D2D3D4 for hIFGP2, $D1D2D3D4(D5)_2$ for hIFGP3, and D5 for hIFGP4. IRTA2 reported by Hatzivassiliou and co-workers (2001) has a structure $D1D2D3(D5)_6$.

The currently described proteins and gp42 are obviously closer to each other than to the leukocyte FcRs, although gp42 appears to be the most distant member of the IFGP family. hIFGP1, hIFGP3 and hIFGP4 share 75–82% identical residues (85–90% identity at the nucleotide level) in the domains of the D5-type, suggesting very recent duplication of their genes. However, similarity in the other domains is weaker (51–57% identical residues in the D4- and D3-type domains of hIFGP1 and hIFGP3). hIFGP2 shows only moderate similarity to the other three human proteins (30–51% identical residues in the domains of the same type). Relationships among the human and mouse proteins remain unclear. The mIFGP1 domains of D4- and D5-types share 60% and 68% iden-

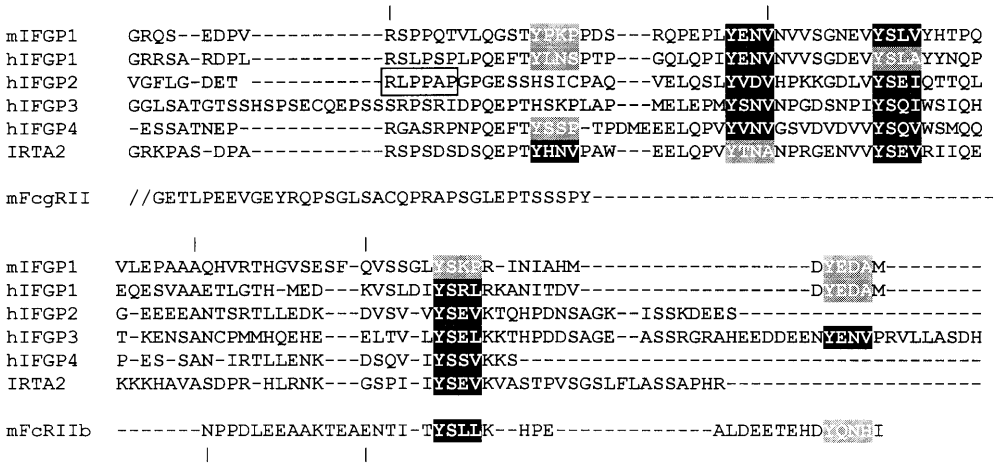

Fig. 3 Alignment of the IFGP cytoplasmic regions. The cytoplasmic tails of mouse FcγRII and human FcγRIIb (truncated at the N-terminus) are given for comparison. *Dashes* indicate gaps introduced to maximize homology. The tyrosine-based motifs are shown by *white letters* on *black* (conforming to the consensus YxxV/I/L) or *gray* (not conforming to it) background. The proline-rich motif in hIFGP2 is *boxed*. *Vertical lines* designate exon/intron boundaries in the human IFGP (*above*) and in the FcγRII (*below*) sequences

tical residues with the corresponding hIFGP1 domains. These two proteins are quite probably orthologs. However, mIFGP1 has one domain less; its transmembrane region, in contrast to hIFGP1, does not contain a negatively charged residue. hIFGP2 is more similar to mIFGP2 than to hIFGP1 and hIFGP3 in the D4-type domain (63, 56 and 58% identity, respectively). In contrast, the D2-type domains of hIFGP2 and mIFGP2 share only 24% identical residues. The presence of the additional SRCR-like domain in mIFGP2 is more evidence that mIFGP2 and hIFGP2 are paralogs. Finally, it was found that FCRL, presumably the secreted protein described previously (Mechetina et al. 2002), is more closely related to FcRs than to the IFGP proteins in the D2 and D3 domains.

No close relatives of D4 were found in other members of IgSF. In the BLAST2 search for such homologs using the D4 consensus sequence, the scores were highest for the sixth domain of mammalian PECAM (E values up to 5e-07). Hatzivassiliou and co-workers (2001) have also indicated that IRTA proteins resemble PECAM, although the similarity was not specified for the domain composition. PECAM may be regarded as an orphan IgSF member for the reason that it has no close relatives (Newman 1997). Upon more careful examination of the primary structure of the PECAM domains, it proved that the fourth PECAM domain shows similarity to the D3 type (Fig. 2). Among the IgSF domains only D3 has the N-terminal motif ELFxxPxL. Furthermore, we assume also that the fifth PECAM domain may be assigned to the D5 type. Analysis of the *Xenopus* and zebrafish PECAM homologs (unpublished data) made this similarity more conspicuous (Fig. 2). Thus, there are good reasons for the suggestion that PECAM might have shared a common evolutionary ancestor with the leukocyte FcRs and IFGPs. However, no definite homology with FcRs or the IFGP proteins was found for the three N-terminal domains of PECAM.

The cytoplasmic tails of all the predicted IFGP transmembrane proteins contain tyrosine-based src-homology type 2 (SH2)-domain binding motifs (Unkeless and Jin 1997; Taylor et al. 2000). Alignment of the cytoplasmic tails clearly indicates five conserved tyrosine positions (Fig. 3). However, the tyrosine-based motifs are not conserved and each member of the family has its own pattern of the motif variants. While some tyrosines are embedded in typical YxxL/I/V sequences, others have serine, proline, or alanine at position +3. Judging by the residues at position –2 relative to tyrosines (hydrophobic or charged) and the motif arrangement, it may be suggested that the mIFGP1 and hIFGP1 motifs are ITAM-like, those of hIFGP4 and hIFGP5/IRTA2 are ITIM-like, whereas hIFGP3 possesses both the ITIM and ITAM-like sequences. The N-terminal motif of the hIFGP2 cytoplasmic tail has a serine residue at position –2. This feature is thought to be a signature of the immunoreceptor tyrosine-based switch motif (ITSM) that is a site for the adaptor protein SH2D1 A regulating SHP-2/SHIP binding (Sayos et al. 1998; Shlapatska et al. 2001). Finally, the proline-rich sequence conforming to the consensus RxPPxP (the putative SH3-domain binding site) in the hIFGP2 cytoplasmic region is noteworthy. The motif is similar to the one found in the cytoplasmic tails of vertebrate CD3ε (Alabyev et al. 2000).

Genomic structure

Recent progress in the genome sequencing project enabled us to determine the genomic structure of the human *IFGP* family (Fig. 4). The identified genes are clustered in a region of approximately 300 kb. The leader peptides are, as a rule, encoded by two exons, while the Ig-like domains and transmembrane regions are encoded by a single exon each. The cytoplasmic tails are invariantly encoded by five exons, thereby confirming that different patterns of the tyrosine-based motifs in the IFGP proteins resulted from mutation accumulation rather than from differential exon loss. Of particular interest is the organization of the *hIFGP1*, *hIFGP4* and *IFGPx* genes. The *hIFGP1* gene, in addition to exons for the domains of the D3, D4 and D5 types, contains two aberrant exons encoding polypeptides matching the D1- and D2-type domains. Both exons have frame-shift mutations. The

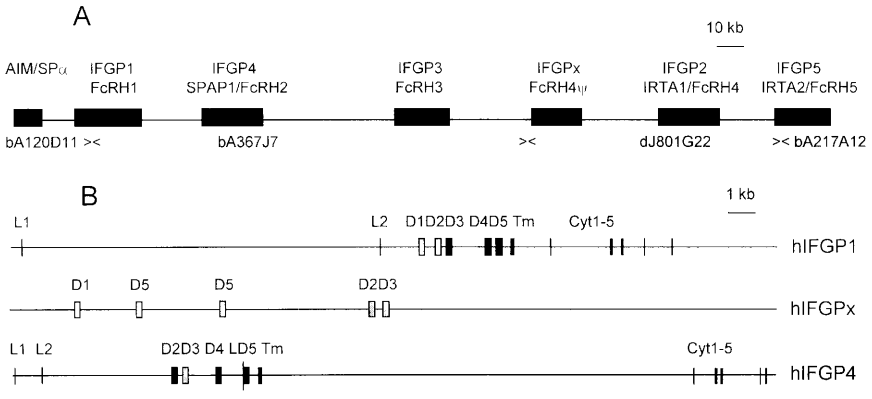

Fig. 4A, B The genomic structure of the human *IFGP* locus. **A** Arrangement of the *AIM/Spα* and *IFGP* genes, as determined using sequences produced by the Sanger Centre. *Arrows* designate the clone borders. **B** Exon/intron organization of the *hIFGP1*, *hIFGPx* and *hIFGP4* genes. Exons are shown by *lines* or *filled rectangles*. Aberrant exons are shown by *open rectangles*. Exons whose functional significance is unclear are shown by *shaded rectangles*. *L* Exons for the leader peptide; *D* the extracellular domains (the numeration corresponds to the domain homology type); *Tm* the transmembrane; *Cyt* cytoplasmic parts of the proteins

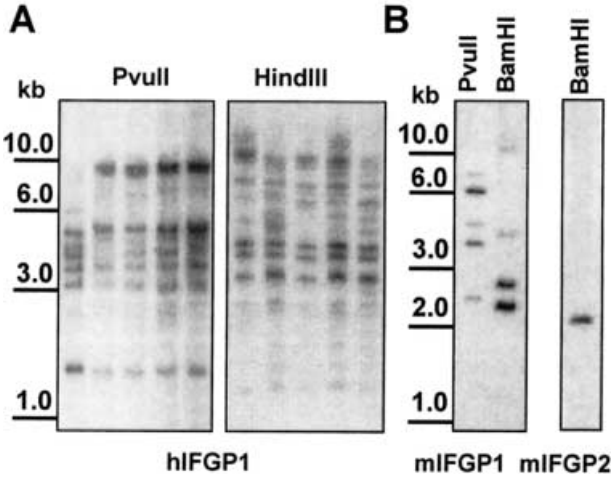

Fig. 5A, B Southern blot analysis of *hIFGP1*, *mIFGP1*, and *mIFGP2*. The DNA was isolated from human (**A**) or mouse (**B**) blood cells and digested with the indicated restriction endonucleases. The probes were the PCR-amplified cDNA fragments for the D5-type domain of hIFGP1, the D5-type domain of mIFGP1, and *Xho*I/*Bam*HI fragment of the *vf84e06* insert coding for two N-terminal domains (D2D3) of *mIFGP2*

IFGPx gene is located between *hIFGP2* and *hIFGP3* and consists of 5 exons (D1D5D5D2D3), of which the first three contain frame-shift mutations and/or stop codons. No exons for the transmembrane or cytoplasmic regions were found using different gene-prediction programs. However, the exons for D2- and D3-type domains have correct splice sites, and it cannot be excluded that the gene is functional and may code for a secreted protein. The *hIFGP4* gene appears to contain exons for the four extracellular domains (D2D3D4D5). Analysis of the *dj801G22* genomic sequence (accession no. AL135929) showed that the exon for the D3-type domain of hIFGP4 has an impaired 5′-splice site (phase 2). In the current experiments, 5′-RACE of individual tonsil cDNA produced only short fragments encoding a protein with a single extracellular domain of the D5-type. Interestingly, both the leader peptide and Ig-like domain are encoded by a single exon in this variant. Xu and co-workers have reported similar results using the same approach (2001). In contrast, Davis and co-workers (2001) have described a cDNA for the four-domain protein (D2D3D4D5) with a leader peptide encoded by two distinct exons. It remains to be determined whether these discrepancies are due to population polymorphism or to other causes.

The genomic DNA from samples of human and mouse blood was analyzed by Southern blotting (Fig. 5). The probe corresponding to the third hIFGP1 domain (D5) revealed multiple hybridization bands. This is consistent with the genomic structure of the human *IFGP* locus and the high (85–90%) nucleotide similarity among the exons for the D5-type domains. Hybridization of the mouse *Bam*HI DNA digest using the *mIFGP2* cDNA fragment as a probe revealed one band. In contrast, the *mIFGP1*-specific probe hybridized with multiple bands, thereby indicating that mouse, like human, possesses a family of *IFGP1*-related genes.

The mouse and human genes differ by their expression

The expression patterns of the identified genes in mouse and human tissues were examined by Northern blot analysis (Fig. 6). The *mIFGP1*-specific probe revealed transcripts of three sizes. Long transcripts (about 5 kb) were detected in brain, liver, intestine, colon, skeletal muscles, peripheral blood leukocytes and spleen. Short transcripts of different sizes were found also in testis, leukocytes and spleen. The mIFGP2 mRNA was expressed in brain, skeletal muscles, colon, testis and thymus. In both cases, brain mRNA produced the most intense signals. This is in contrast to the human genes which were shown to express in the secondary lymphoid but not in nervous tissues (Davis at al. 2001; Hatzivassiliou et al. 2001; Xu et al. 2001; our unpublished data).

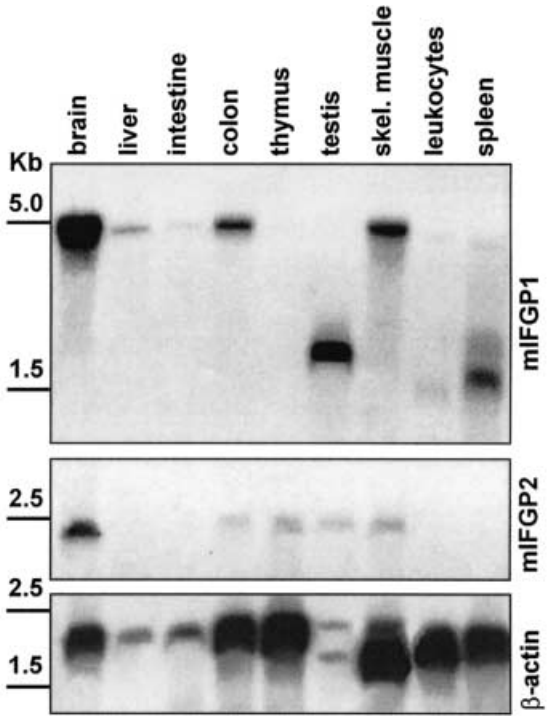

Fig. 6 Northern blot analysis of mIFGP1 and mIFGP2. Mouse tissue mRNA (5 μg/lane) was probed with the *Not*I/*Eco*RI *uc35g02* insert encoding mIFGP1, and the *Xho*I/*Bam*HI fragment of the *vf84e06* insert coding for two N-terminal domains of mIFGP2

Discussion

Sequence comparisons of the IFGP proteins show their derivation by a series of duplications from a common five-domain progenitor. The striking feature of the family is the diversity of its members. Each of the seven (including gp42) proteins has a unique domain composition. Five types of the Ig-like domains (designated D1 through D5) were defined but none is shared by all the proteins. The genomic structure of the human genes indicates that heterogeneity has resulted mainly from post-duplication divergence. Two major mechanisms have obviously contributed to the diversification of duplicated copies: intra- and/or intergenic recombinations and loss of exons due to accumulated mutations. Recombinations presumably generated the *hIFGP2* (D5 loss) and *hIFGP3* (gain of an additional D5) genes. Loss of exons as a result of mutations is apparent in the case of the *hIFGP1* gene and it is perhaps a polymorphic feature of the *hIFGP4* gene. The *IFGPx* gene (D1ψD5ψD5ψD2D3), whose ability to produce proper transcripts is questionable, might have emerged by both mechanisms. With regard to the intergenic shuffling, mIFGP2 bearing a SRCR-like domain is of particular interest. To our knowledge, this is the first instance of a mosaic protein combining together the immunoglobulin and the scavenger receptor superfamilies. The mIFGP2 SRCR-like domain exhibits the highest degree of similarity to the domains of the recently identified mouse and human secreted protein Spα/AIM (Gebe et al. 1997; Miyazaki et al. 1999). This protein functions as an inhibitor of apoptosis in certain cells and also as an inhibitor of B-lymphocyte proliferation (Miyazaki et al. 1999; Yusa et al. 1999). It is of interest that, in the human genome, the *SPα* gene resides 11 kb apart from the *hIFGP1* gene (Fig. 3). Similarly, the mouse *SPα* gene is assigned to the region of Chr 3 that is syntenic with human Chr 1 and contains the *mFcγRI* gene (Blake et al. 2000). Our search in the EST database and genomic sequences for the human mIFGP2 ortholog was unsuccessful. Nevertheless, the SRCR-like domain of mIFGP2 shows a similar degree of homology with human and mouse Spα/AMI (63 and 62% identity, respectively). This suggests that the IgSF-SRCR mosaic protein might have arisen before or at the time of the split between primates and rodents.

Comparisons of the human, mouse, and rat proteins showed also that the evolution of the family was species-specific. The high similarity of the D5-type domains of hIFGP1, hIFGP3, and hIFGP4 suggests their emergence after the separation of primates and rodents. Mouse mIFGP1 is most similar to hIFGP1 (about 60% identical residues in the extracellular domains and 48% in the cytoplasmic regions). The two proteins may be orthologs. The absence of the D3-type domain in the protein encoded by the *mIFGP1* cDNA may be due to alternative splicing. The presence of multiple transcripts revealed by Northern blotting supports this possibility. However, in view of the Southern blotting data, it cannot be ruled out that different transcripts are produced by distinct closely related genes expressed in a tissue-specific manner. Furthermore, mIFGP1 differs from hIFGP1 in the absence of a charged amino acid residue in the transmembrane region and also by the structure of the tyrosine-based motifs in the cytoplasmic tail.

The members of the novel family share domains of the D1-, D2-, and D3-types with the leukocyte FcRs. Of FcRs, only FcγRI has the D3-type domain, the others are composed of the D1, D2 domains. We have recently described FCRL, yet another FcR-like protein containing the D3 domain (Mechetina et al. 2002). The N-terminal domain of FCRL, although possessing conserved regions characteristic of the D1 domain, is devoid of about 40 residues in its middle part. Furthermore, the C-terminal FCRL domain has a unique sequence not found in FcRs or IFGPs. Despite this, the D2 and D3 domains of FCRL are more similar to their FcR than IFGP counterparts. Thus, it may be inferred that, at a certain stage of phylogenesis, the common five-domain progenitor (D1D2D3D4D5) had split into two forms. Of these one, through a series of subsequent duplications, evolved into the FcR family, and the other gave rise to the IFGP family. The distant similarity of the three C-terminal extracellular PECAM domains with the domains of the type D3, D4 and D5 argues that divergence of PECAM and the FcR/IFGP-like proteins might have occurred earlier.

As has been shown, the leukocyte FcR classes and/or forms differ by their signal properties (Daeron 1997; Ravetch and Bolland, 2001). Human FcγRIIb and mouse FcγRII are inhibitory receptors containing intracellular ITIMs. The others activate effector functions through their capability to associate with the FcRγ or TCRζ subunits (FcεRI, FcγRI, FcγRIII) or through the presence of ITAM in the cytoplasmic tail (human FcγRIIa and FcγRIIc). Similarly, all the IFGP receptors contain tyrosine-based motifs conforming to the consensus binding sites of the SH2-domains in their cytoplasmic regions. Different patterns of the tyrosine-based motifs in the intracellular regions of the IFGP proteins suggest their distinct signaling properties. While precise definition of these properties is the subject of future studies, both ITIM- and ITAM-like sequences can be delineated at this juncture. Among the leukocyte FcRs only the FcγRII class possesses a long cytoplasmic tail encoded by 2–3 exons (Ravetch and Kinet 1991). Alignment of the FcγRII and IFGP intracellular sequences shows weak resemblance, mainly in the C-terminal part (Fig. 3). Conservation of the penultimate exons is pertinent to note. Based on this similarity, it may be suggested that evolving FcRs might have lost two cytoplasmic exons relative to the putative progenitor.

Regardless of the molecular events implicated in the divergence of the FcR and IFGP cytoplasmic regions, it appears that reiterative duplications in the rising family were accompanied by selection differentially affecting the extracellular, transmembrane and cytoplasmic parts of the molecules. Changes in the domain composition might have resulted in ligand specificity shifting, whereas divergence or loss of the transmembrane/intracellular parts led to the generation of novel proteins differing by their signaling abilities or by acquiring the role of humoral factors. As an evolutionary outcome, some of the family members, when compared with each other, have no structural elements in common (gp42 versus FcγRII/FcγRIII/FcεRI or mIFGP1versus FcγRI/FCRL).

The whole family may be thus regarded as the most vivid example of the evolutionary plasticity of IgSF to date. It shows that relationships between the mammalian IgSF subsets and distinct members may be hidden not only because of extensive sequence divergence but also because of the differential loss of the Ig-like domains during the duplicative process.

Acknowledgements This work was supported in part by grants to B.Y.A. from the Russian Foundation for Basic Research (1 00-04-49463**)** and from the RAS 6th Program for Fellow Scientists (1 221). The authors are grateful to Mrs. A.N. Fadeeva for help in preparation of the English version of the manuscript and to Dr. M.B. Soares for the NCI_CGAP_GCB1 cDNA library. The invaluable help of E.S. Belousov and V.V. Peremyslov is highly appreciated. All performed experiments comply with the current laws of the Russian Federation.

References

Alabyev BY, Guselnikov SV, Najakshin AM, Mechetina LV, Taranin AV (2000) CD3ε homologues in the chondrostean fish *Acipenser ruthenus*. Immunogenetics 51:1012–1020

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Barclay AN, Brown MH, Law SKA, McKnight AJ, Tomlinson MG, van der Merwe PA (eds) (1997) The leucocyte facts book, 2nd edn. Academic, London

Blake JA, Eppig JT, Richardson JE, Davisson MT (2000) The mouse genome database (MGD): expanding genetic and genomic resources for the laboratory mouse. The mouse genome database group. Nucleic Acids Res 28:108–111

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Daeron M (1997) Fc receptor biology. Annu Rev Immunol 15:203–234

Davis RS, Wang YH, Kubagawa H, Cooper MD (2001) Identification of a family of Fc receptor homologs with preferential B-cell expression. Proc Natl Acad Sci U S A 98:9772–9777

Gebe JA, Kiener PA, Ring HZ, Li X, Francke U, Aruffo A (1997) Molecular cloning, mapping to human chromosome 1 q21-q23, and cell binding characteristics of Spα, a new member of the scavenger receptor cysteine-rich (SRCR) family of proteins. J Biol Chem 272:6151–6158

Hatzivassiliou G, Miller I, Takizawa J, Palanisamy N, Rao PH, Iida S, Tagawa S, Taniwaki M, Russo J, Neri A, Cattoretti G, Clynes R, Mendelsohn C, Chaganti RS, Dalla-Favera R (2001) IRTA1 and IRTA2, novel immunoglobulin superfamily receptors expressed in B cells and involved in chromosome 1q21 abnormalities in B cell malignancy. Immunity 14: 277–289

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Lennon G, Auffray C, Polymeropoulos M, Soares MB (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics 33:151–152

Mechetina LV, Najakshin AM, Volkova OY, Guselnikov SV, Faizulin RZ, Alabyev BY, Chikaev NA, Vinogradova MS, Taranin AV (2002) FCRL, a novel member of the leukocyte FcR family possesses unique structural features. Eur J Immunol 32:87–96

Miyazaki T, Hirokami Y, Matsuhashi N, Takatsuka H, Naito M (1999) Increased susceptibility of thymocytes to apoptosis in mice lacking AIM, a novel murine macrophage-derived soluble factor belonging to the scavenger receptor cysteine-rich domain superfamily. J Exp Med 189:413–422

Newman PJ (1997) The biology of PECAM-1. J Clin Invest 99:3–8

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–6

Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 183:63–98

Ravetch JV, Bolland S (2001) IgG Fc receptors. Annu Rev Immunol 19:275–290

Ravetch JV, Kinet JP (1991) Fc receptors. Annu Rev Immunol 9:457–492

Sambrook T, Fritsch EF, Maniatis T (eds) (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Press, Cold Spring Harbor, NY

Sayos J, Wu C, Morra M, Wang N, Zhang X, Allen D, van Schaik S, Notarangelo L, Geha R, Roncarolo MG, Oettgen H, De Vries JE, Aversa G, Terhorst C (1998) The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. Nature 395:462–469

Seaman WE, Niemi EC, Stark MR, Goldfien RD, Pollock AS, Imboden JB (1991) Molecular cloning of gp42, a cell-surface molecule that is selectively induced on rat natural killer cells by interleukin 2: glycolipid membrane anchoring and capacity for transmembrane signaling. J Exp Med 173:251–260

Shibuya A, Sakamoto N, Shimizu Y, Shibuya K, Osawa M, Hiroyama T, Eyre HJ, Sutherland GR, Endo Y, Fujita T, Miyabayashi T, Sakano S, Tsuji T, Nakayama E, Phillips JH, Lanier LL, Nakauchi H (2000) Fc α/μ receptor mediates endocytosis of IgM-coated microbes. Nat Immunol 1:441–446

Shlapatska LM, Mikhalap SV, Berdova AG, Zelensky OM, Yun TJ, Nichols KE, Clark EA, Sidorenko SP (2001) CD150 association with either the SH2-containing inositol phosphatase or the SH2-containing protein tyrosine phosphatase is regulated by the adaptor protein SH2D1 A. J Immunol 166:5480–5487

Smith RF, Wiese BA, Wojzynski MK, Davison DB, Worley KC (1996) BCM search launcher – an integrated interface to molecular biology data base search and analysis services available on the world wide web. Genome Res 6:454–462

Taylor LS, Paul SP, McVicar DW (2000) Paired inhibitory and activating receptor signals. Rev Immunogenet 2:204–219

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Van Egmond M, Damen CA, van Spriel AB, Vidarsson G, van Garderen E, van de Winkel JG (2001) IgA and the IgA Fc receptor. Trends Immunol 22:205–211

Unkeless JC, Jin J (1997) Inhibitory receptors, ITIM sequences and phosphatases. Curr Opin Immunol 9:338–343

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans C A, Holt RA, Gocayne JD, Amanatides P, Ballew RM, et al (2001) The sequence of the human genome. Science 291:1304–1351

Wong WW, Klickstein LB, Smith JA, Weis JH, Fearon DT (1985) Identification of a partial cDNA clone for the human receptor for complement fragments C3b/C4b. Proc Natl Acad Sci USA 82:7711–7715

Xu MJ, Zhao R, Zhao ZJ (2001) Molecular cloning and characterization of SPAP1, an inhibitory receptor. Biochem Biophys Res Commun 280:768–775

Yusa S, Ohnishi S, Onodera T, Miyazaki T (1999) AIM, a murine apoptosis inhibitory factor, induces strong and sustained growth inhibition of B lymphocytes in combination with TGF-ß1. Eur J Immunol 29:1086–1093