Immunogenetics (2002) 54:570–581
DOI 10.1007/s00251-002-0499-8

ORIGINAL PAPER

Unni Grimholt · Finn Drabløs
Sven Martin Jørgensen · Bjørn Høyheim
René J. M. Stet

The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling

Received: 27 June 2002 / Revised: 19 August 2002 / Published online: 2 October 2002

Abstract A cDNA library screening using the conserved exon 4 of Atlantic salmon *Mhc* class I as probe provided the basis for a study on *Mhc* class I polymorphism in a breeding population. Twelve different alleles were identified in the 82 dams and sires studied. No individual expressed more than two alleles, which corresponded to the diploid segregation patterns of the polymorphic marker residing within the 3′-untranslated tail. Close linkage between the *Sasa-UBA* and *Sasa-TAP2B* loci strengthens the claim that *Sasa-UBA* is the major *Mhc* class I locus in Atlantic salmon. We found no evidence for a second expressed classical or non-classical *Mhc* class I locus in Atlantic salmon. A phylogenetic analysis of salmonid *Mhc* class I sequences showed domains conserved between rainbow trout, brown trout and Atlantic salmon. Evidence for shuffling of the α_1 domain was identified and lineages of the remaining α_2 through the cytoplasmic tail gene segment can be defined. The coding sequence of one allele was found associated with two different markers, suggesting recombination within the 3′-tail dinucleotide repeat itself. Protein modelling of several *Sasa-UBA* alleles shows distinct differences in their peptide binding domains and enables a further understanding of the functionality of the high polymorphism.

Keywords Atlantic salmon · Major histocompatibility complex · Class I · cDNA · Alternative polyadenylation signals

The nucleotide sequences reported in this paper have been submitted to GenBank and assigned the following accession numbers: *UBA*0601* (UA06) AF504013; *UBA*1201* (UA13) AF504014; *UBA*0701* (UA20) AF504015; *UBA*1401* (UA15) AF504016; *UBA*1101* (UA14) AF504017; *UBA*0401* (UA04) AF504018; *UBA*0101* (p30) AF504019; *UBA*0501* (UA05) AF504020; *UBA*0801* (UA08) AF504021; *UBA*0301* (UA03) AF504022; *UBA*0201* (UA21) AF504023; UBA*1001 (UA10) AF504024; UBA*0901 (UA01) AF504025 and *UBA*1501* (AF508864)

U. Grimholt (✉) · S.M. Jørgensen · B. Høyheim
Department of Morphology, Genetics and Aquatic Biology,
Section of Genetics, Norwegian School of Veterinary Science,
P.O. Box 8146 Dep., 0033 Oslo 1, Norway
e-mail: Unni.Grimholt@veths.no
Tel.: +47-22-964787, Fax: +47-22-964758

F. Drabløs
SINTEF Unimed, MR Center, 7465 Trondheim, Norway

R.J.M. Stet
René J.M.Stet, Department of Animal Sciences,
Cell Biology and Immunology Group, Wageningen University,
P.O. Box 338, 6700 AH Wageningen, The Netherlands

Introduction

The major histocompatibility complex (*Mhc*) class I and class II genes encode cell-surface molecules capable of binding and presenting short peptides to T cells. In general, the *Mhc* class I molecule, consisting of one α chain and β_2-microglobulin, presents peptides derived from cytosol to $CD8^+$ T cells. The class II molecule, consisting of one α and one β chain, mainly presents exogenously derived peptides to $CD4^+$ T cells. The classical *Mhc* genes represent some of the most polymorphic genes known to date, with multiple loci and a considerable number of alleles at each given locus. This diversity results in each individual being able to bind and present a wide variety of peptide ligands and has direct functional relevance for immune responses.

Nine years after the first discovery of salmonid *Mhc* sequences (Grimholt et al. 1993; Hordvik et al. 1993), there are still many unanswered questions regarding both number of functional loci, their polymorphic content and their functional relevance. Some of these questions have been addressed by studies in rainbow trout (Aoyagi et al. 2002; Hansen et al. 1999; Shum et al. 2001). Hansen and co-workers (1999) showed that the rainbow trout class I and class II genes are located on different linkage groups, as previously found for zebrafish (Bingulac-Popovic et al. 1997; Graser et al. 1998; Takami et al. 1997). Aoyagi and co-workers (2002) demonstrated the simple nature of trout *Mhc* class I, with one major polymorphic locus. Shum and co-workers (2001) provided evidence for exon shuffling producing “new” alleles.

However, the structural implication of this shuffling remains to be investigated. The primary aim of this study was to study the polymorphism of the major *Mhc* class I locus in Atlantic salmon and to use this information to investigate the differences in the antigen-binding groove using computer modelling.

Materials and methods

Fish

Atlantic salmon (*Salmo salar*) individuals used to identify alleles in this study were derived from a major Norwegian breeder (Aqua Gen, Kyrksæterøra, Norway) and from the Aursunda river close to Trondheim. The cDNA library was made from one individual from the breeding pool. The remaining sequences were detected in another 41 dams and 41 sires belonging to one of four populations (G-1) from this breeding pool. Twenty-five of these sires and dams produced on average 41 offspring, each totalling 1,031 animals, which were included in the segregation analysis. An additional 374 dams and sires belonging to another breeding population (G-4) from the same breeder and 29 animals from a different geographic location (the Aursunda river) were used to investigate microsatellite marker frequencies.

Genetic linkage was investigated in Atlantic salmon from the SALMAP project (FAIR-CT96-1591) including 21 parents tested for marker polymorphism from which two reference families were selected for the linkage analysis with 46 full-sib progeny each.

Nucleic acid isolation and library screening

The cDNA library was generated by oligo-dT priming of mRNA from head kidney, which was inserted into a Lambda-ZAP vector according to the manufacturer's recommendations (Stratagene, La Jolla, Calif.). An Atlantic-salmon-specific exon 4 probe was employed to screen 200,000 cDNA clones, using low stringency conditions, from which 24 positive signals were detected and nine positive cDNA clones were selected for further analysis.

Messenger RNA from the 82 dams and sires was isolated from 100 mg head kidney tissue (QuickPrep micro purification Kit, Pharmacia) and used to generate first-strand cDNA (Ready.To.Go T-primed first strand Kit, Pharmacia) according to the manufacturer's recommendations.

Genomic DNA was isolated from 25 mg muscle tissue (DNeasy Tissue Kit, Qiagen).

PCR and cloning

Allele amplifications were run on GeneAmp 9600/ 9700 machines (Perkin Elmer, Branchbury, N.J.) with a 35-cycle profile of 94 °C denaturation for 60 s, primer-specific annealing temperatures for 30 s and an elongation step at 72 °C for 60–90 s, depending on length of fragment to be amplified. Per 100-µl reaction mix, 100 ng template (cDNA or DNA), 5 pmol each primer and 5 units of AmpliTaq (Perkin Elmer) were used. An Atlantic-salmon-specific *Mhc* class I exon 4, from leucocyte cDNA, was generated by polymerase chain reaction (PCR), based on primers from conserved regions of a known salmon *Mhc* class I sequence (Grimholt et al. 1993). The sense primer sequence was 5′-ACCCCCTCCTCTCCAGTGACC-3′ and the reverse primer sequence was 5′-TTGATGAAGTCCTCCTGGAGACCCG-3′.

The sense primers (5′-TT(CT)AT(TC)(TC)T(TG)-CTG(GC)T-(GT)(CT)T(GA)GGA-3′ and 5′-CTGGGAATAGGCCTTCTACAT-3′) and the antisense primer (5′-CAATTACCACAAGCCCGCTC-3′) used to amplify the coding region from first-strand cDNA were based on the cDNA sequences of p30 (Grimholt et al. 1993) and ID8 identified in the cDNA library screening. An alternative sense primer based on the ICI sequence was later designed (5′-AGCCCTACATTCTTCATCTGC-3′), which amplified two additional *Sasa-UBA* sequences (*UBA*0301*, *UBA*1001*). The resulting PCR fragments were inserted into a vector (TA-cloning Kit, Invitrogen) and DNA from ten clones per individual was isolated (QIAprep Spin Miniprep Kit, Qiagen).

The microsatellite in the 3′-untranslated tail of *Sasa-UBA* was amplified using a fluorescent or radioactively labelled sense primer (5′-GGAGAGCTGCCCAGATGACTT-3′) and a reverse primer (5′-CAATTACCACAAGCCCGCTC-3′) with 80 ng cDNA or genomic DNA per 20-µl total reaction volume and run for 25 cycles at 54–60 °C annealing temperatures. Markers were analysed using automated ABI 377 machines (Applied Biosystems, Foster City, Calif.).

The *Sasa-TAP2* microsatellite repeat in intron 5 (Grimholt 1997) was analysed using the radioactively labelled sense primer sequence 5′-GCGGGACACCGTCAGGGCAGT-3′ and the reverse primer sequences 5′-GTCCTGATGTTGGCTCCCAGG-3′ for *Sasa-TAP2A* and 5′-CCTGATATTGTCTGCCAGATC-3′ for *Sasa-TAP2B*. All radioactively labelled marker analysis was performed using 6% polyacrylamide gels with labelled M13 sequencing ladder and subsequent manual autoradiographic analysis.

Sequencing and handling of sequence data

Sequencing was performed on double-stranded DNA with T3, T7, and internal primers, using automated sequencing on an ABI 377 machine with the Big Dye Terminator Kit (Perkin Elmer). Sequence alignments and analyses were performed using the UWGCG software (Devereux et al. 1984) and ClustalW (Thomson et al. 1994). Phylogenetic trees were created using the neighbour-joining method (Saitou and Nei 1987) based on nucleotide sequences and bootstrapped trees were visualised using Treeview (Page 1996).

Segregation and linkage analysis

The segregation of the class I alleles was assessed in 25 full-sib Atlantic salmon families with an average of 41 offspring each. In total, 1,031 individuals were included in the analysis. Each individual was typed for class I using either informative markers, or a combination of marker analysis and allele-specific PCR using genomic DNA as template. Sequencing of the resulting products in at least two families carrying these particular alleles was performed to assess the specificity of each of the allele-specific primer combinations. Each family was tested for Mendelian segregation of class I alleles using a chi-square test for goodness-of-fit between the observed and expected frequencies. Resulting chi-square values from the 25 families were evaluated with Bonferroni correction for multiple testing.

A linkage analysis was performed using simple sequence repeats residing within the 3′-untranslated tail of *Sasa-UBA* and in intron 5 of *Sasa-TAP2* (Grimholt 1997) in informative families from the SALMAP family panel (John Taggart, personal communication). Additionally, a minisatellite repeat in the *Mhc* class II alpha gene was included in the analysis (Grimholt et al. 2000). Segregation of alleles in the SALMAP families was used to link the *Mhc* markers to the preliminary SALMAP linkage map. Because of the large differences observed in the recombination rates between sexes, sex-specific maps were constructed when markers were informative in both dam and sire. Markers were assigned to linkage groups using the software JoinMap version 2.0 (Stam and Van Ooijen 1995) based on the Kosambi function for map distance calculations. A two-point linkage was used with LOD values larger than 3.5.

Homology-based modelling and visualisation

Three-dimensional models of *Mhc* structures were made by using the Swiss-Model (Guex and Peitsch 1997; Peitsch 1995) web serv-

er (http://www.expasy.ch/swissmod/SWISS-MODEL.html). This server implements an automatic approach for model building. A Blast (Parham et al. 1995) search against protein sequences from a subset of the PDB database (Berman et al. 2000) is used for identification of possible similarities, SIM is used for selection of relevant templates, ProModII (Peitsch 1996) builds initial models based on these templates, and the geometry of the models is optimised using the Gromos96 forcefield (Scott et al. 1999). The quality of the models was evaluated using the SwissPDBViewer (DeepView) (Guex and Peitsch 1997) modelling program, which can also be used for further optimisation of models, if necessary. Atomic charges for potentially charged groups were estimated with Titra (Anthonsen et al. 1994), assuming a pH of 6.5. The molecular surface and the electrostatic potential at the surface were computed with Grasp (Nicholls et al. 1991; Nicholls and Honig 1991), and visualisation of the models was done by using a combination of Molscript (Kraulis 1991), Grasp and Raster3D (Merritt and Bacon 1997).

Results

cDNA screening

To determine the number of expressed sequences and design primers for a population study, an exon 4 probe encoding the α_3 domain was used to screen a cDNA library made from a single Atlantic salmon individual. Twenty-four strongly hybridising signals were identified per 200,000 clones screened. DNA from nine positive clones was isolated and sequenced. Of these, four were partial clones, with identical sequences to the other full-length sequences. The five full-length clones had inserts ranging from 2,167 bp (ID5, ID6, ID8) to 2,477 bp (ID7, IC1) with a 5′-untranslated sequence ranging from 55 bp (ID8) to 124 bp (ID5, ID6), an open reading frame of 1,065–1,074 bp, and a 3′-untranslated tail sequence of 1,100–1,420 bp. The clones 1D5 and 1D6 contained identical sequences to the clones 1C1 and 1D7, with the exception of the additional 310-bp 3′-untranslated tail sequence in the latter two clones. The clone 1D8 contained a different sequence compared with the other clones, with 14% overall sequence divergence from the ICI clone. The 3′-untranslated tail sequences contained 12 nucleotide substitutions over a 1,100-bp range. There are three canonical polyadenylation signals in the 3′-untranslated tail sequences (AATAAA). The clone p30 (Grimholt et al. 1993) terminates at the same position as the ID5/ID6/ID8 sequences and most likely used the polyadenylation signal at position 2157. The ICI clone most likely used the polyadenylation signal at position 2477. The third polyadenylation signal at position 1639 was not used in any of the clones identified. An identical dinucleotide repeat resided within the 3′-untranslated tail of both ICI and ID8 at position 1161. The cDNA screening thus identified two different full-length cDNA sequences and a potentially polymorphic microsatellite marker and provided the necessary information used to design primer sequences for the population study described below.

Expressed *Mhc* class I alleles in farmed fish

First-strand cDNA was used as template for PCR amplification of *Mhc* class I sequences from 41 dams and 41 sires belonging to a major Norwegian breeding pool. Fragments were cloned and sequenced. Ten clones per fish were analysed to exclude as many artefacts as possible. Sequences potentially resulting from PCR artefacts were tested by allele-specific PCR or ultimately by amplification and sequencing of genomic PCR fragments. A total of 11 different allele sequences were identified in the material denoted *Sasa-UBA*0201* to *Sasa-UBA*1201* (Fig. 1). A phylogenetic analysis of the nucleotide sequences identified in the library screening and the population study showed that the ICI and *UBA*0301* sequences were identical (hereafter both denoted *UBA*0301*), while ID8 represents a unique allele hereafter denoted *UBA*1401*. The p30 clone (Grimholt et al. 1993) was renamed to *UBA*0101*. The overall nucleotide sequence diversity between the alleles ranged from 0.16% (*UBA*0401* versus *UBA*1101*) to 18.63% (*UBA*1201* versus *UBA*0901*). Some of these alleles were more frequent than others with allele frequencies ranging from 0.01% (*UBA*1101*) to 0.35% (*UBA*0801*).

Dinucleotide marker

An analysis of the 3′-tail dinucleotide marker in the 41 dams and 41 sires identified considerable polymorphism, but also some discrepancy between markers and amplified allele sequences. Some individuals showed two different markers, but only one allele sequence. A new sense primer was therefore designed purely based on the *UBA*0301* leader sequence to check for additional sequences not amplified by the primary sense primers. This enabled identification of the additional sequences, *UBA*0301* and *UBA*1001*, resulting in a consistent relationship between markers and alleles with one exception. The *UBA*0901* allele was not identified in the marker analysis. It was, however, identified in the cDNA *Sasa-UBA* allele amplification of sires and dams identifying a CA repeat with divergent flanking regions, compared with the other sequences. Segregation of *UBA*0901* was thus tested using allele-specific primers. The identification of the last two alleles, *UBA*0301* and *UBA*1001*, using a different sense primer, resulted in more than 80% of the animals being heterozygous for the *Sasa-UBA* locus. No individual expressed more than two different sequences.

Most marker lengths are linked to one or two alleles only (191=*UBA*0501*, 193=*UBA*1001*, 195=*UBA*0601* and *UBA*1201*, 201=*UBA*0701*, 218=*UBA*1101*), while one marker is linked to seven different alleles (199=*UBA*0201*, *UBA*0301*, *UBA*0401*, *UBA*0801*, *UBA*1401*). The marker nomenclature relates to the size of the amplified fragments run on a denaturing gel. Amplification of the marker on a genomic level from

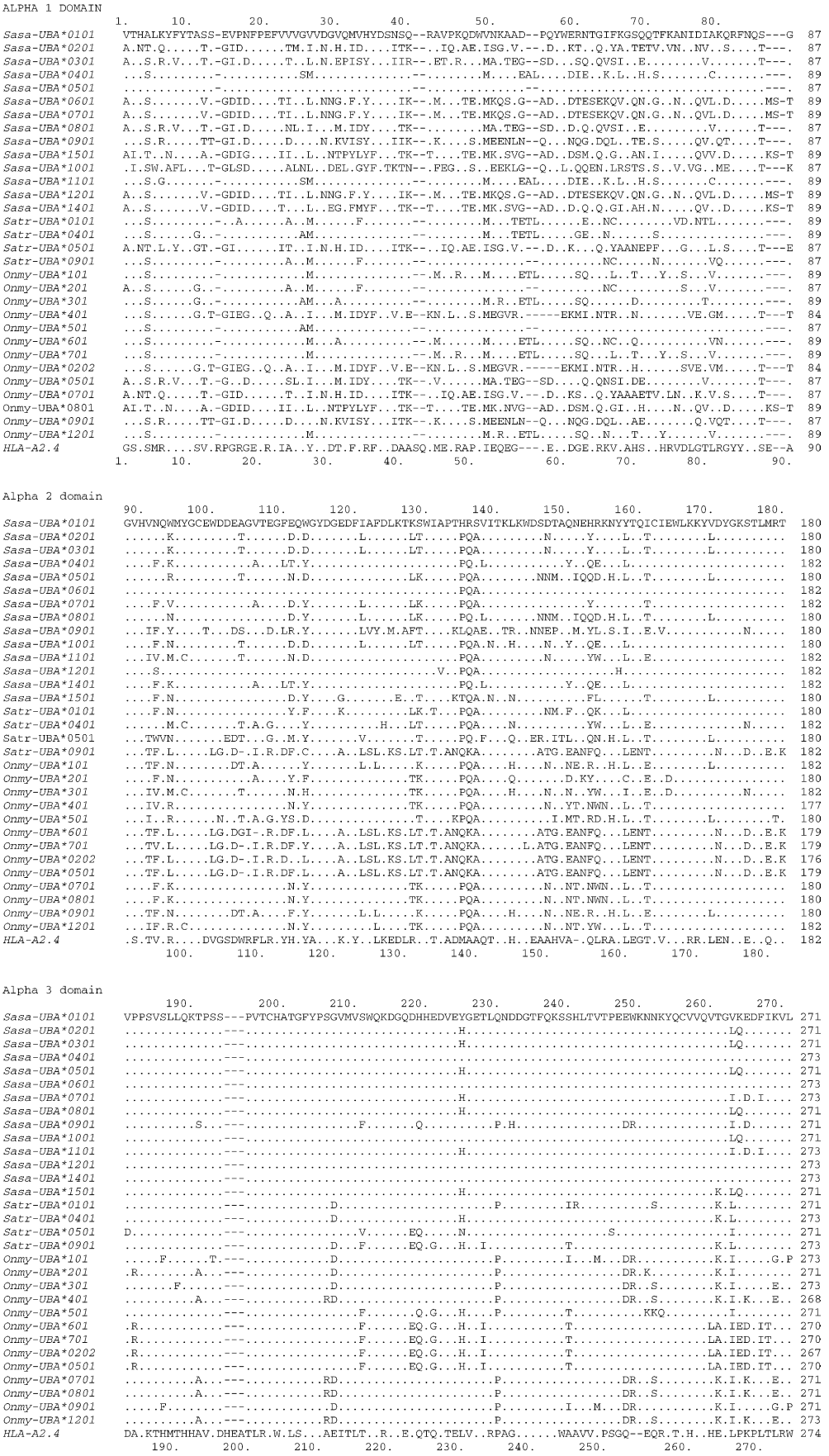
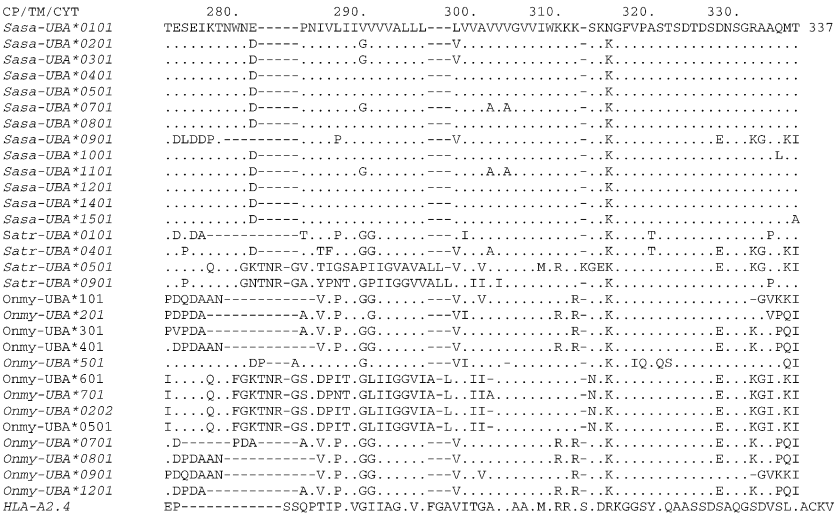


Fig. 1 Alignment of *Mhc* class I amino acid sequences. *Dots* indicate identity, while gaps introduced to maximise alignment are indicated by *dashes*. *Satr-UBA* sequences (*Salmo trutta*) and four digit *Onmy-UBA* (*Oncorhynchus mykiss*) sequences are taken from Shum and co-workers (2001) (AF296359–AF296383). Three digit *Onmy* sequences derive from Aoyagi et al. (2002) (AF287483–AF287492). *HLA-A2.4* α_1 and α_2 domain included for comparison of peptide binding region (Parham et al. 1989); *Sasa-UBA*0101* (p30) (Grimholt et al. 1993)

Fig. 1 (continued)



these individuals, identified different fragment sizes resulting from an intron of 124 bases within the amplified region. The marker sizes on a genomic level thus ranged from 314 to 340 bases. An additional breeding population belonging to the same breeder (G-4) was also typed for the *Sasa-UBA* marker. Two additional markers were identified using genomic DNA (310 and 312) and the marker allele frequencies were quite different. In our population (G-1) the 322 marker is dominant, while in the new population (G-4) the 318 marker is by far the most frequent. Marker analysis of the 29 fish from the Aursunda river also identified two new markers (334 and 336).

Two animals from the Aursunda river containing the 310 and 336 markers not found in the G-1 population (310/318 and 336/336) were chosen for further investigation of expressed *Mhc* class I cDNAs. Two different allele sequences were identified from each individual. The homozygous 336/336 animal had one allele completely identical to *Sasa-UBA*0801*, which in the G-1 population is linked to a 322 marker. The other allele in this animal had not been found previously and was designated *Sasa-UBA*1501*. In the animal with 310/318 markers, the two identified expressed sequences had sequence similarities to *Sasa-UBA*0901* and *Sasa-UBA*1101*.

Amino acid sequences

Comparison of the deduced amino acid (aa) sequences with *Mhc* class I aa sequences from other vertebrates identified a leader peptide of 18–19 aa, an α_1 domain of 87–89 aa, an α_2 domain of 93 aa, an α_3 domain of 91 aa, a connecting peptide (CP) of 9–13 residues, a transmembrane region of 24 aa, and a cytoplasmic region of 29 aa (Fig. 1). The α_1 domain has by far the highest number of variable aa positions (71) when compared with the α_2 domain (48) and the α_3 domain (12). The α_1 domain is also the only domain, apart from the CP, that varied in length. The sequences *UBA*0101* (p30), *UBA*0201*, *UBA*0301*, *UBA*0501*, *UBA*0801*, *UBA*0901* and *UBA*1001* all have two amino acids less than the remaining α_1 domain sequences. The CP through cytoplasmic region has six variable amino acid positions, disregarding the *UBA*0901* sequence, which contains a four-residue deletion in the CP, as well as many variable positions not identified in any of the other *Sasa-UBA* sequences. There is one cysteine located in the leader sequence of the *UBA*0301* aa sequence. In the α_1 domain there is one cysteine at aa position 79 in the sequences *UBA*0401* and *UBA*1101* (all positions relating to *UBA*0101* sequence, Fig. 1). All the α_2 and α_3 domains contain two cysteines each at aa positions 98, 162, 198 and 256, with the exception of *UBA*1101*, which has an additional cysteine at position 96. The potential glycosylation site NQS or NQT at aa position 84 is conserved in all sequences.

Phylogenetic analyses

Based on the data generated by Shum and co-workers (2001) and Aoyagi and co-workers (2002), a phylogenetic analysis was performed using individual α_1, α_2 and α_3 domain sequences from Atlantic salmon, rainbow trout and brown trout. The sequence identity between salmonid aa domain sequences ranges from 25% to 100% for the α_1 domains, from 51% to 100% for α_2 and from 75% to 100% for α_3. The sequence identity solely between Atlantic salmon domains range from 38% to 81% identity for α_1 domain aa sequences, from 66% to 100% for the α_2 domains and from 88% to 100% for the α_3 domains. *Sasa-UBA*1001* has only 30% identity to the α_1 domain of *Sasa-UBA*0201* and 25% sequence identity to *Satr-UBA*0501* representing the most distant α_1 domains. Among the most divergent α_2 domains are those of *Onmy-UBA*0701* and *Sasa-UBA*0901* with

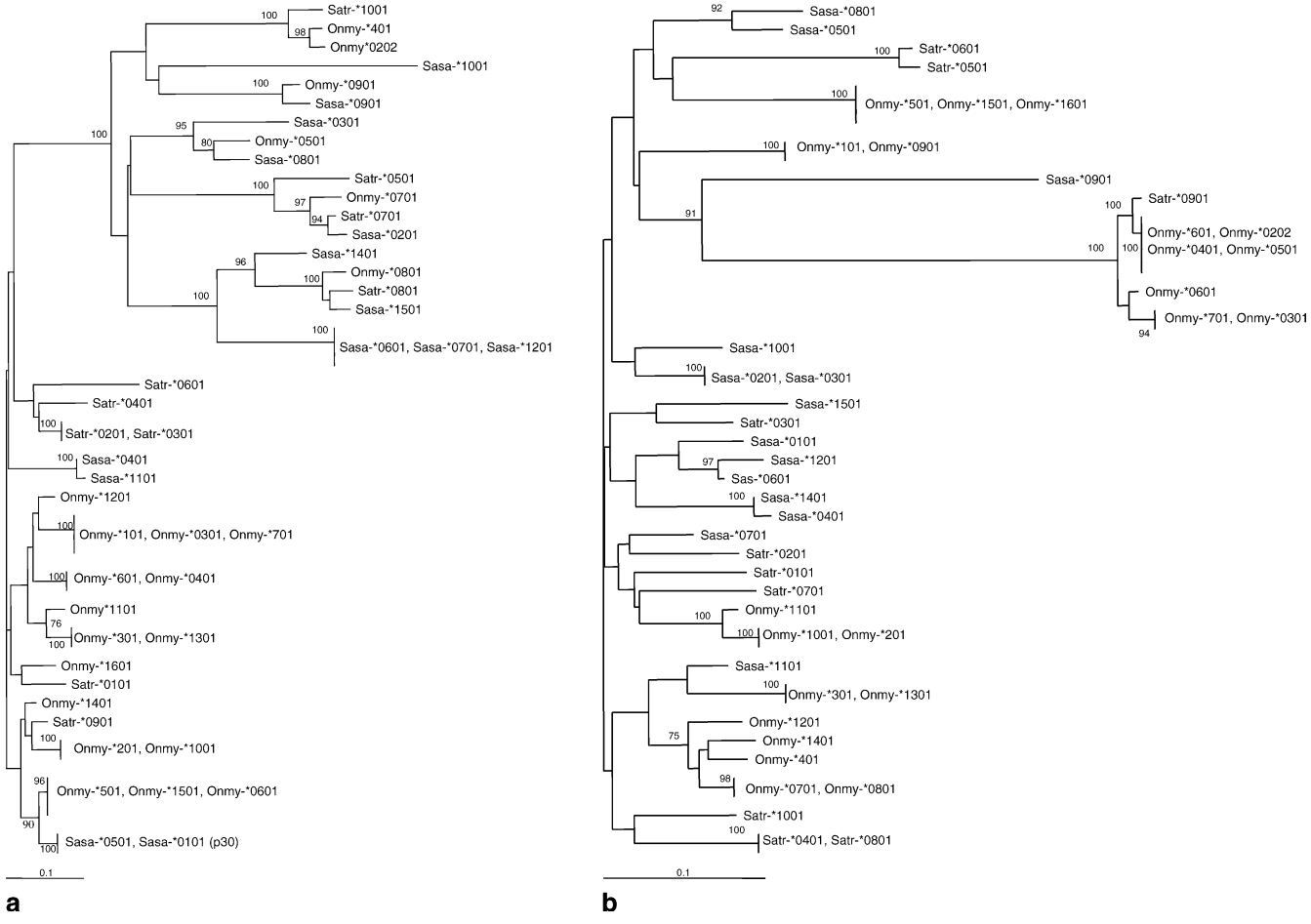


Fig. 2 Phylogenetic tree of individual α1 (**a**) and α2 (**b**) domains based on salmonid amino acid sequences aligned in Fig. 1. Genetic distance is indicated at the *bottom*, while the reliability of the cluster analyses are tested by bootstrap confidence limits and indicated as percent success per 1,000 bootstrap trials with values above 70% presented on nodes. See legend to Fig. 1 for sequence references and accession numbers

only 51% sequence identity. For the α_1 and α_2 domains (Fig. 2a, b), the *Sasa* sequences intermingle in between the *Onmy* and *Satr* sequences. For the α_3 domain, *Sasa* sequences define separate clusters, with the exception of *Sasa-UBA*0901*, but the bootstrap values are low rendering the α_3 domain tree somewhat questionable (data not presented). The CP, transmembrane and cytoplasmic tail sequences of the Atlantic salmon sequences are much more conserved than those of rainbow trout and brown trout (Fig. 1).

Some *Sasa* sequences have identical α_1, α_2 or α_3 domains (Fig. 2a, b). *Sasa-UBA*0601*, *UBA*0701* and *UBA*1201* share identical α_1 domains, as do *Sasa-UBA*0101* and *UBA*0501*. *UBA*0201* and *UBA*0301* have identical α_2 domains. For the α_3 domain, the sequences *UBA*0101*, *UBA*0401*, *UBA*0601*, *UBA*1201* and *UBA*1401* are completely identical, as are the α_3 domains of *UBA*0201*, *UBA*0301*, *UBA*0501* and *UBA*0801*. As a matter of fact, the entire region encompassing the α_2, α_3 and transmembrane/cytoplasmic domains of *UBA*0201* and *UBA*0301* are completely identical (Fig. 1). The α_1 domain from the two sequences identified in the 310/318 marker animal from the Aursunda river were identical to *Sasa-UBA*0201* (Aur22#2) and *Sasa-UBA*0401* (Aur22#4) (data not presented). The rest of their sequences from α_2 through cytoplasmic domain were identical to *Sasa-UBA*0901* (Aur22#4) and *Sasa-UBA*1101* (Aur22#4). For the animal homozygous for the 336 marker, one allele sequence was entirely new, and designated *Sasa-UBA*1501*, while the other showed complete aa sequence identity to *UBA*0801*.

Linkage and segregation analyses

Based on both male and female recombination the *Sasa-UBA*, *Sasa* class II α and *Sasa-TAP2* markers were assigned to three different linkage groups (Fig. 3). An expected linkage disequilibrium was observed between *Sasa-UBA* and *Sasa-DAA* (class II α), which is consistent with previous studies in other teleosts (Bingulac-Popovic et al. 1997; Graser et al. 1998; Hansen et al. 1999; Takami et al. 1997). The *Sasa-TAP2B* marker co-segregated with the class I region, while the *Sasa-TAP2A* locus resides within a different linkage group.

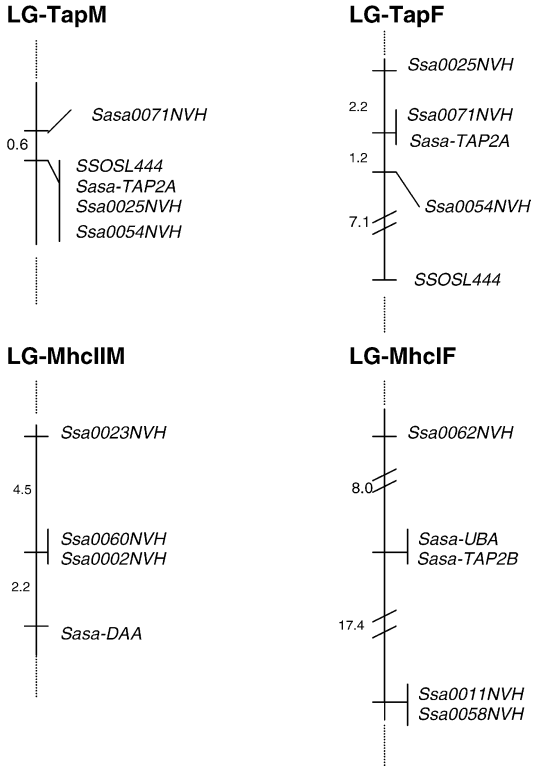


Fig. 3 Linkage of *Sasa-UBA*, *Sasa-DAA*, *Sasa-TAP2A* and *Sasa-TAP2B* to three linkage groups and their respective markers (*SsaNVH*) on the preliminary SALMAP map. Linkage groups are here denoted *LG-MhcI*, *LG-MhcII* and *LG-Tap*. When possible, comparative male (*M*) and female (*F*) linkage groups were constructed. Distances between markers are given in centiMorgans

The calculated *Sasa-TAP2B* recombination frequencies were based on marker segregation in offspring from one sire only, since the microsatellite marker contained limited polymorphism. In the male and female map the *Sasa-TAP2A* marker showed varying distances and a different order relative to the flanking SALMAP markers (LG-TapM/F; Fig. 3).

Segregation of the different *Sasa-UBA* alleles was assessed in 25 families with an average of 41 offspring each for discrepancies between observed and expected number of individuals carrying paternal and maternal class I alleles. Bonferroni correction for multiple testing indicated that no families showed segregation distortion. This detailed segregation analysis confirms that the identified alleles belong to one locus.

Protein modelling

Four protein sequences were selected for model building: *Sasa-UBA*1401*, *UBA*0301*, *UBA*0101* (p30) and *UBA*0901*. These sequences represent a significant part of the sequence diversity in the data set, as they are found in separate clusters in the phylogenetic tree (data not shown). The model building was successful, for all four sequences several templates with approximately 37% sequence identity could be found, and this is normally sufficient for building good quality models. The superimposed backbone traces of the antigen binding region for these models is shown in Fig. 4a. The electrostatic profile of the molecular surface at ±10 kT/e for each model is shown in Fig. 4b, using the same orientation as in Fig. 4a.

The superimposed backbone traces of the *Mhc* models (Fig. 4a) show that the antigen binding sites most likely are structurally conserved with respect to backbone trace. This is not surprising, as the sequence identity from pairwise alignment is between 68 and 79% for these four sequences. Although the sequence identity is lower for the antigen binding region, where most of the sequence diversity is found, it is still within the range where structural conservation of the backbone normally is assumed (Sander and Schneider 1991). Some structural variation is seen in loop regions, which does not affect the conformation of the binding cleft. The two-residue insertion of *UBA*1401* between residues 86 and 87 (*UBA*0101* numbering, Fig. 1) is found in one of these loop regions, and the four-residue deletion of *UBA*0901* in the CP is just outside the modelled region, these indels will therefore not affect the general conformation of the backbone. It is, therefore, reasonable to assume that the models are robust with respect to backbone conformation, and that most of the relevant structural diversity in the antigen-binding region can be analysed by looking at side chain variation.

The electrostatic profiles of these models (Fig. 4b) show very clear differences, and these differences can most likely be associated with specificity of antigen binding. The electrostatic properties have to be analysed with some care. The explicit charges of potentially charged residues are assigned by Titra based on certain assumptions about accessibility and local environment. Also, the actual effect of a specific charge depends upon solvent accessibility, hydrogen bonding and the local environment in general. Therefore the electrostatic properties at a detailed level are sensitive to the quality of the model. These models have been optimised by a molecular mechanics approach, which normally removes unrealistic interactions. However, this does not guarantee that all side chain orientations or interactions are optimal, or as seen in the true structure. Therefore the electrostatic models should be analysed by looking at general features, rather than individual interactions.

All four *Sasa-UBA*0301*, *UBA*1401*, *UBA*0901* and *UBA*0101* protein models show characteristic electrostatic properties in the antigen-binding cleft. *Sasa-UBA*0301* has an overall negatively charged binding region, whereas *UBA*1401* and *UBA*0101* are mainly positively charged. There are some differences between *Sasa-UBA*1401* and *UBA*0101* as well; *UBA*1401* seems to have some explicit strong positive charges in the actual binding cleft, whereas *UBA*0101* seems to have a more diffuse positive charge. *Sasa-UBA*0901* seems to have mixed properties, where half of the bind-

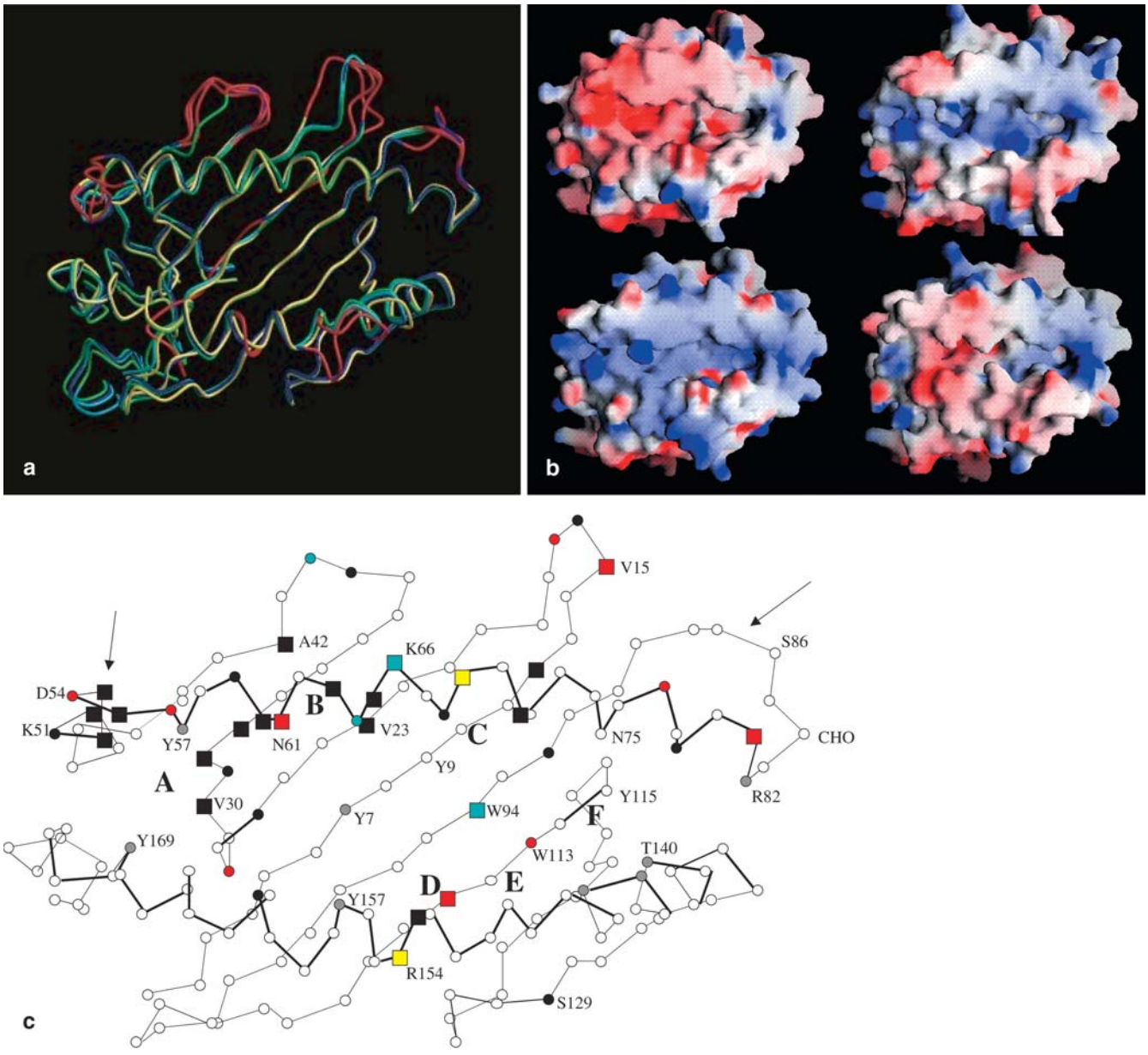


Fig. 4 **a** Superimposed backbone traces of the four *Mhc* models. The models of *Sasa-UBA*1401*, *Sasa-UBA*0101* and *Sasa-UBA*0901* have been superimposed on the *Sasa-UBA*0301* model by minimisation of the residual means squared (rms) at 53 corresponding Cα positions in the α-helices forming the antigen binding cleft (helix 1 and 2 in the α_1 and α_2 domains, respectively), excluding the central irregular part of the lower helix (helix 2). The four models are coloured individually (*Sasa-UBA*0301 – green*, *Sasa-UBA*1401 – blue*, *Sasa-UBA*0101 – cyan*, *Sasa-UBA*0901 – yellow*), except for the sections re-modelled by loop building during automatic modelling, which are shown in *red*. Most of these loop regions are outside the binding cleft. A cutting plane has been inserted behind the central β-sheet structure, masking most of the α_3 domain from the display. **b** Presentation of the electrostatic charges of the peptide binding groove of four allele sequences. *Sasa-UBA*0301 top left*, *Sasa-UBA*1401 top right*, *Sasa-UBA*0101 below left*, *Sasa-UBA*0901 below right*. Negatively charged residues are shown in *red*, while positively charged residues are shown in *blue*. Positioning of residues as presented in Fig. 4a and Fig. 4c. See text for details. **c** Schematic presentation of the structural backbone of the *Sasa-UBA*0101* (p30) peptide binding groove. Positions with *Sasa-UBA* positive and neutral amino acids are shown in *blue*, positions with negative and neutral amino acid are shown in *red* and those with positive, negative and neutral amino acids are shown in *yellow*. Positions with variability index higher than seven are shown as *squares* (var_l=number of different residues in a given position divided by the frequency of the most common residue (Parham et al. 1988). Conserved positions assumed to be involved in binding of peptide N and C termini are shown as *grey circles*. Regions with insertions/deletions are indicated by *arrows*

ing region is negative, the other half is positive. In all structures, a positive charge is seen towards the occluded end of the binding cleft. This is a totally conserved arginine (R82); it is found in all sequences analysed in this study.

In order to understand the variation in electrostatic properties at a single-residue level the residues most likely to be important for this variation were identified. The four models are based on a total of 13 template structures, and all of these structures have been determined with bound peptide. By using the CSU (Contacts of Structural Units) program (Sobolev et al. 1999) all side chains in contact with the peptide were identified in all these structures. This gave a total of 45 potential contact residues, which could be related to corresponding positions in the models by using the sequence alignments from the model building. Out of these 45 positions it was found that 20 positions are totally conserved within the four modelled sequences, and 12 of these are totally conserved also in the full set of sequences described here. An additional nine positions have only non-charged residues, and although they will contribute towards variation in, for example, hydrophobic properties, they are less important for electrostatic variation. The remaining 16 positions can be separated into three groups, ten positions with neutral or negative residues (V15, N17, D28, D54, Q56, N61, D77, Q81, E111, W113), four positions with neutral or positive residues (Q40, I64, K66, W94), and two positions with negative, neutral and positive residues (Q69, R154) (Fig. 4c). There is a cluster of two to three residues at the bottom of the binding cleft (D116, E118, potentially D119) that may be important for defining the electrostatic component of peptide binding. Some variation is also seen along the α-helices forming the ridges of binding cleft.

Discussion

The deduced aa sequences have all the characteristics of functional *Mhc* class I molecules. An α_1 domain without conserved cysteines, the conserved cysteines in the α_2 and α_3 domains, a glycosylation site in the α_1 domain, as well as a CP, a transmembrane and a cytoplasmic tail sequence. The *UBA*0901* allele encodes a shorter CP, as seen in some *Mhc* class I sequences of other salmonids (Fig. 1). In other vertebrates, the CP segment is assumedly involved in the association of *Mhc* class I molecules with calnexin and the suggested kinetic association of class I molecules with calnexin is likely to contribute to the different maturation rate between several class I alleles (Carreno et al. 1995; Qian and Chen 2000). This is one potential explanation for the observed variability in the salmonid CP.

Twelve different *Sasa-UBA* alleles were identified in 82 dams and sires with overall sequence diversity ranging from 0.1% to 18%. The polymorphic marker residing within the 3′-untranslated sequence showed diploid Mendelian segregation patterns in all 25 families studied, and α_1-allele-specific PCR in families with uninformative markers identified a stable inheritance of marker versus coding region. The linkage analysis shows a close linkage of the *Sasa-UBA* and *Sasa-TAP2B* loci, as previously identified for the major class I region in rainbow trout (Hansen et al. 1999), supporting the claim of *Sasa-UBA* representing the main MHC class I region in Atlantic salmon. The *Sasa-TAP2A* could be linked to a second uncharacterised class I locus although we have no data supporting this. The microsatellite in the 3′-tail sequence segregates in a diploid manner and no animals display more than two different expressed sequences. Perhaps we did not amplify these alleles because their expression is repressed or they have very distant nucleotide sequences, thus avoiding detection by PCR and hybridisation. Assuming there is a second *Mhc* class I region in salmonids, it seems to play a different or minor role in the immune system similar to that found in both chicken and frog (Flajnik et al. 1999; Kaufman et al. 1999). The final verdict on the number of functional *Mhc* class I genes in Atlantic salmon will have to come from detailed physical mapping of the *Mhc* and/or *Mhc*-related regions.

The phylogenetic analyses, including sequences from rainbow trout, brown trout and Atlantic salmon, identified clusters containing all three salmonid species. The sequence similarity among the three salmonid species is particularly profound for α_1 domains (Fig. 2a). The relationship between α_2 domains is more obscure as represented by weaker bootstrap values (Fig. 2b). For the α_3 domains through cytoplasmic tail, the Satr and Onmy sequences share more similarities between themselves than to Atlantic salmon (data not presented). The divergence between rainbow trout and brown trout/Atlantic salmon is estimated to be 15–20 Myr (Devlin 1993; McKay et al. 1996), whereas brown trout and Atlantic salmon are estimated to have diverged between 2.5 and 6 Myr (McVeigh et al. 1991). None of the aligned sequence data support a closer relationship between *Salmo trutta* and *Salmo salar,* but the selected fish may represent a biased sample.

Two studies in rainbow trout (Aoyagi et al. 2002; Shum et al. 2001) identified more polymorphic positions in the second exon than in the third as found in the Norwegian Atlantic salmon alleles. A Canadian study of the *Mhc* class I polymorphism in salmonids revealed a much lower variability both concerning number of alleles, as well as the sequence diversity between them and identified more exon 3 than exon 2 diversity (Miller and Withler 1998). Lineage definition is difficult in salmonids, since some alleles are completely identical in one domain and then highly divergent in others. Shum and co-workers (2001) suggested recombinational shuffling of the α_1 domain as a possible explanation using recombination signals in the still unidentified second intron sequence. Thus, α_1 domains would be shifted between different α_2 through cytoplasmic region sequences increasing the number of different antigen binding domains available to the population. Of the 32

Onmy and Satr sequences included, 19 sequences share eight different α_2 through α_3 domains, but have highly divergent α_1 domains. This is also found in Atlantic salmon sequences, with two pairs of sequences with virtually identical α_2 through cytoplasmic tail regions identified in the G-1 population (*0201/*0301, *1401/*0401) and two additional pairs identified in the two Aursunda animals (*0901/aur22#4, *1101/aur22#2).

The sequence conservation seen in the α_3 domain is extended to the remaining part of the sequence. Only 12 variable nt positions are present over a 1,100-bp region in the 3′-untranslated tail sequences of *Sasa-UBA*0101*, *Sasa-UBA*0301* and *Sasa-UBA*1401*, sequences that appear as quite divergent in the phylogenetic analysis of their coding regions. The *Sasa-UBA*0101* and the *Sasa-UBA*1401* sequences have only six variable nt positions between them. Why such limited sequence diversity in a region that is not traditionally under selective pressure? Perhaps sequence conservation in the 3′ region is a prerequisite for recombination or gene conversion events, both phenomena assumed to promote variability in warm-blooded vertebrates (Carrington 1999; Yoshino et al. 1995). The identification of the *Sasa-UBA*0801* allele in the Aursunda population linked to a new marker suggests this marker is indeed involved in recombination. In higher vertebrates, microsatellite are suggested to assist in or even promote the recombination events (Pardue et al. 1987). Such unexpected sequence conservation in the 3′-untranslated regions has also been observed in carp (Van Erp et al. 1996) and chicken (Jim Kaufman, personal communication). If or how it affects the coding region and how recombination within *Sasa-UBA* would influence functionality of its linked TAPs are both unanswered, but no less intriguing questions.

The calculated genetic distances on the linkage group LG-Tap revealed large sex-specific differences in recombination rates. Based on male recombination, *Sasa-TAP2A* clusters with five other markers covering a region of only 0.6 cM, while female recombination rates give a linkage group spanning 10.5 cM, with only one SALMAP marker segregating with *Sasa-TAP2A* (Fig. 3). The compressed male map distances seen on LG-TapM versus LG-TapF is typical for salmonid linkage maps, and are believed to arise from multivalent formations occurring in male meiosis, which affects regions proximal to centromeric positions (Sakamoto et al. 2000). These sex-specific differences in recombination rates may also affect the evolution of specific genes such as the *Mhc*.

Presently there are no crystallographic models of fish *Mhc* molecules and no data on peptide specificity, leaving us with computer modelling and comparison to those of the human analogous. Due to the insertions/deletion in the salmonid α_1 domain, the position and thus numbering of Sasa-UBA residues analogous to the HLA-A residues become slightly different. The positioning of the conserved Y84 in HLA-A2 is similar to a completely conserved R82 in salmonids, as well as chicken, lizard, frog and shark (Kaufman et al. 1994), which is most likely involved in binding the negatively charged C-terminus of the antigen peptide (Fig. 4c). Modelling Atlantic salmon class I peptide binding grooves exhibiting quite different electrostatic charges most likely influence the kind of peptides they are capable of binding. The insertions/deletions between residues 86 and 87 (UBA*0101 numbering) in the α_1 domain are buried in a loop probably not affecting peptide binding, while the two additional amino acids between reside 54 and 55 may have an effect. The structures of *Sasa-UBA*0101* and *UBA*1401* in the residue 53–56 region are also different (Fig. 4a). An analysis of the *Sasa-UBA* α1 positions with a variability index of seven or more [var_i (Parham et al. 1988)] identify 17 positions with an average var_i of 15 (numbering according to UBA*0101; 23, 30, 31, 33, 42, 50, 52, 53, 55, 60, 61, 63, 65, 66, 69, 71, 81). In the α_2 domain only three such highly variable positions are identified with an average var_i of nine (94, 111, 154). A similar analysis of the Onmy sequences aligned in Fig. 1 showed nine α_1 positions (23, 30, 33, 42, 50, 53, 61, 65, 68) and three α_2 positions (84, 111, 150) with var_i=11. The slight differences between the Sasa and Onmy variability profiles are most likely attributable to random variation in a small sample. For comparison, the highly variable HLA-A, B, and C positions are 9, 24 (23), 45 (42), 62 (60), 65 (63), 66 (64), 67 (65), 70 (68), 71 (69), 74 (72), 77 (75), 80 (78) in the α_1 domain and 95 (92), 97 (94), 114 (111), 116 (113), 156 (154), and 163 (161) in the α_2 domain all with var_i=13 (numbering according to *HLA-A2* in alignment Fig. 1, *Sasa-UBA*0101* numbering in parenthesis).

The structures of the peptide-binding specificity pockets of human and mouse *Mhc* class I molecules have been widely analysed (Madden et al. 1991; Matsumura et al. 1992; Saper et al. 1991). The composition and shape of the six pockets designated A–F differs among individual alleles and this structural polymorphism is thought to contribute to the peptide binding specificity of different class I molecules. Deep, but highly conserved pockets at each end of the groove bind the amino (pocket A, residues 7, 59, 159, 171) and carboxyl termini (pocket F, residues 84, 143, 146, 147) of the peptides, thus dictating the orientation of peptide binding. The comparative positions in salmonid class I sequences (positions 7, 57, 82, 140, 143, 144, 157, 169; *Sasa-UBA*0101* numbering) are identical and invariable with the exception of the human Y84 being replaced by an arginine in the salmonid sequences. Apart from pockets A and F, each of the remaining pockets vary in both size and depth between different human and mice class I molecules. Pocket B, mostly determined by the size and chemical nature of residues at positions 7, 9, 24, 34, 45, 63, 66, 67, 70 and 99, is a major pocket in many class I molecules and accommodates the anchor residues of the second amino acid (P2) of the peptide.

Information about general sequence variation in *Mhc* sequences was extracted from the Kabat Database of Sequences of Proteins of Immunological Interest

(Johnson and Wu 2001) by using the Variability web interface. Several variability plots were generated, both for the complete database (1,042 sequences) and for subsets selected by species [e.g. human (394 sequences), mouse (177 sequences) or Atlantic cod (28 sequences)]. In each plot, positions showing a variability clearly above a loosely defined "noise level" were identified, and compared to the variable positions indicated in Fig. 4c. In particular, positions 63, 66, 94, 111 and 153 were always identified as highly variable, and to some extent positions 42 and 60. These positions correspond approximately to binding pockets B and E, and this variation is most likely essential for achieving a broad repertoire in antigen binding. Most of the additional variability observed in Fig. 4c may be associated with other known binding pockets, in particular pockets A and C. One possible exception is the high variability observed in positions 50–55. This region is generally not used for antigen binding, and high variability in this region seems to be unique to this data set, compared with sequences available in the Kabat database. It is known that *Mhc* class I molecules can bind peptides that are too long to fit into the actual binding pocket, and that such peptides may extend out of the pocket (Collins et al. 1994). Additional variation close to binding pocket A may modulate binding of such antigens, through direct interaction with the antigen or by modified flexibility of the helices defining the binding cleft. It is also known that the natural killer cell receptor may bind to *Mhc* class I in this region (Tormo et al. 1999; PDB code 1qo3). Natural killer cells are capable of lysing virally infected cells and are in part regulated through interactions with the *Mhc* class I molecule. Knowledge about cytotoxic cells in fish is limited, but channel catfish has been shown to possess an array of different cytotoxic cells, some potentially resembling the human NK cells (Shen et al. 2002). A rapid screening of the zebrafish and pufferfish genome databases also show multiple sequences with similarity to receptor molecules traditionally used to discern NK cells. An automatic screening of the 3D structure (1qo3) for protein-protein interactions, using the Protein-Protein Interaction Server (Jones and Thornton 1996), identified residues 47 and 50–56 as involved in receptor binding in this region. It is therefore possible that the observed variation is associated with modulation of protein-protein interactions taking place in this area.

In conclusion, we have identified highly divergent alleles belonging to the major *Mhc* class I locus of Atlantic salmon, *Sasa-UBA*. Shuffling of α_1 domains between alleles was identified and the microsatellite marker residing within the 3′-untranslated tail seems to participate in recombination events. Modelling the antigen-binding groove identifies both structure and charge differences between alleles, suggesting major differences in peptide repertoire. A polymorphic region unique to salmonid sequences was identified, potentially representing an NK-receptor interaction site.

Acknowledgements This study was supported by grants from the Norwegian Research Council and an European Commission FAIR CT97-3643 contract. The material presented in this paper is available upon request. The experiments performed in this study comply with the current laws of Norway and The Netherlands.

References

Anthonsen HW, Baptista A, Drablos F, Martel P, Petersen SB (1994) The blind watchmaker and rational protein engineering. J Biotechnol 36:185–220

Aoyagi K, Dijkstra JM, Xia C, Denda I, Ototake M, Hashimoto K, Nakanishi T (2002) Classical MHC class I genes composed of highly divergent sequence lineages share a single locus in rainbow trout (*Oncorhynchus mykiss*). J Immunol 168:260–273

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242

Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, Gates M, Postlethwait JH, Klein J (1997) Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. Immunogenetics 46:129–134

Carreno BM, Schreiber KL, McKean DJ, Stroynowski I, Hansen TH (1995) Aglycosylated and phosphatidylinositol-anchored MHC class I molecules are associated with calnexin. Evidence implicating the class I-connecting peptide segment in calnexin association. J Immunol 154:5173–5180

Carrington M (1999) Recombination within the human MHC. Immunol Rev 167:245–256

Collins EJ, Garboczi DN, Wiley DC (1994) Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. Nature 371:626–629

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–95

Devlin RH (1993) Sequence of sockeye salmon type 1 and 2 growth hormone genes and the relationship of rainbow trout with Atlantic salmon and Pacific salmon. Can J Fish Aquatic Sci. 50:1738-1741

Flajnik MF, Ohta Y, Greenberg AS, Salter-Cid L, Carrizosa A, Du PL, Kasahara M (1999) Two ancient allelic lineages at the single classical class I locus in the *Xenopus* MHC. J Immunol 163:3826-3833

Graser R, Vincek V, Takami K, Klein J (1998) Analysis of zebrafish *Mhc* using BAC clones. Immunogenetics 47:318–325

Grimholt U (1997) Transport-associated proteins in Atlantic salmon (*Salmo salar*). Immunogenetics 46:213–221

Grimholt U, Hordvik I, Fosse VM, Olsaker I, Endresen C, Lie O (1993) Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). Immunogenetics 37:469–473

Grimholt U, Getahun A, Hermsen T, Stet RJ (2000) The major histocompatibility class II alpha chain in salmonid fishes. Dev Comp Immunol 24:751–763

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723

Hansen JD, Strassburger P, Thorgaard GH, Young WP, Du PL (1999) Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. J Immunol 163:774–786

Hordvik I, Grimholt U, Fosse VM, Lie O, Endresen C (1993) Cloning and sequence analysis of cDNAs encoding the MHC class II beta chain in Atlantic salmon (*Salmo salar*). Immunogenetics 37:437–441

Johnson G, Wu TT (2001) Kabat Database and its applications: future directions. Nucleic Acids Res 29:205–206

Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proc Natl Acad Sci USA 93:13–20

Kaufman J, Salomonsen J, Flajnik M (1994) Evolutionary conservation of MHC class I and class II molecules – different yet the same. Semin Immunol 6:411–424

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923-925

Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 24:946–950

Madden DR, Gorga JC, Strominger JL, Wiley DC (1991) The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. Nature 353:321–325

Matsumura M, Fremont DH, Peterson PA, Wilson IA (1992) Emerging principles for the recognition of peptide antigens by MHC class I molecules. Science 257:927–934

McKay SJ, Devlin RH, Smith MJ (1996) Phylogeny of Pacific salmon and trout based on growth hormone type-2 and mitocondrial NADH dehydrogenase subunit 3 DNA sequences. Can J Fish Aquatic Sci 53:1165-1168

McVeigh HP, Barlett SE, Davidson WS (1991) Polymerase chain reaction/direct sequence analysis of the cytochrome b gene in *Salmo salar*. Aquaculture 95:225–233

Merritt EA, Bacon DJ (1997) Raster3D: photorealistic molecular graphics. Methods Enzymol 277:505–524

Miller KM, Withler RE (1998) The salmonid class I MHC: limited diversity in a primitive teleost. Immunol Rev 166:279–293

Nicholls A, Honig B (1991) A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J Comput Chem 12:435–445

Nicholls A, Sharp KA, Honig B (1991) Protein folding and associations: insight from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11:281–296

Page RDM (1996) An application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Pardue ML, Lowenhaupt K, Rich A, Nordheim A (1987) (dC-dA) n.(dG-dT)n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. EMBO J 6:1781–1789

Parham P, Lomen CE, Lawlor DA, Ways JP, Holmes N, Coppin HL, Salter RD, Wan AM, Ennis PD (1988) Nature of polymorphism in HLA-A, -B, and -C molecules. Proc Natl Acad Sci USA 85:4005–4009

Parham P, Lawlor DA, Lomen CE, Ennis PD (1989) Diversity and diversification of HLA-A, B, C alleles. J Immunol 142: 3937–3950

Parham P, Adams EJ, Arnett KL (1995) The origins of HLA A, B, C polymorphism. Immunol Rev 143:139–180

Peitsch MC (1995) Protein modeling by E-mail. Biotechnology 13:658–660

Peitsch MC (1996) ProMod and Swiss-model: internet-based tools for automated comparative protein modelling. Biochem Soc Trans 24:274–279

Qian SB, Chen SS (2000) Blocked transport of soluble K(b) molecules containing connecting peptide segment involved in calnexin association. Int Immunol 12:1409–1416

Saitou N, Nei M (1987) The Neighbor-Joining method: a new method for reconstucting phylogenetic trees. Mol Biol Evol 4:406–525

Sakamoto T, Danzmann RG, Gharbi K, Howard P, Ozaki A, Khoo SK, Woram RA, Okamoto N, Ferguson MM, Holm LE, Guyomard R, Hoyheim B (2000) A microsatellite linkage map of rainbow trout *(Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. Genetics 155:1331–1345

Sander C, Schneider R (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 9:56–68

Saper MA, Bjorkman PJ, Wiley DE (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 ångstrøm resolution. J Mol Biol 219:277–319

Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF (1999) The GROMOS biomolecular simulation program package. J Phys Chem 103:3596–3607

Shen L, Stuge TB, Zhou H, Khayat M, Barker KS, Quiniou SM, Wilson M, Bengten E, Chinchar VG, Clem LW, Miller NW (2002) Channel catfish cytotoxic cells: a mini-review. Dev Comp Immunol 26:141–149

Shum BP, Guethlein L, Flodin LR, Adkison MA, Hedrick RP, Nehring RB, Stet RJ, Secombes C, Parham P (2001) Modes of salmonid MHC class I and II evolution differ from the primate paradigm. J Immunol 166:3297–3308

Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M (1999) Automated analysis of interatomic contacts in proteins. Bioinformatics 15:327–332

Stam P, Van Ooijen JW (1995) JoinMap version 2.0: software for the calculation of genetic maps. cPRO-DLO, Wageningen

Takami K, Zaleska-Rutczynska Z, Figueroa F, Klein J (1997) Linkage of LMP, TAP, and RING3 with Mhc class I rather than class II genes in the zebrafish. J Immunol 159:6052–6060

Thomson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tormo J, Natarajan K, Margulies DH, Mariuzza RA (1999) Crystal structure of a lectin-like natural killer cell receptor bound to its MHC class I ligand. Nature 402:623–361

Van Erp SHM, Dixon B, Figueroa F, Egberts E, Stet RJM (1996) Identification of a novel major histocompatibility complex class I gene from carp (*Cyprinus carpio*). Immunogenetics 44:49–61

Yoshino M, Sagai T, Lindahl KF, Toyoda Y, Moriwaki K, Shiroishi T (1995) Allele-dependent recombination frequency: homology requirement in meiotic recombination at the hot spot in the mouse major histocompatibility complex. Genomics 27:298–305