Innate immunity / Immunité innée

Innate immunity in early chordates and the appearance of adaptive immunity

Louis Du Pasquier

Institute of Zoology, University of Basel, Rheinsprung 9, CH-4051 Basel, Switzerland

Abstract

In the urochordate *Ciona intestinalis* some membrane Immunoglobulin superfamily members with ancestral features of antigen receptors are homologs of vertebrate adhesion molecules acting as virus receptors. They include the following: the junction adhesion molecule (reovirus receptor) (JAM), the Cortical thymocyte marker of *Xenopus* (CTX family) (Coxsackie's virus receptor) and the poliovirus receptor (PVR). In humans these genes belong to the same linkage group, of which 4 paralogous groups exist. This situation is consistent with the notion that the *Ciona* set of genes would correspond to a preduplication state. In addition, the human region 3q13 and its paralogs, harbour genes remotely related to the nectin family that can be detected in Protostomes (human CRTAM and CD80-86 related to *Drosophila* Beat). In addition, this linkage group contains several CDs important for the immune system CD166, CD47 and many members of the tetraspanin family. The VC1-like core of the nectin is homologous to the VC1 core of the MHC-linked tapasin and to the VC1 segments of, for example, specific antigen receptors of vertebrates, and could be related to a primitive antigen receptor gene. It is suggested that the virus binding property of the members of this family was exploited, and that they were recruited in the vertebrate immune system following the introduction of the somatic rearrangement machinery. In this way the adaptive immune system could have developed from a set of receptors involved in a primitive local innate immunity involving NF-κB-mediated apoptosis. *To cite this article: L. Du Pasquier, C. R. Biologies 327 (2004).*
© 2004 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Résumé

L'immunité innée chez les chordés primitifs et l'apparition de l'immunité adaptative. Chez l'Urochordé *Ciona intestinalis*, trois membres de la superfamille des immunoglobulines, ressemblant au récepteur des cellules T ancestral, ont été trouvées au sein de deux familles de molécules d'adhésion. Ces molécules possèdent un domaine de type variable codé par un gène V complet sans réarrangement somatique, en position distale associé au gène d'un domaine constant, soit de type C2 (famille CTX et JAM), le plus ancien, soit de type C1, typique des molécules du système immunitaire adaptatif (Famille nectine). Chez les vertébrés, les homologues de ces trois gènes sont tous groupés sur un seul segment de chromosome et ses paralogues (3q13, 11q23, 21q21, 1q22). Chez l'homme, ces trois types de molécules d'adhésion peuvent servir de récepteur pour des virus (reovirus, Coxsackie's virus et poliovirus) et leur interaction avec le ligand peut entraîner l'apoptose de la cellule par une voie

E-mail address: dupasquier@dial.eunet.ch (L. Du Pasquier).

NF-κB. Il est possible qu'au cours de l'évolution des chordés une telle propriété ait favorisé le recrutement de cette famille de molécules dans le système immunitaire. Les gènes V de ces molécules constituaient une cible idéale pour l'introduction du réarrangement somatique. Sur le même groupe de liaison, i.e., le chromosome 3q13 de l'homme, se trouve aussi CD166 et des tétraspanines, dont on retrouve des homologues chez *Ciona*. Sur le même groupe de liaison et ses paralogues, on trouve encore les gènes de molécules apparentées, CRTAM, CD80 et 86, qui ont des homologues dans la famille Beat de la drosophile, ce qui permettrait de retrouver la trace des récepteurs spécifiques du système immunitaire des vertébrés chez les protostomes. ***Pour citer cet article : L. Du Pasquier, C. R. Biologies 327 (2004).***

1. Introduction

During evolution many defense strategies have been invented to fight pathogens and parasites, whether these parasites belong to a different phylum or kingdom (bacteria, virus, protozoa etc.) or to the same species (cases of allorecognition in many invertebrates) as the host. A great variety of germline-encoded innate defense mechanisms, characterized by their rapid intervention, are found across the animal kingdom. In contrast, adaptive immunity, grounded upon immunoglobulins (Ig) and T cell receptors (TCR) with the variable part of the receptor generated somatically by random recombination of elements scattered in the genome, and the major histocompatibility complex (MHC)-encoded antigen presenting molecules are present only in the jawed vertebrates. Because of clonal selection, positive and negative selection in the thymus, MHC-regulated initiation of all adaptive responses, etc., the essential elements of the adaptive immune system are locked in a coevolving Ig-TCR-MHC unit. This complex system was in fact superimposed onto innate systems inherited from invertebrates, of which some elements are conserved, such as the complement system and others, such as the Toll-like receptor cascade, are still required for the initial phases of the adaptive response [1]. Why and how did this adaptive system arise is a fascinating evolutionary question, the speculative answer of which may be revealed as we investigate the genomes of non-vertebrate chordates. This article deals with the origin of the rearranging receptor genes.

2. Potential ancestors of antigen specific receptors in the genome of an early chordate *Ciona intestinalis*

How was the somatic generation of diversity introduced in vertebrates? One needed the introduction of a rearranging mechanism in an Immunoglobulin superfamily (Igsf) V gene without jeopardizing the life of the individual. So presumably the introduction of RAG mediated rearrangement came after the expression of the ancestral receptor was confined to a cell such as the modern effector of adaptive immunity, i.e., a primitive lymphocyte. If things went wrong, it would only affect one clone of cells and not the organism.

For the enzyme machinery, the apparent abrupt appearance of RAG1 and 2 in elasmobranchs suggests that a transposon and its transposase, presumably from a unicellular organism, were introduced in the primordial V genes of the receptor [2].

An easier task perhaps would be to find out where the MHC, T-cell receptor and antibodies structural elements come from and in which order. The idea is to examine non-vertebrates chordates for such molecular characteristics. Now the genome of the urochordate *Ciona intestinalis* is available and no 'real' genes of Ig TCR in the modern sense have been found [3]. It has the reputation of not having adaptive immune responses and to have natural killers [4] and some complement components as parts of its innate immune system have been reported [5]. It is an ideal deuterostome order to search for genes related to the ancestors of the modern effectors of adaptive immunity.

First one has to predict what could have been the immediate ancestor of an Ig or a TCR polypeptide

chain. The simplest precursor would have one complete V domain and one C domain, similar to a light chain or to a TCR chain. The C domain should be related the very specific C1 type, the Igsf domain type, seen only in vertebrates that achieve somatic rearrangement [6]. C1 domains are encountered in a very small number of molecules compared to the other Igsf I-set and C2 domains. In fact, other than TCR and Ig, C1 domains are found in the MHC class I and II molecules, the MHC-linked tapasin, perhaps in the MHC-linked butyrophilin, and in a small family of signal inhibitory proteins called the SIRPs [7]. However, we shall see that there are other putative candidates, and that they are precisely the ones that have homologues in the *Ciona* genome.

2.1. Immunoglobulin superfamily members structurally related to the vertebrate immune system

Several Igsf molecules related to the adaptive immune system were found in typical Blast searches. In addition to TCR and Ig V regions, the V domains of CTX (cortical thymocyte receptor of *Xenopus*) and Tapasin (a molecule involved in the class I peptide loading pathway) were used as *in silico* probes. These were known by experience to have yielded many non-rearranging V genes homologs in vertebrates. The sequences of the genes mentioned below can be retrieved from the *Ciona* database with their id numbers given after each. However some of the genes mentioned were not found under this form in the JGI database, where the reconstruction is not always accurate (see below).

2.1.1. Igsf members with V domains poorly related to the antigen specific receptor

Homology was determined with the Blast program. The type of Ig domain often can be predicted from the primary sequence by a variety of programs. Five were used: Psired, Predict, j pred, gor 1 and gor 4. All are available on the Expasy server (http://www.expasy.ch/). If none of these precisely identify the V or C nature of a segment, all of them detect differences between domains and permit grouping. This characteristic, in addition to the detection of specific residues [8] allows a reasonable estimate of the type of domain encountered.

One Homolog of CD166 (for a complete set of reference about the CDs mentioned in this paper please consult (http://www.ncbi.nlm.nih.gov/PROW/) (ci0100151841) V-V-C2-C2-C2-TM-Cy has been selected because it is expressed in hematopoietic tissues and has 2 V domains. In humans it belongs to a linkage group that may have phylogenetic relevance (see below).

Several homologs of Amphioxus VCBP V-V-chitin binding domain [9] grail 41-64.1, 41 65.1, 41 66.1 (ci0100131230). Grail 324.11.1 (ci0100148111) and Grail 7418.1 (chitinase domain only) are made of 2 V domains that do not rearrange somatically, followed by one chitin-binding domain (Fig. 1). In *Ciona* those VCBPs form a smaller multigene family than in *Branchiostoma*, the more evolved cephalochordate. In *Ciona* the V domains are assembled each from two exons spliced by a type 0 joint like for the vertebrate CTX V domain. In the case of the best characterized gene 324.11.1 (still incomplete at the 5'end) the splicing side in the V2 domain is at the same position as in CTX, i.e., 28 amino acids after the first cystein, even though there is low sequence identity with CTX. Their expression pattern in the gut of the amphioxus would fit with a role in innate immunity in that species, based on studies done in other species. The chitinase domain shows some conservation with a human Chitotriosidase secreted by activated human macrophages and involved in degrading chitin-rich pathogens [10]. A similar function could exist in Amphioxus and *Ciona*. The architecture of VCBP molecules remind us of that seen in FREPs (V-V-Fibrinogen) of the molluscan *Biomphalaria*, where 2 Ig domains, belonging to a diverse multigene family, are connected to a fibrinogen domain and play a role in the defense against the parasite *Schistosoma* [11,12].

None of these molecules is close to the ancestor of TCR single chain VC-tm-cy proposed architecture, not only because they lack C1 domains, but also because they are apparently soluble molecules, and not likely to function as receptors. Therefore using CTX, an Igsf receptor with a non rearranging V, as an *in silico* probe I searched the *Ciona* genome and the *Ciona* EST data bases and uncovered two other categories of putative Igsf membrane receptors.
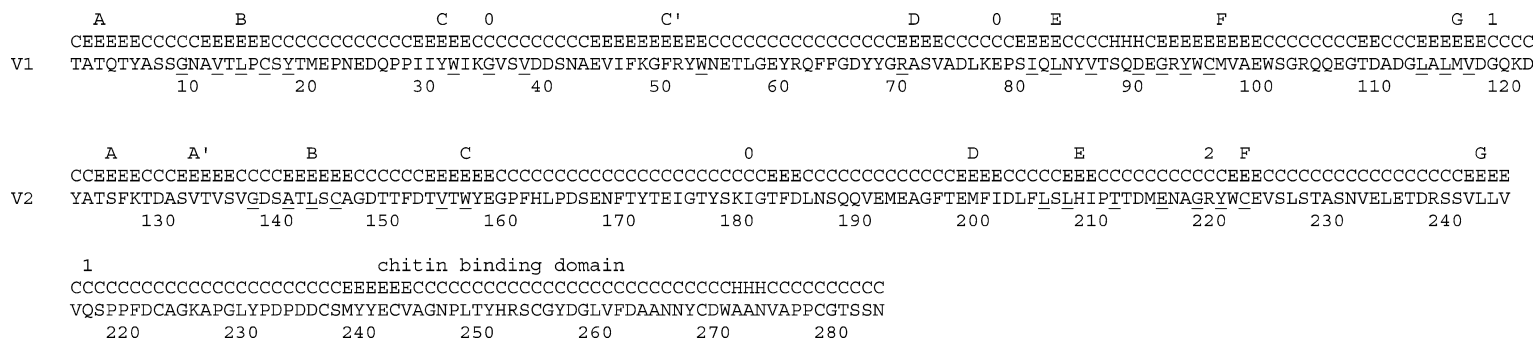
```
                A               B                 C       0              C'                      D       0     E               F                  G    1
                CEEEEECCCCCEEEEEECCCCCCCCCCCCCCEEEEECCCCCCCCCCEEEEEEEEEECCCCCCCCCCCCCCCCCCCCEEEECCCCCCEEEECCCCHHCEEEEEEEEEECCCCCCCCCEECCCCEEEEEECCCC
        V1      TATQTYASSGNAVTLPCSYTMEPNEDQPPIIYWIKGVSVDDSNAEVIFKGFRYWNETLGEYRQFFGDYYGRASVADLKEPSIQLNYVTSQDEGRYWCMVAEWSGRQQEGTDADGLALMVDGQKD
                         10        20        30        40        50        60        70        80        90       100       110       120


                A      A'       B              C                                     0               D        E           2  F                        G
                CCEEEEECCCEEEEEECCCCEEEEEEECCCCCEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCEEECCCCCCCCCCCCCCCCEEEECCCCCEEECCCCCCCCCCCCCCCEEECCCCCCCCCCCCCCCCCCCCEEEE
        V2      YATSFKTDASVTVSVGDSATLSCAGDTTFDTVTWYEGPFHLPDSENFTYTEIGTYSKIGTFDLNSQQVEMEAGFTEMFIDLFLSLHIPTTDMENAGRYWCEVSLSTASNVELETDRSSVLLV
                        130       140       150       160       170       180       190       200       210       220       230       240


                1                             chitin binding domain
                CCCCCCCCCCCCCCCCCCCCCCEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCC
                VQSPPFDCAGKAPGLYPDPDDCSMYYECVAGNPLTYHRSCGYDGLVFDAANNYCDWAANVAPPCGTSSN
                        220       230       240       250       260       270       280
```

Fig. 1. *Ciona* VCBP 324.11.1 – Reconstructed from grail 324. V1 V2 Immunoglobulin superfamily domain with V fold Amino acid sequence deduced from the genomic sequence and secondary structure prediction. The underlined residues represent the V frame fingerprint residues. The figures 0, 1, 2 correspond to the position of the splicing sites and the number indicates the splicing law. The ? indicates that the gene is not complete in the data base. For the secondary structure prediction C = coil, E = strand, H = helix, Phosp = phosphorylation site, ABCC′DEFG = Igsf strands.

```
                        1
                        CCHHHHHHHHHHHHHHHH
        leader          MNFNVLLCVVLAFVCCC


                A               A'          B           2   C              C'       C"      1  D          E                 F                  G
                CCCCCEEEEEEECCCEEEEEECCCCEEEEEEEECCCCCCCCCCEEEEEECCCCCCCEEEEEEEEEECCEECCCCCCCCCEEECCCCEEEECCCCCCCCCEEEEEEEEEECCCEEEEEECEEEEEEEE
        V       NGVRVTLLNGAASPLPIEVTLGNPITLACSYQPENGNAQDARLEWLYSTNMDGTSTSGQNIVLDGIIQPNFNNGRFSIVNPSSLQISESIKTDTGSYTCLVTLPNDTPSKGSGTYHLTVN
                         10        20        30        40        50        60        70        80        90       100       110       120


                1       A                   B  1       C                   C'          E           F             G
                CCCCCEEEECCCCCCCCCCCCEEEEEEEECCCCCCCEEEEEECCCCCCCCCCCCEEEEECEEEEECCCCEEEEECCCHHHCEEEEEEEEEECCCCEEEEEEEEEE
        C2      VPPSPPQCQYTSDHALTIGFAAEFRCLSVEGIPTPTYTWFKNGRQLPLSAQDDVTYKNASFSYSATTGVLKFSKVTTGDAGTYYCSSSNKIATESCLPTVITT
                         130       140       150       160       170       180       190       200       210       220


                    transmembrane  phosph                                               phosph
                CCCCCEEEEEEECCCCEEEEECEECCCCCCCEEECCCCCCCCCCCCCCCCCEEEECCCCEEEEEECCCEECCCCCCCCCCCCCCCEEECC
                KNQDVGMIVGIVFGVLFGLILIGVLVWWTWRKGYCDGLSKEDDPDEIEDDGSNDIMLDTNGPTVHKPPSEVSSTRKQEASMII
                         230       250       260       270       280       290       300       310
```
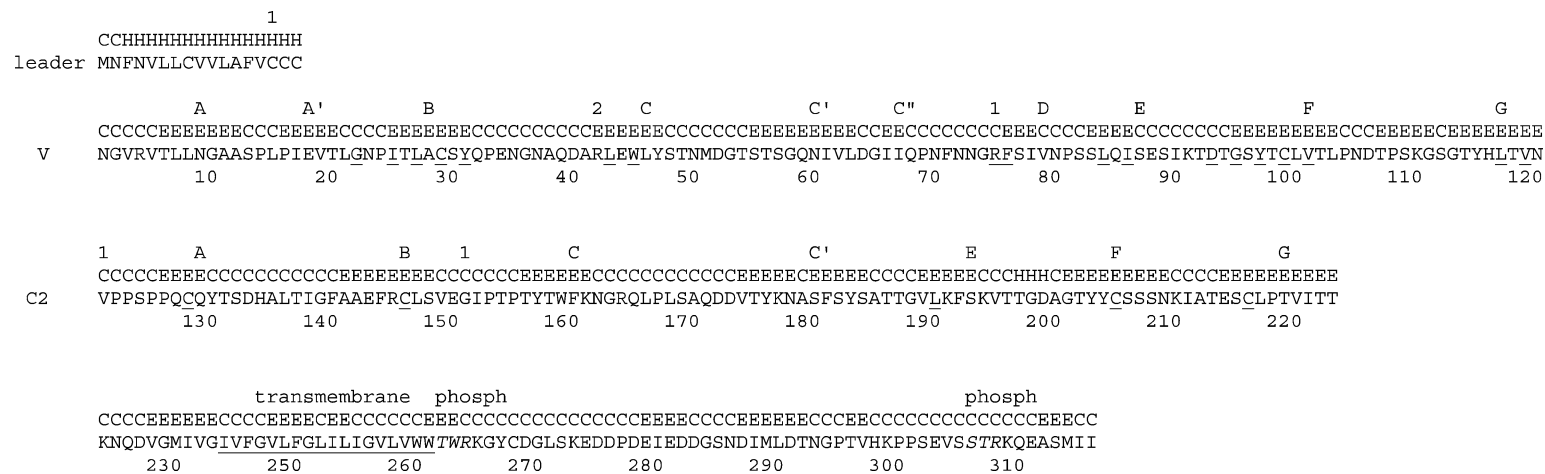
Fig. 2. A *Ciona* JAM CTX member ci0100131719. Amino acid sequence deduced from the genomic sequence and secondary structure prediction. The underlined residues represent the V frame fingerprint residues. The figures 0, 1, 2 correspond to the position of the splicing sites and the number indicates the splicing law. For the secondary structure prediction C = coil, E = strand, H = helix, ABCC′DEFG = Igsf strands, Phosp = phosphorylation site. The four Cysteins of the constant region, a shared characteristic with CTX family members have been underlined.

2.2. Igsf transmembrane receptors

2.2.1. A JAM CTX family member (V-C2-TM-Cy)

One gene (>ci0100131719) was reconstructed (Fig. 2) and completed with help of a partial CDNA (BW203636). It resembles a member of the JAM CTX family. With an Igsf V domain, a C2 domain, a transmembrane and cytoplasmic segments, it has the hallmarks of a cell surface receptor. The overall homology puts the *Ciona* gene close to JAM3 and JAM2, the two forms that unlike JAM1 and the related Z39Ig, share the extra pair of cysteins in the C2 domain with CTX (Figs. 2 and 4). The unique *Ciona* gene looks like apreduplicated form, ancestral to modern JAM and to CTX. The gene architecture so well conserved in vertebrate CTX and to a certain extent in the JAM genes (Fig. 4) is not maintained in the *Ciona* gene, as new introns were introduced in the urochordate gene and introns have disappeared in the vertebrate genes. Yet there are striking similarities. The last two exons of the C domain splice to eachother via a type 0 splice site in the same position as in the JAM3 gene (Fig. 4).

The V domain is made of 3 exons (Fig. 4) but they are not superimposable to those of JAM. As in JAM, the transmembrane segment is encoded by an exon that encodes only a few cytoplasmic residues; this is not the case in CTX.

2.2.2. Two nectin/poliovirus receptor (PVR) family members (V-C1 like-C2-TM-Cy)

The other family detected in *Ciona* having the architecture of a cell surface receptor was that of nectins and poliovirus receptors (PVR), molecules relatively ignored by immunologists, represented by two specimens in *Ciona* (Fig. 3). Those genes are made of one distal V domain followed by 2 C domains linked to a transmembrane and a cytoplasmic segments. The anterior C domain of the nectins is more like a C1 than a C2 domain, and the VC1 segment represents a core that has good homology with many molecules, or segments of molecules, of the adaptive immune system in the following ways: (*i*) the TCR and Ig basic architecture; and (*ii*) the core of the MHC-linked Tapasin and its paralogs and to a certain extent the MHC-linked butyrophilin. The two *Ciona* genes differ markedly at the level of the first constant domain. In gene 39 the C domain is more like a C1 domain than that in the gene 34.

In phylogenetic trees the two genes do not cluster together; thus they may be derived from an ancient duplication.

In summary, the only two membrane *Ciona* Igsf forms with a reasonable membrane receptor architecture correspond to two vertebrates families already suspected of having ancestral features [5]: JAM [13]/ CTX family [14–16] and the nectins [7].

3. Linkage studies

The above mentioned members found in *Ciona* belong to a single linkage group in humans, with its several paralogs. Moreover, members of the extended VC-like family belong to the MHC linkage group (tapasin) [17,18] and some of the human nectin CTX JAM chromosomal segments are on bona fide MHC paralog as well (CTX is next to MHC proper in *Xenopus*; JAM1 and the CTX family member A33 are on 1q, a MHC paralog in human; Tapasin is in the MHC and its paralog is on the MHC related group on 12 p13). In summary, the search in *Ciona* led not only to one or a few homologous genes but also to conserved genetic regions. In humans these genetic regions contain many immunologically relevant genes some of which are known CDs, including CD80 and 86 with a good homology with CTX V [14] and several tetraspanins on chromosome 3. Other genes such as the pentaspanin, CD47 with an Ig domain clustering with the V domain of the linkage group CD47 could not be found in *Ciona*. Interesting enough the V domain of CD47 is most similar to the members of that linkage group as if its V domain was related to the nectin, tapasin, JAM/CTX group. It looks like a possible duplicate of the family just like P0 and Eva are CTX related segments on their respective linkage groups [7]. In addition, CD166 mentioned above, because of its homolog in *Ciona*, is also in the same linkage group 3q13 (Fig. 5). Other genes in these regions are linked to innate immunity (Tollip, Mail).

Some of the non-Igsf genes can be found in the *Ciona* genome: for instance many tetraspanins and CD36. Some of the TM4 turned out to be on a *Ciona* segment (grail 149) that contains several human homologues that all map to the human region 11p 15, which we had not pointed out so far in this article. However, what looked like the pseudogene of a nectin-

```
                          CCHHHCCHHHHHHHHHHHHHHH
             leader       MILRRIFIPRILKAALLLGCFC
                          1    A       B              C    0     C'         C''       D      E              F              G
             39V          CCCCCCEEEEECCCEEEEEEEEECCCCCCCCCCEEEEEEEEEEECCCCCEEEEEECCCCCCCCCEEECCCCCCCEEEEECCCEEEECCCHHHCEEEEEEEEEEECCCCEEEEEEEEE
                          CYGVHGTNVIQGESAFIPCYNPETASQFPNMLVILEFQPTPTAPQQLVSHYQPAQLKDPVVFATGSRYRGRLQFMNTGQLEIKFISLPDEGTFTCTAQVYSLAPSKIITSVIVY
                                                                                                                                        116

                          1         A        B     1      C        2 C'       D      E              F        G
             39C1-L       CCCCCCCCCCCCCCEEEEECCCCCEEEEEEEECCCCCCCCCEEEEECCCCCCCCCCCCEEECCCCCEEEEEECCEEEEECCCCCCCCCEEEEEEEEECCCCCEEEEEEEEEE
                          VAPQFEVRKVNDPLLISDGGTMVNAAECIASNGKPPAVVFWGTNRHLDFDQTATNSSTTPLLTTAVSTLHVEPTKFYNKQSFYCEMRHPSIPNSKRETIYLNVT
                                                                                                                              239

                          1   A      B          C          C'     E              F 1      G
             39C2         EEEEEEEEEEECCCCCEEEEEECCCCCCEEEEEEECCCCCEEEECCCEEEEEEEECCHHHCEEEEEEEEEECCCCCEEEEEEEEEECC
             1            YPPETPTITVTADGLSLTCISHANPPATISWLLPASAPSRTFMGPTVSMMSPNTAERSNDTYTCTANNGIGVAVQSTITAGEASA
                                                                                                           325

                                         Transmembrane        phosph
                          CEEECCCCCCCCEEEEECCCCCEEEEEEEEEEEEEEEEECCCCCCCCCCCCCCCCCCCCCC
                          VIVVTKESNVGIMVGAIIGSLVVLAIIIALIYFFIIKPARKDTGPYDQTIQPFKADR
                                                                         364




                                       1
             leader       CCCHHHHHHHHHHHHHHHHCCCCCCEE
                          MKLILIQSFLYLCFSCYSTYGEGVV
                                      A       B          C        C'    C''    2D       E              F          G     1
                          EEECCCCCEEECCCCCCEEEEEEEEECCCCCCCCCEEEEEEEECCEEEEEECCCCEECCCCCCEEEECCCCCCCCEEEEEECCCHHHCEEEEEEEEEECCCCCEEEEEEEEEECC
             34V          LLSGSTVNGVVTPAPTKTVTVYCDIVGQLPVDQSYHFTWSVEGEGLISYDDDVVSKGYQSRFIVANSPNEASRLALEINPVKLSDVGKYDCSIVGLYDHNVISSSFLEVNQL
                                                                                                                                           115

                                a?            B        C2           C'    D      E              F        G
                          CCCCCCCCCCCCCCCCCEEEEEEEEECCCCCCCEEEEECCCCCCCCCCCCEEEEECCCEEEEEECCEEEECCCCCCCCCCEEEEEEECCCCCCCEEEEEEEECC
             34C?          PNITVKVNDLTVVEGNEVILGTCYANASKPQPEIWIEDSAGHRVESLSSPSIVNHEEHAGVSDSVLKFSKVFNREENGESYRCMVQHPDNLETFSSYIGQVD
                                                                                                                         218

                          1      A      B      C                   C'     1 E              F        G
                          EECCEEEECCCCCEEEEECCCCEEEEEEEEEECCCCEEEEECCCCCCCCCCCCCCCCCCCCCEEEEEEEEEEECCCEEEEEECCCCCEEEEEEEEEECCCCCEEEEEEEEEE
             34C2         QFSPTVSIMGDESDISMEEGQSLSVVSLTTGNPTPVVVWTFQPDTTQPPATSADSTNPTDPPTYGLPANFQAAGNTLIGNPLSMIGNDGTFTCTVSNGLGSPASISFHLTIT
                                                                                                                                         329
                          1        1      1                   Transmembrane     phosph 1                                 phosph.
                          CCCEEECCCCCCCCCEEECCCCCCCCCCCCCCCCCCEEEECCEEEEEEHHHHHHHHHHHHHHHHCCCEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHCCC
             34C2         EKPTTTAAPTTPVATVGGHDSGEATKPLPRQAGADAAVIGGVVAVAVFVLIVAIIFLLRYFMTHKGEYYTNELKATDDARLENNPIDDEDDDLDGDEDQPLDERKRTEHFL
                                                                                                                                          440
```

Fig. 3. Nectins from *Ciona* from grail 34 and 39. Amino acid sequence deduced from the genomic sequence and secondary structure prediction. The underlined residues represent the V frame fingerprint residues. The figures 0, 1, 2 correspond to the position of the splicing sites and the number indicates the splicing law. For the secondary structure prediction C = coil, E = strand, H = helix, ABCC′DEFG = Igsf strands, Phosph = phosphorylation site. The nectin 34 has a second C domain that is not typical.
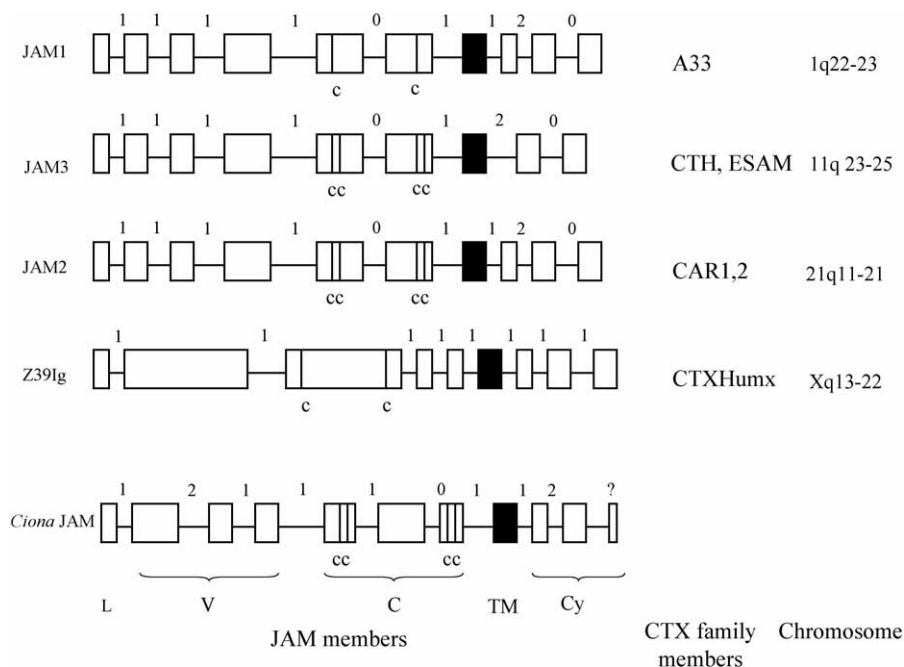
Fig. 4. The JAM/CTX family in *Ciona* and human. Exon intron organization and chromosomal localization of JAM and CTX in human. Data assembled from the ucsc server http://www.genome.ucsc.edu/ in April 2003. For the CTX architecture, see Chrétien et al. [15]. The cystein residues of the constant region are labeled c and a corresponding bar has been drawn in the C exons. The figures 0, 1, 2 indicate the type of splicing. The *Ciona* JAM is that from Fig. 2.

tapasin related gene, namely a single V-C1 segment for which no est has been found; its C1-like domain always lined up best with cold-blooded vertebrates class I and II genes located earlier exactly in this area i.e., between Tollip and RAG (!) [19].

4. Expression functions and properties of the JAM, CTX, and nectin molecules

The relevance of this finding to the origin of the lymphocyte antigen receptor is further suggested by the fact that some of these genes can be expressed on lymphocyte surface such as CTX, and CHT1, or on monocytes and other hematopoietic cells such as the nectin family members: CD111 and CD112 for two nectins and CD155 for the poliovirus receptor.

Another feature that could put these molecules on the evolutionary track toward TCR is the function that the nectin and JAM/CTX families assume in modern vertebrates. I mean, not necessarily, so much their function but their *properties* could be a telltale sign.

They are not ubiquitously expressed and seem to play an important role in the nervous system. These molecules function as adhesion molecules in tight junctions where for instance nectin and JAM interact with each other [12]. Indeed, there might be good reasons for keeping the JAM/CTX and nectin family together (adaptive clustering) on the chromosome. One wonders, of course, what they would do on single cells of the hematopoietic system where they have also been identified. There are no functional data on their roles in this lineage.

No commitment to immunity has been found except perhaps for CTX and CHT1 in non-mammalian vertebrates, where they are expressed at very specific stage of thymocyte development [14,20,21].

The link of these molecules with the adaptive immune system might be their propensity to act as virus receptors. Nectins and PVR are known to allow endocytosis of different viruses via an interaction with their V domain. Mutations affecting those interactions have been localized to the equivalent of framework 2 and CDR2. The Coxsackie's virus receptor is a
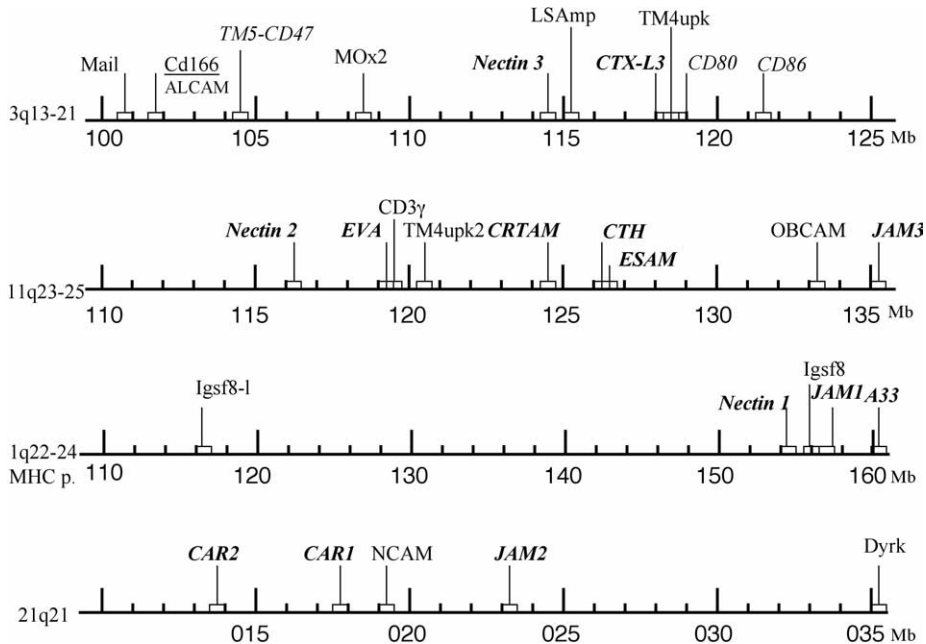
Fig. 5. Linkage of Nectin-JAM-CTX on four human chromosomal regions. Data compiled from ucsc http://www.genome.ucsc.edu/ in March 2003. The precise positions may not fit with newer releases but the relative positions of the genes are respected. The scale is in megabases. The CTX-L3 sequence was not originally recognized as a CTX family member despite its gene architecture. Accession number Igsf11, NM_152538. CTXhumx was identified earlier [7], accession number BC043216. MHCp indicates that this region is paralogous to the MHC region on chromosome 6p21 (review in [34]). There are other PVR-L, PVR and nectin or nectin-like genes on chromosome 19q13 and 3p12.

homologue of CTX. But this feature is not precisely an immune property. However, a paper by Barton in Cell [22] a few months ago showed that JAM was also a virus receptor and that the interaction of the virus with the V domain of the receptor could lead, whether the binding involve or not the sugar moiety, to apoptosis via the NF-κB pathway. Apoptosis can be considered as a local, presystemic, innate response. It was the sacrifice of one cell before the virus can replicate. As further proof that apoptosis is detrimental to the virus and that it could be therefore considered as a defense mechanism, the virus often inhibits these apoptotic pathways [23]. Provided that viruses existed at the time of the genesis of the adaptive immune system I would therefore not be surprised if that property was co-opted early in the build up of the adaptive immune system as already proposed [24]. It provides a link with the innate system with the conserved signaling cascades and with the structural requirements. From an adhesion molecule recognized by a virus, the introduction of somatic rearrangement might have

made a diversifiable set of receptors recognizing the virus.

The question of dimerization is not resolved for the immediate ancestor of the specific receptor, nor is the mode of binding resolved. Still, JAM can form dimers [25] as well as CTX [26], and as well as CARs and nectins [27]. If this molecule resembles a TCR ancestor does it recognize a peptide-MHC-like ligand as $\alpha\beta$ TCR does, or does it recognize a native epitope as $\gamma\delta$ TCR does? The homologies of CTX V and of related members place the V domain in the line to TCR V delta [12]. It is therefore conceivable that the prochordate molecule related to the ancestor of a primitive TCR molecule used a mode of recognition resembling that of antibody and that persisted in the modern $\gamma\delta$ receptors.

Was the *Ciona* or rather the *Ciona*-like genome of the true ancestor ready to accept Rag in one or the other of the above mention related genes? Has *Ciona* some of the other elements like MHC and 'pre lymphocytic' – coelomocytes that could indicate what was there 400 million years ago in its related

species that finally made the somatic rearrangement. A recent survey of the *Ciona* genome shows regions with conserved synteny of the class III MHC region but no inducible proteasomes [28]. This matches well the complementary finding on the cephalochordate Amphioxux by Abi Rached and Pontarotti [29]. For the lymphocyte arsenal, quite a few markers have been identified in *Ciona* [28]. *Ciona* with its complement system and its arsenal of innate immunity cascades was well defended and we can imagine the related genome of our ancestor was somehow similar, with its VC1 domain, coelomocytes and proto-MHC region not yet invaded by RAG1 and 2 and in its absence of course, choosing other solutions such as VCBP multiplicity and the complement components, the Toll-like receptors and probably other mechanisms. Some of those such as VCBP would be abandoned in later species using somatic rearrangement.

5. What about earlier Protostomes?

So far we have spent time trying to put ourselves in the tunique of a sea squirt looking forward in the most obnoxious finalistic attitude towards the adaptive immune system. What about the past and the origin of the VC1 lineage?

In other words, when looking further back in evolution, can one trace this VC1 lineage in protostomes?

Nectins V-C-C-TM-CY as such are not found in *Drosophila* but a molecule carrying a V-C-hydrophilic segment-TM-Cy, where the V-C part shows homology to nectin (that could be the equivalent of this supposedly wandering VC1 core) can be found in *Drosophila* Beat genes [30] (V-C?-hydrophilic insertion-TM-Cy). Indeed the databases present the first domain as a V fold and the C domain lines up reasonably well with nectin first C domain (database annotation takes the trouble of pronouncing it a nectin-like fragment). This Beat is part of a large family of at least 14 V carrying domain known for along time (another example of expansion of V families in invertebrates [31]) and with membrane forms more recently identified. It plays a role in the guidance of neurons towards their target.

Beat as such does not seem to exist in *Ciona* but a homologue with the same solvent exposed insertion exists in Vertebrates (birds and mammals) that was also selected it as being related to ancestral

TCR gene because of grouping in some phylogenetic trees [7]. Its name is CRTAM in human for 'class-I MHC-restricted T cell associated molecule' [32] and CTADS in *Gallus* for 'chicken thymocyte activation and developmental protein' [33]. It is a T cell-specific molecule in those two vertebrate species, and perhaps no surprise, it belongs to the same human linkage group on chromosome 11q 23-25.

6. Conclusion

I would like to propose the hypothesis that a unique set of Igsf genes now belonging to a single vertebrate linkage group (human 3q13 and its paralogs) has been instrumental for the genesis of the adaptive immune system specific receptors. Since round of duplications are likely to have occurred between the urochordate stage and the early vertebrate stage [34] and since some duplication of these genes could occur by other mechanism (there are already 2 nectins in Ciona) it is conceivable that one of the duplicates could have served as a target for experimenting somatic rearrangement. The homology of the V of these families with TCR V delta would fit with the idea that this primitive receptor recognized native antigen and not peptide in a class I or class II molecule. The trimming toward a V-C-TM-Cy architecture is not seen in this early chordate. The sequestration of expression to lymphocyte-like cells has not been investigated yet. Perhaps work in the amphioxus and in jawless fish will lead to discovery of molecules intermediary between the one described here and the true antigen specific receptor.

References

[1] M. Flajnik, K. Miller, L. Du Pasquier, Evolution of the Immune System, in Fundamental immunology, in: W. Paul (Ed.), Lippincott Williams and Wilkins, Philadelphia, 2003, pp. 519–570.

[2] J.D. Hansen, J.F. McBlane, Recombination-activating genes, transposition, and the lymphoid-specific combinatorial im-

mune system: a common evolutionary connection, Curr. Top. Microbiol. Immunol. 248 (2000) 111–135.

[3] P. Dehal, Y. Satou, R.K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D.M. Goodstein, N. Harafuji, K.E. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I.A. Meinertzhagen, S. Necula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H.G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chin-Bow, R. DeSantis, S. Doyle, P. Francino, D.N. Keys, S. Haga, H. Hayashi, K.W. Habe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, L. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M.M. Suzuki, O. Tassy, N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P.D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y. Kohara, M. Levine, N. Satoh, D.S. Rokhsar, The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins, Science 298 (2002) 2157–2167.

[4] C.M. Peddie, V.J. Smith, In vitro spontaneous cytotoxic activity against mammalian target cells by the hemocytes of the solitary ascidian, *Ciona intestinalis*, J. Exp. Zool. 267 (1993) 616–623.

[5] R. Marino, Y. Kimura, R. De Santis, J.D. Lambris, M.R. Pinto, Complement in urochordates: cloning and characterization of two C3-like genes in the ascidian *Ciona intestinalis*, Immunogenetics 53 (2002) 1055–1064.

[6] L. Du Pasquier, I. Chrétien, CTX, a new lymphocyte receptor in *Xenopus* and the early evolution of Ig domains, Res. Immunol. 147 (1996) 218–226.

[7] L. Du Pasquier, The phylogenetic origin of antigen-specific receptors, Curr. Top. Microbiol. Immunol. 248 (2000) 160–185.

[8] Y. Harpaz, C. Chothia, Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains, J. Mol. Biol. 238 (1994) 528–539.

[9] J.P. Cannon, R.N. Haire, G.W. Litman, Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate, Nat. Immunol. 3 (2002) 1200–1207.

[10] R.G. Boot, G.H. Renkema, A. Strijland, A.J. van Zonneveld, J.M. Aerts, Cloning of a cDNA encoding chitotriosidase, a human chitinase produced by macrophages, J. Biol. Chem. 270 (1995) 26252–26256.

[11] C.M. Adema, L.A. Hertel, R.D. Miller, E.S. Loker, A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection, Proc. Natl Acad. Sci. USA 94 (1997) 8691–8696.

[12] A. Fukuhara, K. Irie, H. Nakanishi, K. Takekuni, T. Kawakatsu, W. Ikeda, A. Yamada, T. Katata, T. Honda, T. Sato, K. Shimizu, H. Ozaki, H. Horiuchi, T. Kita, Y. Takai, Involvement of nectin in the localization of junctional adhesion molecule at tight junctions, Oncogene 21 (2002) 7642–7655.

[13] I. Martin-Padura, S. Lostaglio, M. Schneemann, L. Williams, M. Romano, P. Fruscella, C. Panzeri, A. Stoppacciaro, L. Ruco, A. Villa, D. Simmons, E. Dejana, Junctional adhesion molecule, a novel member of the immunoglobulin superfamily that distributes at intercellular junctions and modulates monocyte transmigration, J. Cell. Biol. 142 (1998) 117–127.

[14] I. Chrétien, J. Robert, A. Marcuz, J.A. Garcia-Sanz, M. Courtet, L. Pasquier, CTX, a novel molecule specifically expressed on the surface of cortical thymocytes in *Xenopus*, Eur. J. Immunol. 26 (1996) 780–791.

[15] I. Chrétien, A. Marcuz, M. Courtet, K. Katevuo, O. Vainio, J.K. Heath, S.J. White, L. Du Pasquier, CTX, a *Xenopus* thymocyte receptor, defines a molecular family conserved throughout vertebrates, Eur. J. Immunol. 28 (1998) 4094–4104.

[16] F. Eberle, P. Dubreuil, M.G. Mattei, E. Devilard, M. Lopez, The human PRR2 gene, related to the human poliovirus receptor gene (PVR), is the true homolog of the murine MPH gene, Gene 159 (1995) 267–272.

[17] J.A. Herberg, J. Sgouros, T. Jones, J. Copeman, S.J. Humphray, D. Sheer, P. Cresswell, S. Beck, J. Trowsdale, Genomic analysis of the Tapasin gene, located close to the TAP loci in the MHC, Eur. J. Immunol. 28 (1998) 459–467.

[18] M.S. Teng, R. Stephens, L. Du Pasquier, T. Freeman, J.A. Lindquist, J. Trowsdale, A human TAPBP (TAPASIN)-related gene, TAPBP-R, Eur. J. Immunol. 32 (2002) 1059–1068.

[19] L. Du Pasquier, Relationships among the genes encoding MHC molecules and the specific antigen receptors, in: M. Kasahara (Ed.), Major histocompatibility complex, evolution, structure and function, Springer, Tokyo, 2000, pp. 53–65.

[20] F.-K. Kong, C.-L.H. Chen, M.D. Cooper, Thymic function can be accurately monitored by the levels of recent T cell emigrants in the circulation, Immunity (1997), in press.

[21] K. Katevuo, B.A. Imhof, R. Boyd, A. Chidgey, A. Bean, D. Dunon, T.W. Göbel, O. Vainio, ChT1, an Ig superfamily molecule required for T cell differentiation, J. Immunol. 162 (1999) 5685–5694.

[22] E.S. Barton, J.C. Forrest, J.L. Connolly, J.D. Chappell, Y. Liu, F.J. Schnell, A. Nusrat, C.A. Parkos, T.S. Dermody, Junction adhesion molecule is a receptor for reovirus, Cell 104 (2001) 441–451.

[23] E.A. Tolskaya, L.I. Romanova, M.S. Kolesnikova, T.A. Ivannikova, E.A. Smirnova, N.T. Raikhlin, V.I. Agol, Apoptosis-inducing and apoptosis-preventing functions of poliovirus, J. Virol. 69 (1995) 1181–1189.

[24] D.L. Vaux, G. Haecker, A. Strasser, An evolutionary perspective on apoptosis, Cell 76 (1994) 777–779.

[25] D. Kostrewa, M. Brockhaus, A. D'Arcy, G.E. Dale, P. Nelboeck, G. Schmid, F. Mueller, G. Bazzoni, E. Dejana, T. Bartfai, F.K. Winkler, M. Hennig, X-ray structure of junctional adhesion molecule: structural basis for homophilic adhesion via a novel dimerization, EMBO J. 20 (2001) 4391–4398.

[26] J. Robert, I. Chrétien, C. Guiet, L. Du Pasquier, Cross-linking CTX, a novel thymocyte-specific molecule, inhibits the growth of lymphoid tumor cells in *Xenopus*, Mol. Immunol. 34 (1997) 133–143.

[27] M.J. van Raaij, E. Chouin, H. van der Zandt, J.M. Bergelson, S. Cusack, Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 A resolution, Structure Fold. Des. 8 (2000) 1147–1155.

[28] K. Azumi, R. De Santis, A. De Tomaso, I. Rigoutsos, F. Yoshizaki, M.R. Pinto, R. Marino, K. Shida, M. Ikeda, M. Arai, Y. Inoue, T. Shimizu, N. Satoh, D.S. Rokhsar, D.L. Du Pasquier, M. Kasahara, M. Satake, M. Nonaka, Genomic Analysis of Immunity in a Basal Chordate and the Emergence of the Vertebrate Immune System: Waiting for Godot, Immunogenetics (in press).

[29] L. Abi Rached, M.F. McDermott, P. Pontarotti, The MHC big bang, Immunol. Rev. 167 (1999) 33–44.

[30] D. Fambrough, C.S. Goodman, The *Drosophila* beaten path gene encodes a novel secreted protein that regulates defasciculation at motor axon choice points, Cell 87 (1996) 1049–1058.

[31] G.C. Pipes, Q. Lin, S.E. Riley, C.S. Goodman, The Beat generation: a multigene family encoding IgSF proteins related to the Beat axon guidance molecule in *Drosophila*, Development 128 (2001) 4545–4552.

[32] J. Kennedy, A.P. Vicari, V. Saylor, S.M. Zurawski, N.G. Copeland, D.J. Gilbert, N.A. Jenkins, A. Zlotnik, A molecular analysis of NKT cells: identification of a class-I restricted T cell-associated molecule (CRTAM), J. Leukoc. Biol. 67 (2000) 725–734.

[33] D.M. Ruble, D.N. Foster, Molecular characterization of a new member of the immunoglobulin superfamily that potentially functions in T-cell activation and development, Immunogenetics 51 (2000) 347–357.

[34] M.F. Flajnik, M. Kasahara, Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system, Immunity 15 (2001) 351–362.