THE MOLECULAR GENETICS OF HLA

Wolfgang R. Mayr and Dieter W.M. Schwartz

Klinische Abteilung für Blutgruppenserologie, Universität Wien, Währinger Gürtel 18-20, A-1090 Wien, Austria

1. THE HLA SYSTEM

The HLA system (for review, see Tsuji et al, 1992) which represents the major histocompatibility complex (MHC) of man, encompasses approximately one thousandth of the human genome (3.8 centi-Morgans corresponding to 3800 kilobases, kb) and is localised on the short arm of chromosome 6 (band 6p21).
In the HLA region, three major classes of genes can be found:
1) the HLA class I and class II genes,
2) immune function related genes (C2, C4A, C4B, BF, TNFA and TNFB, transporter and proteasome genes, HSP70, ...) and
3) other genes apparently not related to immune functions (e.g. CYP21, valyl-tRNA synthetase, ...).
This article will consider only the HLA class I and II genes, as well as immune function related genes with the exception of the complement genes C2, C4A, C4B and BF.

The HLA class I (HLA-ABC) loci code for glycosilated polypeptides with a molecular weight of 44000 daltons (44 kDa). The HLA complex contains around 30 such loci; most of them, however, are pseudogenes which are not expressed. The loci HLA-A, HLA-B and HLA-C are class I loci as well as other loci defining alloantigens and possibly differentiation antigens on T lymphocytes. HLA-E, HLA-F and HLA-G are further class I loci with expressed products which might be polymorphic. The size of each class I locus represents around 10 kb, the whole class I region which is situated on the telomeric end of the MHC corresponds to 1800 kb. ß2-microglobulin (ß2m), a monomorphic product of chromosome 15 (band 15q21 - 15q22) with 12 kDa, is non-covalently bound to the gene products of the class I loci. The allodeterminants of class I molecules are carried on the two distal domains of the heavy chain; for the formation of these sites, the presence of ß2m is necessary as the union of both molecules deeply influences the quaternary structure of the heterodimer.

The HLA class II (HLA-D) loci code for A and B chains (glycosilated polypeptides with 33 and 29 kDa, respectively). Both chains are non-covalently associated; the alloantigenic epitopes are expressed on the distal domain of the extracellular region. Each of the class II loci comprised approximately 10 kb; the size of the whole region is 1000 kb. The HLA class II region can be subdivided into at least 3 subregions: HLA-DR, HLA-DQ and HLA-DP (205 kb, 100 kb and 90 kb, respectively); within these subregions, several loci define A and B chains.
The HLA-DR alloantigens are situated on molecules consisting of the gene products of HLA-DRA and DRB1 (DR1 - DR18), DRB3 (DR52), DRB4 (DR53) and DRB5 (DR51). The product of DRA shows a rather limited variability, DRB2 as well as DRB6 - DRB9 are pseudogenes. The number of DRB genes on a haplotype (haplotype is the designation for the combination of the alleles of each locus situated on one chromosome and inherited en bloc) is variable and depends on the main DR specificity defined by DRB1.

In the HLA-DQ subregion, there are 5 loci: DQA1, DQB1, DQA2, DQB2 and DQB3. Both DQA1 and DQB1 chains are polymorphic; the serologically defined epitopes are situated on the B chain. The products of DQA2, DQB2 and DQB3 are not known to be expressed.
The HLA-DP subregion contains 4 loci: DPA1, DPB1, DPA2 and DPB2. The DP molecules are made up of the DPA1 and the DPB1 chain; DPA2 and DPB2 are pseudogenes.
Four other HLA class II loci, DMA, DMB, DNA and DOB, are also situated in the HLA-D region; the function of their products is not yet known.

Furthermore, besides the class I and II loci, there are several other loci situated in the MHC of man:
- at least 4 loci defining proteins of the complement system (C2, C4A, C4B and BF),
- the structural gene for the 21-hydroxylase, CYP21, and an inactive pseudogene, CYP21P,
- two loci for tumor-necrosis factors, TNFA and TNFB,
- loci for the heat-shock proteins HSP70 and several loci coding for HLA-B associated transcripts (BAT1-9); the products of these loci may belong to a cell stress response system,
- genes involved in the processing of antigens: transporter genes, TAP1 and TAP2 (TAP = transporter associated with antigen processing), and proteasome-related genes, LMP2 and LMP7 (LMP = large multifunctional protease).

The probable order of these loci on the short arm of chromosome 6 is: centromere - ... - DPB2 - DPA2 - DPB1 - DPA1 - DNA - DMA - DMB - LMP2 - TAP1 - LMP7 - TAP2 - DOB- DQB2 - DQA2 - DQB3 - DQB1 - DQA1 - (DRB1 - DRB9) - DRA - ... - CYP21 - C4B - CYP21P - C4A - BF - C2 - (HSP70, BAT2 - 9) - TNFA - TNFB - BAT1 - ... - HLA-B - HLA-C - HLA-E - HLA-A - (HLA-F, HLA-G).

A strong linkage disequilibrium can be seen between various HLA alleles which is due to the preferential gametic association of some of the alleles, that means that some haplotypes are found more frequently than expected and other ones more rarely. The cause of the linkage disequilibrium is not known, the easiest explanation for its existence is the assumption that selective forces habe been acting in favour of or against some allele combinations. Because of the strong linkage disequilibrium found in the HLA system, especially in the HLA-D region between DQB1 and DRA, not all haplotypes which can be constructed by combining the alleles of the various loci are found in the population. A recombinational hot-spot is localised between HLA-DQA1 and DPA1; there is no strong positive gametic association between HLA-DP and HLA-DQ/DR alleles. The amount of the linkage disequilibrium between HLA-DQ/DR and HLA-B or between the class I loci is considerably smaller than the gametic association of the HLA-DR and DQ alleles.

2. HLA POLYMORPHISMS

2.1. Polymorphisms defined by conventional methods

2.1.1. Serologically defined factors

The class I loci HLA-A, B and C show a remarkable multiple allelism: taking into account only the specificities accepted by the HLA Nomenclature Committee 23 HLA-A, 46 HLA-B and 9 HLA-C factors are recognised. From these figures, 277 HLA-A, 1082 HLA-B and 46 HLA-C phenotypes can be computed; the number of possible HLA-ABC phenotypes amounts to 13.8×10^6. The inclusion of new factors described during the 10th and 11th International

Histocombatibility Workshops increases the number of HLA-ABC phenotypes to more than 10^8.
The class II molecules defined by serology are slightly less polymorphic: 17 HLA-DR and 7 HLA-DQ alleles are officially recognised, corresponding to 154 DR and 29 DQ phenotypes. With possibly new alleles defined in the 10th and 11th International Histocombatibility Workshops, these figures increase to more than 200 and 40, respectively. Because of the extremely strong linkage disequilibrium between HLA-DR and DQ alleles, the DQ factors do not increase the HLA polymorphism; the same applies to DR51, DR52, and DR53 genes which show an extreme linkage disequilibrium with some of the DR alleles DR1 - DR18.
The number of possible HLA phenotypes which can be detected by serology taking into account the products of the loci HLA-A, B, C and DR is higher than 20 x 10^9. This figure, however, is an underestimation as the class I loci coding for antigens on T lymphocytes and some subtypic factors of well defined determinants are not included. Furthermore, the phenotypic polymorphism is increased by the fact that the DQA1 molecule of one haplotype can combine with the DQB1 product of the other haplotype ("trans-complementation").

2.1.2. Cellularly defined specificities
It is possible to measure the response of T lymphocytes of one individual against determinants defined by the HLA-D haplotypes of another one using the one-way mixed lymphocyte culture (MLC). With the help of homozygous typing cells, several specificities, the HLA-Dw factors, could be defined. Most of them are included in HLA-DR specificities and it is highly probable that the response in the MLC reflects disparities of the whole HLA class II region; with regard to the strength of the stimulation, the following hierarchy can be envisaged: DRB > DQ > DP.
Other polymorphisms can be detected by a secondary response of T lymphocytes which have been primed in a first incubation step. The main targets detectable by this technique, the primed lymphocyte typing (PLT), are the HLA-DP factors provided that the responding cells have not been primed against HLA-DR/DQ. For the moment, 6 HLA-DP specificities (corresponding to 22 possible HLA-DP phenotypes) are officially recognised.
Besides these two methods, T cells, especially cells which have been expanded in clones, can be used in cytotoxic or proliferative assays for the definition of class I or class II polymorphisms. The subtypic factors of several class I determinants (e.g. of A2, B7 od B27) which have been detected in this way correspond very well to differences demonstrable by one-dimensional isoelectrofocusing (1D-IEF). The same applies for cellularly defined class II splits which can also be detected by one-dimensional and two-dimensional gels.

2.1.3. Biochemically defined polymorphisms
The use of 1D-IEF for the detection of class I molecules during the 10th International Histocompatibility Workshop demonstrated the existence of many subtypes of HLA-A, B and C gene products. These subtypes increase the class I polymorphism by a factor of 16.
The analysis of two-dimensional gels for HLA-DRB, DQA1 and DQB1 factors showed at least 23, 9 and 9 different patterns, respectively, in 71 highly selected homozygous cell lines. These figures clearly show that class II molecules are more polymorphic than expected on the basis of serologic data.

2.2. Polymorphisms defined by molecular genetics

2.2.1. HLA class I and class II alleles
2.2.1.1. RFLPs
The restricton fragment length polymorphisms (RFLPs) detected by class I probes are not very helpful for the definition of HLA-ABC factors: some patterns correlate well with already

known molecules (e.g. HLA-A3) and several supertypic patterns can be observed, but it is not possible to type for all class I specificities with RFLPs.
In HLA class II, the situation is different: using only 3 detection systems consisting of a specific DNA probe and a given restriction enzyme (DRB/TaqI, DQB/TaqI and DQB/BamHI), it is possible to detect polymorphisms which are highly correlated with the known DR, DQ and Dw specificities. Furthermore, the application of other systems reveals several splits, thus giving a better definition of the various HLA-D haplotypes. As the proportion between exons and introns is only 1 to 9, it is probable that the majority of the class II RFLPs is due to variations in the non-coding sequences. These polymorphisms, however, must show a very strong linkage disequilibrium with phenotypically expressed HLA factors detectable by serology, cellular techniques, biochemical methods and/or PCR.

2.2.1.2. Sequence data
The ultimate definition of an HLA factor must be the nucleotide sequence of the allele or the aminoacid sequence of the gene product. Up to now, the sequences of 41 HLA-A, 61 HLA-B, 17 HLA-C, 4 HLA-E, 2 DRA, 60 DRB1, 4 DRB3, 1 DRB4, 4 DRB5, 3 DRB6, 13 DQA1, 19 DQB1, 8 DPA1 and 37 DPB1 factors are known and they received an official designation by the HLA Nomenclature Committee (each allele was given a unique numerical designation placed after an asterisk following the gene name). The sequence data demonstrate that several molecules which appeared to be homogeneous can be further subdivided: for instance, HLA-A2 can be coded for by at least 12 different alleles, HLA-B27 by 7 alleles or HLA-DR4 by 12 alleles.
The knowledge of many class II sequences is the basis for the application of the polymerase chain reaction (PCR) in order to define HLA class II alleles. By constructing specific primers, it is possible to amplify selectively the polymorphic parts of the second exon of HLA class II alleles from genomic DNA; the amplified DNA is mainly analysed by hybridisation with sequence-specific oligonucleotides (PCR-SSO) although other detection methods can be used, e.g. PCR-SSP (PCR with sequence-specific primers), PCR-RFLP (digestion of the amplified DNA with restriction enzymes and subsequent electrophoresis in polyacrylamide gel) or direct sequencing of the amplified DNA. These techniques are also used in order to define HLA class I alleles.

2.2.1.3. Polymorphisms in the promoter region
The sequencing of the promoter regions of HLA class II loci revealed marked polymorphisms in these regions. For the moment, such polymorphisms are known for DRA, DRB, DQA1 and DQB1; they show a strong linkage disequilibrium with the corresponding HLA class II alleles.

2.2.2. Other polymorphisms of the HLA system
2.2.2.1. TNF
In the TNF region, there are at least 5 complex microsatellite polymorphisms (TNFa - TNFe) and 3 biallelic RFLPs. The alleles of these polymorphisms are in linkage disequilibrium with class II alleles.

2.2.2.2. TAP
Polymorphic residues have been described in TAP1 and TAP2. Four TAP1 alleles and 8 TAP2 alleles are possible, 3 and 5 of them have been observed. There is also a strong gametic association between TAP and HLA class II alleles.

2.2.2.3. HSP70
The three HSP70 loci (HSP70-1, HSP70-2 and HSP70-Hom) show only a restricted polymorphism in the coding or in the promoter regions.

2.2.2.4. RFLPs and microsatellites between HLA-B and HSP70
Three polymorphic genes (with 4, 3 and 7 alleles, respectively) situated between HLA-B and BAT1 can be recognised with RFLPs. Again, there is a strong gametic association between the alleles of these loci and HLA class I alleles. Furthermore, several microsatellite loci coding for a variable number of CA repeats are located between HSP70 and HLA-B.

3. CONCLUSIONS

By using molecular genetic methods, it is possible to increase the HLA polymorphism, thus reaching a better definition of the HLA haplotypes. If only the class II alleles are considered (tested e.g. by PCR-SSO), the chance of paternity exclusion reaches 90%.
Due to the very strong linkage disequilibrium between the alleles at the various loci, the informations obtained by molecular genetics are in most cases, especially if "common" haplotypes are involved, not significantly higher than the informations arising from serology. Nevertheless, the definition of HLA class I and class II alleles, of other RFLPs and of microsatellites governed by the MHC can be helpful in some paternity cases; the chance of paternity exclusion being 98% if HLA-A,B,C and DR factors are typed by serology might increase to 99%. If two putative fathers, however, cannot be excluded by the serology of HLA, it is obviously more informative to include other systems not linked with HLA (e.g. VNTR-AMPFLP or STR DNA polymorphisms) than to analyse extensively the molecular structure of the HLA haplotypes observed in such a case.

REFERENCE

Tsuji K, Aizawa M, Sasazuki T: HLA 1991, Oxford University Press 1992