# **HLA Diagnostic Sequencing – Conception, Application and Automation**

Diagnostische HLA-Sequenzierung – Konzept, Anwendung und Automatisierung

R. Blasczyk

**Summary:** Sequencing gives the most reliable and accurate information of the DNA sequence of a gene and is therefore of particular interest for fully characterising the genetic complexity and allelic diversity of the HLA genes. The full complexity of allelic diversity in HLA class I and class II genes, as well as recent improvements in the convenience and quality of automated sequencing, makes sequencing the method of choice for HLA low- and high-resolution typing. HLA typing by means of sequencing should be done whenever the HLA type of an individual is needed. The current developments with regard to HLA-sequencing systems, sequencing instruments and software solutions for data interpretation have made sequencing equally simple and robust, providing formats for each level of throughput and automation in order to match each laboratory's individual requirements. Modern sequencing is based on direct sequencing of PCR products from genomic DNA with the help of fluorescent-labelling methods using dye terminator cycle sequencing chemistries with specialised DNA polymerases and automated capillary sequencers. The presently available HLA sequencing systems differ mainly in regard to their PCR amplification strategies and software solutions for allele identification.

As the diversity of all HLA loci developed through extensive sequence exchanges, resulting in chimeric structures of each HLA allele, the allele-specific sequencing approach is the most successful for high-resolution typing (4 digits), as it is capable of determining the cis linkage of sequence motifs. This is achieved by applying multiple group-specific amplifications in parallel, thus splitting the alleles into individual groups. This approach increases the capacity for cis linkage definition and in turn decreases the number of ambiguities. When sequencing shall be established as the only method for 4-digit HLA typing, these group-specific approaches must be used. When sequencing shall be used also for

2-digit HLA typing, a gene-specific amplification approach for sequencing of both alleles simultaneously should be applied. Thus, the progress made in diagnostic HLA sequencing now offers a real alternative to the conventional molecular typing techniques, and will increasingly replace them for the patients' benefit.

Redaktion: Mariam Klouche

**Keywords:** HLA analysis; sequencing; automation.

Zusammenfassung: Die Sequenzierung liefert die zuverlässigste und genaueste Information über die DNA-Sequenz eines Gens und ist daher für die Charakterisierung der genetischen Komplexität und allelischen Vielfalt der HLA-Gene von besonderem Interesse. Die hohe Komplexität der allelischen Diversität der HLA Klasse I und II Gene und die aktuellen Verbesserungen in der Anwendung und Qualität der automatischen Sequenzierung machen die Sequenzierung zur Methode der Wahl für niedrig- und hochauflösende HLA-Typisierungen. Eine Sequenzierung sollte daher immer in Betracht gezogen werden, wenn eine HLA-Typisierung erforderlich ist. Die modernen Entwicklungen der HLA-Sequenziersysteme, Sequenziergeräte und Softwarelösungen zur Dateninterpretation haben die Sequenzierung gleichermaßen einfach und robust werden lassen, bieten jede gewünschte Stufe für Hochdurchsatz und Automatisierung und können die individuellen Anforderungen jedes einzelnen Labors erfüllen. Die moderne Sequenzierung beruht auf der direkten Sequenzierung von PCR-Produkten aus genomischer DNA unter Verwendung von Fluoreszenzmarkierungen mit 4-Farben-Terminatoren, Zyklus-Sequenzierungen mit spezialisierten DNA-Polymerasen und automatischen Kapillarsequenzierern. Die zur Zeit verfügbaren HLA-Sequenziersysteme unterscheiden sich hauptsächlich hinsichtlich ihrer PCR-Amplifikationsstrategien und der Software-Lösungen für die Allelidentifizierung. Da die Diversität aller HLA-Gene durch intensive segmentale Rekombinationen entstanden ist, stellt jedes HLA-Allel eine Chimäre anderer HLA-Allele dar. Die Allelspezifische Sequenzierung ist daher für eine hochauflösende (4-stellige) Typisierung am besten geeignet, da sie die cis-Kopplung der Sequenzmotive erkennt. Eine Allel-spezifische Sequenzierung wird erreicht, indem multiple Gruppen-spezifische Amplifikationen parallel durchgeführt werden und dadurch die beiden Allele

Institut für Transfusionsmedizin, Medizinische Hochschule Hannover

Correspondence: Prof. Dr. Rainer Blasczyk, Institut für Transfusionsmedizin, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

Str. 1, 30625 Hannover, Germany. Tel.: +49 51 15 32 67 00

Fax: +49 51 15 32 20 79

E-mail: blasczyk.rainer@mh-hannover.de

einer Probe in zwei Gruppen getrennt werden. Dieser Ansatz steigert die Fähigkeit eines Sequenziersystems die cis-Kopplungen von Sequenzmotiven zu definieren erheblich und minimiert auf diese Weise die Anzahl uneindeutiger Sequenzierergebnisse. Wenn die Sequenzierung als alleinige Methode für 4-stellige HLA-Typisierungen etabliert werden soll, müssen Gruppenspezifische Amplifikationen verwendet werden. Wenn die Sequenzierung auch für 2-stellige HLA-Typisierungen eingesetzt werden soll, können Genort-spezifische Amplifikationen zur simultanen Sequenzierung beider Allele verwendet werden. Die Fortschritte der diagnostischen Sequenzierung bieten daher bereits jetzt eine wirkliche Alternative zu konventionellen molekularen Typisierungstechniken und werden diese zum Wohle des Patienten zunehmend ersetzen.

**Schlüsselwörter:** HLA-Analyse; Sequenzierung; Automatisierung.

# **HLA-typing methods**

with respect to the clinical relevance and the extensive polymorphism of HLA antigens, adequate typing methods are extremely important. Early work on the definition of HLA antigens was aggravated by the poor reproducibility of the leukocyte agglutination test. The replacement of leukocyte agglutination by lymphocytotoxicity tests in 1964 [1] and the systematic analysis of antibodies formed in pregnancy have brought substantial progress in clinically applied HLA typing. In the 1970s, homozygous typing cells were used in the mixed lymphocyte culture (MLC) to define the polymorphism at the HLA-D locus, which enabled the discrimination of 26 antigens named HLA-Dw1-Dw26 [2]. In the late 1970s, sera obtained from multiparous women, containing HLA antibodies, showed similar specificities as the MLC for HLA-D typing. The use of these sera permitted the definition of HLA-DR (D related) and later also the identification of a second HLA class II locus, which was named HLA-DQ [3]. The third HLA class II locus, HLA-DP, was detected in the late 1970s as a weak MLC locus, the polymorphism of which was defined by a secondary MLC or primed lymphocyte test (PLT). The application of the PLT allowed the discrimination of further six antigens: HLA-DPw1-DPw6 [4, 5]. Of note, HLA-DP could not be defined by means of the serological lymphocytotoxicity test.

In the course of the 1980s, the antibody-mediated micro-lymphocytotoxicity test was the most widely used assay for HLA-A, B, C, DR and DQ typing. In addition, for bone marrow transplant programmes, the PLT was used to define the HLA-DP specificities and the MLC to define serologically undefined variants of the HLA-DR and DQ loci.

These conventional typing methods suffer from a limited accuracy and reliability, especially when they

are routinely used for clinical purposes [6–8]. With the discovery of the polymerase chain reaction (PCR), the polymorphism of the MHC genes could be studied directly on the nucleotide level. With the increasing availability of sequenced alleles, the development of PCR-based typing techniques in the late 1980s and the 1990s increasingly replaced conventional typing methods. The PCR-based methods for HLA class II genes have meanwhile completely superseded serological typing and MLC. For HLA class I, a few laboratories are still content with the continuing use of serology, but as the beneficial effects of PCR-based typing approaches have been clearly demonstrated, the latter are now also considered as the standard typing technique for HLA class I.

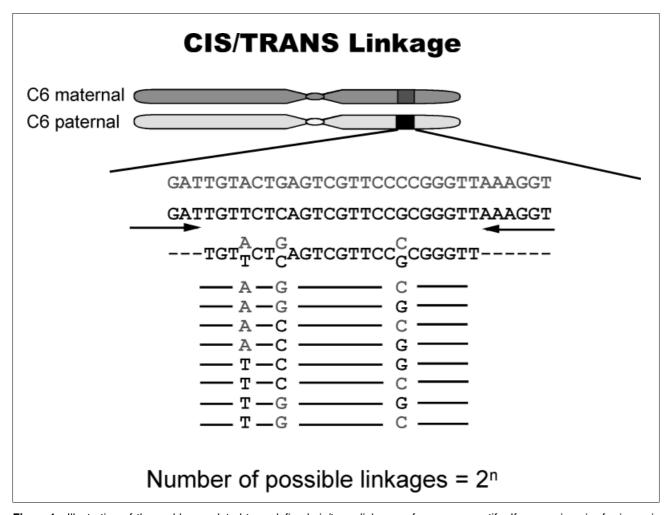
# Nature of HLA diversity

The PCR technology has made it possible to investigate fragments of genes and has emerged as one of the most important molecular techniques for HLA typing. As the polymorphism of HLA genes is mainly restricted to only a few exons, the PCR-based methods for HLA-typing purposes focus on only small fragments of HLA genes. These are mainly exon 2 in HLA class II genes and exons 2 and 3 in HLA class I genes.

The PCR-based diagnostics of the polymorphic HLA system requires the consideration of some general rules concerning the polymorphic structure of HLA genes. The diversity of HLA genes at the antigen-presenting sites has evolved on the one hand by the gradual accumulation of point mutations during successive ancestral species, and on the other hand by gene conversion events or recombinations [9, 10]. The last mechanism appears to be the most important factor for the enormous diversification of HLA genes. This has consequences for their polymorphic structure. When scrutinising the sequences of the different alleles per locus, it becomes apparent that the allelic diversity is not characterised by allele-specific point mutations but by different combinations of sequence motifs from a common pool of different motifs. This patchwork structure of allelic sequences is reconciled with the polymorphism being generated by segmental sequence exchanges. Thus, most of the alleles can be considered as chimeras formed from various segments of other alleles. In most cases, newly identified alleles are characterised by a new combination of already existing sequence motifs. From these observations, it can be concluded that the number of alleles at each locus is represented by all theoretically possible combinations of known sequence motifs.

# **Principle of PCR-based typing methods**

These chimerical features of the polymorphic HLA gene structure have consequences with respect to PCR-



**Figure 1** Illustration of the problems related to undefined cis/trans linkages of sequence motifs. If a generic pair of primers is used for amplification, the PCR product contains the alleles of both haplotypes. Direct sequencing of this PCR product generates sequencing fragments of both alleles simultaneously, resulting in heterozygous peak patterns in the electropherogram at the indicated positions. Since the cis/trans linkage is not defined, the number of possible linkages is  $2^n$ , in which n is the number of heterozygous positions. The HLA typing is done by comparing all theoretically possible linkages with the most relevant HLA sequence database. This is the major reason for HLA typing ambiguities. In unambiguous typings, this bears the risk that the typing result may become ambiguous over the years, in case new alleles are identified.

based HLA-typing strategies. More than 90% of the individuals are heterozygous for each HLA locus. Therefore, for HLA typing at the allelic level, it is generally not sufficient to prove the presence of certain sequence motifs, but to demonstrate their cis/trans linkage, i. e. to demonstrate how the motifs are linked on each of the two haplotypes (Fig. 1). Exclusively this approach is capable of delivering unambiguous typing results. If the cis/trans linkage of variable sites is not defined, multiple heterozygous allele combinations cannot be distinguished from each other. Additionally, new alleles may be mistyped as a heterozygous combination of known alleles.

In terms of PCR methodology, those techniques that lead to the amplification of both alleles present on a particular HLA locus are called a *generic* PCR. The re-

sulting PCR product contains the gene fragments of both alleles, ideally in a more or less equal amount. In practice, however, depending on the PCR conditions, the quality of the amplification primers and the analysed locus, there is usually a preferential amplification of one allele over the other. Preferential amplification may cause problems if the generated PCR fragments are used for direct sequencing. Those techniques that lead to the amplification of a limited number of related alleles are called a group-specific PCR. The group-specific PCR takes advantage of the closer sequence homology of a certain cluster of alleles. This sequence homology usually reflects the fact that these groupspecifically amplified alleles belong to the same serologically defined group. Thus, the term 'group specific' normally indicates a PCR amplification that is, with regard to specificity, related to the serological variability of a locus, e.g. specific for all alleles having the serological phenotype HLA-A1 or HLA-A2. At least one of the primers used in group-specific PCR is located on a sequence motif that is present only in a limited number of alleles.

The group-specific amplification strategy is one of the most important approaches to defining the cis/trans linkage of sequence motifs. If the technique used for PCR amplification is capable of amplifying the alleles of a HLA locus according to the serological variability, then the alleles of all serologically heterozygous individuals can be amplified separately. This way of amplification is therefore often called an *allele- or haplotype-specific* PCR. However, the term 'allele-specific' should be reserved for those PCR amplifications that allow an allele assignment just from the information delivered by the presence or absence of a PCR fragment.

A group-specific PCR amplification is not in all cases haplotype-specific. In those individuals who are heterozygous for different alleles belonging to the same amplification group, the PCR leads to the amplification of both alleles simultaneously, a situation which is similar to the generic PCR. However, due to the close sequence homologies, the quantitative relationship between the amplified fragments of the different alleles is usually more homogeneous than in a generic PCR, facilitating the post-PCR specificity steps. In addition, these closely related alleles sometimes differ only at a single nucleotide, which allows the definition of the cis/trans linkage even without haplotype-specific PCR.

Thus, the principle of modern DNA-based HLA-typing methods is the haplotype-specific PCR amplification of the polymorphic exons of the HLA genes, followed by postamplification specificity steps determining the hypervariable regions for allele assignment.

# PCR-based approaches to HLA typing

In the course of the development of PCR-based methods for HLA typing, a large variety of techniques has been described, which mainly differ in their postamplification specificity strategy for determining the individual sequence of the amplified fragment. The techniques used to analyse the amplified fragments include the restriction fragment length polymorphism (PCR-RFLP) [11], the single-strand conformation polymorphism (PCR-SSCP) [12], the heteroduplex formation (PCR-HDF) [13], the use of sequence-specific oligonucleotide probing (PCR-SSOP) [14], the use of sequence-specific priming (PCR-SSP) [15], and the determination of the nucleotide sequence of the PCRamplified DNA by sequencing-based typing (PCR-SBT) [16–18]. In the course of the 11th International Histocompatibility Workshop (IHWC), the PCR-SSOP method for HLA class II genes had been standardised, which allowed the introduction of PCR-based typing methods to histocompatibility laboratories worldwide [19]. The routine use of PCR-based techniques in many laboratories all over the world has brought deep insights into the variability of HLA genes, which emerged to be much more variable than could have been expected from the conventional typing methods. This increasing diversity has complicated more and more the PCR-based typing methods, so that in the course of the last years, only three methods (PCR-SSOP, PCR-SSP, and PCR-SBT) have survived and are mainly used for HLA typing on the nucleotide level.

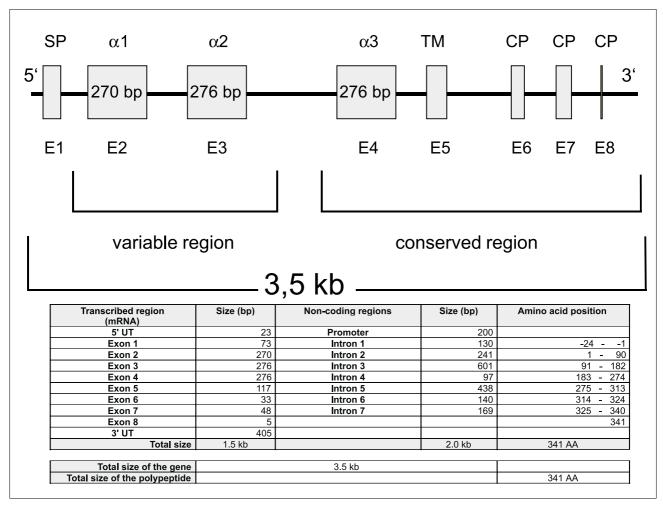
# Sequencing-based typing

Sequencing gives the most reliable and accurate information of the DNA sequence of a gene and is therefore of particular interest for the determination of the full extent and complexity of the HLA polymorphism. In contrast to other PCR-based typing approaches, sequencing is not limited to the detection of known sequence polymorphisms in the hypervariable regions. This may also apply to conformation analysing methods (single-strand conformation polymorphism and heteroduplex analysis) [20, 21]. However, the PCR design for these techniques is rather complicated and they neither provide 100% sensitivity, nor reveal the nature of the variabilities in terms of the DNA sequence. Moreover, depending on the methodical set-up, sequencing may offer the possibility of defining the cis/trans linkage of sequence motifs.

Sequencing for HLA-typing purposes does not usually require the consideration of the complete gene, but is restricted to those parts of the genes displaying the polymorphic sites, i.e. exon 2 in HLA class II and exons 2 and 3 in HLA class I genes (Figs 2 and 3). Therefore, the method of sequencing exclusively the typing-relevant regions as a post-PCR specificity step has been designated sequencing-based typing (PCR-SBT).

Over the last years, sequencing-based typing of HLA genes has progressed considerably, allowing the implementation of PCR-SBT for routine HLA typing. Although all sequencing proposals are based on the dideoxynucleotide chain termination technique of enzymatically synthesised DNA fragments, first described by Sanger et al. [22], the term PCR-SBT initially covered a very heterogeneous field of methodical approaches with respect to the starting material, the amplification strategies, the template preparation, and the sequencing chemistry [16–18, 23–25]. Currently, much of this heterogeneity has now been overcome, and modern sequencing is based on direct sequencing of PCR products from genomic DNA with the help of fluorescent-labelling methods using dye terminator cycle sequencing chemistry with specialised DNA polymerases and automated capillary sequencers [26].

Among the PCR-based HLA-typing methods, PCR-SBT requires the most expensive equipment. This is due to the fact that sequencing for clinical purposes can

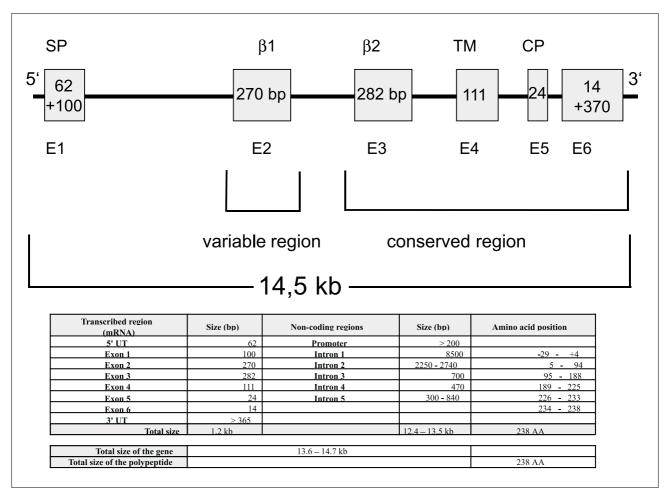


**Figure 2** Structure of HLA class I genes. The polymorphism is nearly exclusively restricted to exons 2 and 3.  $\alpha$ 1- $\alpha$ 3,  $\alpha$ 1- $\alpha$ 3 domains; CP, cytoplasmic part; FR, flanking region; UT, untranslated region; E, exon; I, intron; SP, signal peptide; TM, transmembrane part.

only be done reasonably with the help of an automated sequencer. On the other hand, sequencing offers several advantages over other methods (Table 1): (i) sequencing gives the most accurate information because it provides not only the nucleotide sequence but also has the highest potential for defining the cis-linkage of sequence motifs, (ii) sequencing has a very high potential for miniaturisation, and (iii) sequencing has a very high potential for automation. In contrast to all other PCRbased HLA-typing methods, sequencing is the only method capable of fulfilling all HLA-typing requirements without the need of any additional typing method. This applies to low- (2-digits) as well as to highresolution (4-digits) typing, and to low- as well as to high-throughput formats, making sequencing equally attractive to small, intermediate, and large laboratories.

#### Amplification strategies in PCR-SBT

PCR amplification of polymorphic systems for direct sequencing can be performed gene or haplotype specific. The latter approach has two distinct advantages. First, the cis/trans linkage of sequence motifs can be defined and, second, problems related to preferential amplification of one haplotype are avoided. In contrast, PCR-SBT, after generic PCR amplification, is unable to define the cis/trans linkage of sequence motifs, which means, for typing uses, it is similar to oligotyping. The rapidly growing number of newly identified alleles confirms that new alleles arise mainly from gene conversion events that take place between different alleles of the same locus. Newly identified alleles are not characterised by new sequence motifs, but rather by new combinations of already existing sequence motifs. The conclusion from this observation is that the possible number of alleles at each locus is the theoretical number of all recombinations of known sequence motifs.



**Figure 3** Structure of HLA class II DRB genes. The polymorphism is nearly exclusively restricted to exon 2.  $\beta$ 1,  $\beta$ 2,  $\beta$ 1 and  $\beta$ 2 domains; CP, cytoplasmatic part; FR, flanking region; UT, untranslated region; E, exon; I, intron; SP, signal peptide; TM, transmembrane part.

This suggests that, although many of these theoretically possible alleles will probably be subject to negative selection, the number of as yet unidentified alleles is enormous. If the cis/trans linkage of the analysed polymorphic regions is not defined, then some new alleles may be mistyped as a heterozygous combination of known alleles.

This has consequences with respect to SBT strategies, i.e. an unambiguous typing result of SBT after generic PCR amplification is only unambiguous with regard to the presently known HLA database. However, as more and more new alleles will be detected and defined in the course of the next few years, the unambiguous result obtained today can become ambiguous in the future. Such tendencies were observed, in the past, with PCR-based HLA-DRB1 typing, and are likely to reoccur with PCR-based class I typing. To support the above, unambiguous typing results only remain unambiguous if the cis/trans linkage is defined. Since these results will remain unambiguous regardless

of an ever-growing HLA sequence database, this typing strategy is of great importance with regard to clinical PCR typing.

If generic amplification is used and both alleles are amplified simultaneously, it often occurs that one allele is amplified more efficiently than the other [27–29]. This problem of preferential amplification can lead to an allelic drop out during the sequencing reaction, leading to problems such as overlooking one haplotype or falsely assigning homozygosity.

The currently available PCR-SBT systems (Abbott Laboratories, Abbott Park, IL, USA; Forensic Analytical, Hayward, CA, USA; Protrans, Ketsch, Germany) differ mainly with regard to their amplification strategies. The most advanced product line for HLA sequencing is presently offered by Protrans, covering gene-specific approaches for complete HLA genes up to group-specific approaches for each HLA gene using 4, 8 or 16 group-specific primer mixes for maximal cislinkage definition.

Table 1 Differences between the major PCR-based typing methods				
	PCR-SSOP	PCR-SSP	PCR-SBT gene specific	PCR-SBT group specific
Complete exon sequences	-	_	+	++
Definition of cis-linkages	_	+	-	++
Automation	+	_	+	++
Miniaturisation	+	++	++	++
SBT = sequencing-based typing, SSOP = sequence specific oligonucleotide probing, SSP = sequence specific priming				

#### **Detection of PCR amplification**

All approaches working with gene-specific amplification do not require the application of a PCR product detection method, as the PCR should always be positive. All other approaches working with several group-specific primer mixes in parallel do require the application of a PCR product detection method. Only when group-specific primer mixes are selectively used based on an already existing low-resolution typing PCR product detection can be left out.

Usually the presence or absence of PCR products is proven by an agarose gel-based read-out. When for example eight group-specific primer mixes are used for HLA-DRB1 sequencing, 96 PCRs have to be checked for 12 DNAs to be sequenced. With the help of microplate-compatible gels and 8-channel pipettors, a reasonable throughput is achievable. One has to keep in mind that the time-consuming necessity of PCR product detection in group-specific amplification approaches is later compensated by the considerably facilitated allele assignment and by the almost completely abrogated ambiguities after sequencing.

However, the requirement of agarose gel-based readout prevents the implementation of a high-throughput of samples, thus creating the demand for alternatives, in particular when complete automation is desired. An alternative way of PCR product detection is achieved by fluorescence-based read-out using the 5' nuclease technology. This has been evaluated for sequence-specific primed PCR [30, 31] and is currently adapted to the group-specific amplification approaches in PCR-SBT. As soon as this will be available, the complete process of sequencing from DNA isolation to allele assignment will be automatable.

#### **Sequencing chemistries and sequencing instruments**

For automated Sanger-sequencing, different methods were available, which in turn varied significantly with respect to the effort required to carry them out and the accuracy of the results obtained. Only a few years ago, the dye primer technique, using fluorescent-labelled primers and sequenase as a sequencing enzyme, gave the highest reliability for the detection of heterozygous positions [27, 28]. With the availability of the new enzymes AmpliTaq<sup>TM</sup> FS (Applied Biosystems, Foster

City, CA, USA) or Thermo Sequenase™ (Amersham Bioscience, Freiburg, Germany), almost identical results were obtained with less effort (cycle sequencing) [17]. With the development of terminators with different fluorescent labels (dye terminator technique) and the use of one-lane sequencers, the workload was reduced further [32]. In particular, the recent developments of modern dye terminators and of multicapillary sequencers with 4–384 capillaries (Amersham Bioscience, Applied Biosystems, Beckman Coulter, Allendale, NJ, USA) have made sequencing at the same time simple and robust, making it attractive for clinical laboratories and thus bringing automation of sequencing closer to reality.

The currently available PCR-SBT systems for HLA typing (Abbott Laboratories, Forensic Analytical, Protrans) are exclusively working with modern dye terminator cycle sequencing chemistry and can be run on most (Abbott Laboratories, Forensic Analytical), possibly all (Protrans) capillary sequencing instruments currently available.

The number of sequencing reactions to be carried out per sample is identical in gene- and group-specific amplification approaches. In all gene-specific approaches, both alleles are sequenced simultaneously, which requires the reliable detection of heterozygous positions. Here, very often peak shifts can be observed [33], mimicking insertions and making it difficult to accurately assign heterozygous positions. In order to overcome this problem, sequencing in both orientations has to be performed to compensate for the deficiencies that may have occurred in each direction. In hemizygote sequencing after PCR-based separation of the alleles, peak shifts do not occur, therefore sequencing need only be carried out in a single direction. Thus, the necessity of sequencing both haplotypes separately after group-specific PCR is compensated by not needing to sequence in both orientations.

#### **Database and allele assignment**

The nucleotide sequences of HLA alleles are accessible via public nucleotide sequence databases such as GenBank (http://www.ncbi.nlm.nih.gov/Web/Search/index.html), EMBL (European Molecular Biology Laboratory; http://www.embl-heidelberg.de) or DDJB (DNA

Data Bank of Japan; http://www.ddbj.nig.ac.jp), which offer their services via the World Wide Web. The WHO Nomenclature Committee reports and updates for factors of the HLA system list the database accession numbers for each specificity, so that their sequence can easily be retrieved from the servers. The most relevant alignment of HLA sequences is available through the gopher of the 'HLA Informatics Group' (http://www.anthonynolan.com/HIG/index.html). The sequences on this server are compiled by experts in the field, are available as static alignments or via interactive tools, are updated quarterly and are accessible to everybody free of charge.

The most important bottleneck of applying sequencing for diagnostic purposes was the final allele assignment due to the lack of convenient bioinformatics. For clinical laboratories performing HLA sequencing, allele assignment can be done reasonably only with the help of specialised software. Currently, two software packages are available for automated allele assignment (Matchmaker, Abbott Laboratories; Sequence Pilot, Protrans). Both Matchmaker and Sequence Pilot work with the sequence database provided by the HLA Informatics Group, and both companies offer regular updates of their sequence databases according to the quarterly updates provided by the HLA Informatics Group. For clinical laboratories with high sample throughput, Sequence Pilot offers advantageous features with regard to heterozygous detection, sample management, interface to laboratory information systems, client-server applications and integration in automated applications.

# **Ambiguities**

The introduction of molecular typing methods for HLA around 1990 has led to the elucidation of a continuously increasing number of alleles. The recent release of the International Immunogenetics Project (IMGT)/HLA Sequence Database (July 2003) currently lists 1699 allele sequences (Table 2). The allelic variability of HLA class I genes is primarily located in exons 2 and 3, that of HLA class II genes in exon 2. Molecular typing techniques must at least take into account the polymorphic sites of these exons. However, the large number of alleles and the polymorphisms outside the above-mentioned exons has led to an increasing number of ambiguities in all PCR-based typing techniques, including sequencing. The database-related ambiguities are listed under http://www.ebi.ac.uk/imgt/hla/ambig.html and are updated with each new release of the database. The document includes a list of all alleles that are identical over exons 2 and 3 for HLA class I and exon 2 for HLA class II; in addition, all ambiguous results obtained when both alleles of a locus are analysed simultaneously are included.

The ambiguities in sequencing-based typing have two reasons. (a) The probably less important reason is the diversity located outside the region sequenced, if 2

**Table 2** Number of alleles per locus as recognised by the WHO Nomenclature Committee for factors of the HLA system (www.anthonynolan.com/HIG/index.html; release 2.2.0, July 2003)

Lacus	Nember of allalas
Locus	Number of alleles
HLA-A	282
HLA-B	540
HLA-C	136
HLA-DRA	3
HLA-DRB	418
HLA-DQA1	24
HLA-DQB1	53
HLA-DPA1	20
HLA-DPB1	103

alleles have identical sequences over this region. These ambiguities may be solved by extending the sequenced region, but from a clinical point of view, these polymorphisms outside the functionally relevant exons are probably less important. (b) The major reason for ambiguities is the heterozygous nature of most sequencing approaches, which leaves the cis linkage of sequence motifs undefined due to simultaneous sequencing of both alleles. Over 90 % of the ambiguities are caused by undefined cis linkages and can be solved exclusively by separating the alleles. Moreover, today's unambiguous typing results obtained after simultaneous sequencing of both alleles may be an ambiguous result of tomorrow. This is due to the increasing number of alleles, which denotes the requirement to include the date of the sequence database used in the HLA-typing result reports. As the sequence database will continuously increase, it can be expected that unambiguous allele identifications will soon only be obtained by methods applying allele-separating sequencing strategies.

Thus, when simultaneous sequencing after gene-specific PCR is chosen in clinical laboratories for HLA-typing purposes, additional methods, in particular sequence-specific primed PCR, are required to sort out ambiguities. When sequencing shall be established as the only method for HLA typing, the group-specific approaches must be used. However, if sequencing is used as the sole typing technique, it might be reasonable to use gene-specific amplification, i.e. simultaneous sequencing of both alleles, for 2-digit typing additionaly. This 2-digit sequencing is the most important indication for gene-specific amplification approaches in sequencing-based typing.

#### Automation

Among the molecular typing techniques available, sequencing has the highest potential for being automated. For each step of sample processing, including DNA iso-

lation, PCR set-up, PCR product purification, sequencing set-up, sequencing reaction purification, electrophoresis and allele assignment, automated solutions have been developed. The growing number of known HLA alleles requires a cis-linkage assessment of sequence motifs in order to get unambiguous typing results. This is best done by multiple PCRs in a sequencespecific primed amplification format with subsequent sequencing of selected amplicons. Besides the software-based final allele assignment, these are the greatest challenges in automating sequencing based on PCR and sequencing set-up robots. This has been achieved by establishing flexible pipetting workstations equipped with a 4-8 channel pipettor for sample preparation and a 96-channel pipettor for high-throughput sample distribution.

To lead the user through the different steps of sequencing and to integrate the robotic platforms, a bioinformatic tool - Sequence Assistant (Protrans) - has been developed. Sequence Assistant controls the complete sample management following DNA isolation. After assigning the processing profile to each DNA, the software generates pipetting schemes, which can be used for manual processing or which can be transferred as a worktable to the PCR set-up robot. Following PCR, the agarose gel-based or fluorescence-based read-outs of the multiple primer mixes run in parallel are then further processed by the software, which generates worktables for further manual or robotic processing. The electronic worktable is transferred to the sequencing set-up robot, which then selectively picks the PCR products for sequencing. This robot performs all steps following the PCR reactions, including PCR product purification, pipetting, running and clean-up of the sequencing reactions. The ready to run sequencing reactions are simply transferred to the sequencer, which performs electrophoretic separation according to the plate record provided by Sequence Assistant. Thus, sequencing can be carried out manually, fully automatically, or at any level of automation depending on the laboratory's individual requirements.

Taken together, the complete process of allele separating sequencing can be subjected to automation. Indeed, this also applies to gene-specific approaches. However, the multiple ambiguities obtained after simultaneous analysis of both alleles prevent reasonable automation, as resolving these ambiguities requires extensive manual intervention.

# Clinical applications, 2-digit and 4-digit typing

HLA typing by means of molecular techniques should be done whenever the HLA type of an individual is needed. Reasons for tissue typing include matching for solid-organ and bone-marrow transplantation, transfusion of platelets, disease association, forensic and anthropological studies, and studies of T-cell-mediated immunity. The need for HLA matching of donor and recipient for solid-organ and stem-cell transplantation has mainly driven the requirement for identification of specificities in the HLA system. In clinical circumstances, HLA typing is primarily essential for transplantation and transfusion purposes [34–38]. Depending on the clinical demand – transfusion, solid-organ or stem-cell transplantation – different levels of resolution are required. In solid-organ transplantation and platelet transfusion, HLA typing is generally restricted to the serology equivalent or 2-digit level of resolution.

In allogeneic stem-cell transplantation, a high-resolution typing on the allelic or 4-digit level is required at least for HLA-DRB1 [37, 38]. Though high-resolution typing of DRB1 has been done in most cases by PCR-SSOP and PCR-SSP, the rapidly growing number of known alleles and the continuous methodical advancements have made this a domain of PCR-SBT. Even though the importance of individual basepair mismatches of HLA class I alleles for the clinical outcome has yet to be established and is still a matter of discussion, PCR-SBT may become the method of choice also for routine high-resolution HLA class I typing.

It is noteworthy that sequencing is mostly considered to be a high-resolution or 4-digit HLA-typing method. However, in view of the expanding HLA database, also low-resolution or 2-digit typing requires the consideration of more and more sequence information. Together with the improvements of convenience and quality in automated sequencing, PCR-SBT is also the best choice for 2-digit typing. Except for tissue-typing of cadaver donors, which demands PCR-SSP due to time constraints, sequencing should always be considered if the HLA type of an individual is needed, independent of the level of resolution required.

#### References

- 1. Terasaki PI, McClelland JB. Microdroplet assay for human serum cytotoxins. Nature 1964;204:998–999.
- **2.** Mempel W, Grosse-Wilde H, Baumann P, Netzel B, Steinbauer-Rosenthal I, Scholz S, Bertrams J, Albert ED. Population genetics of the MLC response: typing for MLC determinants using homozygous and heterozygous reference cells. Transplant Proc 1973;5:1529–1534.
- **3.** Tosi R, Tanigaki N, Centis D, Ferrara GB, Pressman D. Immunological dissection of human Ia molecules. J Exp Med 1978;148: 1592–1611.
- **4.** Mawas C, Charmot D, Sivy M, Mercier P, North ML, Hauptmann G. A weak human MLR locus mapping at the right of a crossing-over between HLA-D, Bf and GLO. J Immunogenet 1978;5: 383–395.
- **5.** Shaw S, Johnson AH, Shearer GM. Evidence for a new segregant series of B cell antigens that are encoded in the HLA-D region and that stimulate secondary allogenic proliferative and cytotoxic responses. J Exp Med 1980;152:565–580.
- **6.** Mytilineos J, Scherer S, Dunckley H, Trejaut J, Chapman J, Fischer G, Fae I, Middleton D, Savage D, Bignon JD, et al. DNA HLA-DR typing results of 4000 kidney transplants. Transplantation 1993;55:778–781.

- 7. Yu N, Ohashi, M, Alosco S, Granja C, Salazar M, Hegland J, Yunis E. Accurate typing of HLA-A antigens and analysis of serological deficiencies. Tissue Antigens 1997;50:380–386.
- **8.** Lorentzen DF, Iwanaga KK, Meuer KJ, Moritz TL, Watkins DI. A 25% error rate in serological typing of HLA-B homozygotes. Tissue Antigens 1997;50:359–365.
- **9.** Parham P, Lawlor DA, Lomen CE, Ennis PD. Diversity and diversification of HLA-A,B,C alleles. J Immunol 1989;142:3937–3950.
- **10.** Klein J. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum Immunol 1987;19:155–162.
- 11. Uryu N, Maeda M, Ota M, Tsuji K, Inoko H. A simple and rapid method for HLA-DRB and -DQB typing by digestion of PCR-amplified DNA with allele specific restriction endonucleases. Tissue Antigens 1990;35:20–31.
- **12.** Blasczyk R, Hahn U, Wehling J, Huhn D, Salama A. Complete subtyping of the HLA-A locus by sequence-specific amplification followed by direct sequencing or single-strand conformation polymorphism analysis. Tissue Antigens 1995;46:86–95.
- 13. Savage DA, Tang JP, Wood NA, Evans J, Bidwell JL, Wee JL, Oei AA, Hui KM. A rapid HLA-DRB1\*04 subtyping method using PCR and DNA heteroduplex generators. Tissue Antigens 1996;47: 284–292.
- **14.** Bignon JD, Fernandez-Viña MA. Protocols of the 12th International Histocompatibility Workshop for typing of HLA class II alleles by DNA amplification by the polymerase chain reation (PCR) and hybridization with sequence specific oligonucleotide probes (SSOP). In: Charron D (ed.): Genetic diversity of HLA, functional and medical implications, vol. 1. Paris, EDK, 1997, pp 584–632.
- **15.** Bunce M, O'Neill CM, Barnardo MC, Krausa P, Browning MJ, Morris PJ, Welsh KI. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). Tissue Antigens 1995;46:355–367.
- **16.** Norgaard L, Fugger L, Jakobsen BK, Svejgaard A. Sequencing-based typing of HLA-A locus using mRNA and a single locus-specific PCR followed by cycle-sequencing with AmpliTaq DNA polymerase, FS. Tissue Antigens 1997;49:455–465.
- **17.** Voorter CEM, Rozemuller EH, de Bruyn-Geraets D, van der Zwan AW, Tilanus MG, van den Berg-Loonen EM. Comparison of DRB sequence-based typing using different strategies. Tissue Antigens 1997;49:471–476.
- **18.** Kotsch K, Wehling J, Köhler S, Blasczyk R. Sequencing of HLA class I genes based on the conserved diversity of the noncoding regions: sequencing-based typing of the HLA-A gene. Tissue Antigens 1997;50:178–191.
- 19. Kimura A, Sasazuki T. Eleventh international histocompatibility workshop reference protocol for the HLA DNA-typing technique. In: Tsuji K, Aizawa M, Sasazuki T (eds): HLA 1991. Proceedings of the Eleventh International Histocompatibility Workshop and Conference, vol. 1. Oxford, Oxford Science Publications, 1992, pp 397–419.
- **20.** Arguello JR, Madrigal JA. HLA typing by reference strand mediated conformation analysis (RSCA). Rev Immunogenet 1999;1: 209–219.
- **21.** Buchler T, Gallardo D, Rodriguez-Luaces M, Pujal JM, Granena A. Frequency of HLA-DPB1 disparities detected by reference strand-mediated conformation analysis in HLA-A, -B, and -DRB1 matched siblings. Hum Immunol 2002;63:139–142.

- **22.** Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977;74:5463–5467
- **23.** Rozemuller EH, Bouwens AG, Bast BE, Tilanus MG. Assignment of HLA-DPB alleles by computerized matching based upon sequence data. Hum Immunol 1993;37:207–212.
- **24.** Versluis LF, Rozemuller E, Tonks S, Marsh SG, Bouwens AG, Bodmer JG, Tilanus MG. High-resolution HLA-DPB typing based upon computerized analysis of data obtained by fluorescent sequencing of the amplified polymorphic exon 2. Hum Immunol 1993;38:277–283.
- **25.** Spurkland A, Knutsen I, Markussen G, Vartdal F, Egeland T, Thorsby E. HLA matching of unrelated bone marrow transplant pairs: direct sequencing of in vitro amplified HLA-DRB1 and -DQB1 genes using magnetic beads as solid support. Tissue Antigens 1993;41:155–164.
- **26.** Motz M, Paabo S, Kilger C. Improved cycle sequencing of GCrich templates by a combination of nucleotide analogs. Biotechniques 2000;29:268–270.
- **27.** Rozemuller EH, Eliaou JF, Baxter-Lowe LA, Charron D, Kronick M, Tilanus MG. An evaluation of a multicenter study on HLA-DPB1 typing using solid-phase Taq-cycle sequencing chemistry. Tissue Antigens 1995;46:96–103.
- **28.** McGinnis MD, Conrad MP, Bouwens AG, Tilanus MG, Kronick MN. Automated, solid-phase sequencing of DRB region genes using T7 sequencing chemistry and dye-labeled primers. Tissue Antigens 1995;46:173–179.
- **29.** Voorter CEM, Kik MC, van den Berg-Loonen EM. High-resolution HLA typing for the DQB1 gene by sequence-based typing. Tissue Antigens 1998;51:80–87.
- **30.** Slateva K, Camps MA, Blasczyk R. Fluorotyping of HLA-A by sequence-specific priming and fluorogenic probing. Tissue Antigens 1998;52:462–472.
- **31.** Slateva K, Elsner HA, Albis-Camps M, Blasczyk R. HLA-DRB fluorotyping by dark quenching and automated analysis. Tissue Antigens 2001;58:250–254.
- **32.** Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA, Baumeister K. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science 1987;238:336–341.
- **33.** Rozemuller EH, Chadwick B, Charron D, Baxter-Lowe LA, Eliaou JF, Johnston-Dow L, Tilanus MG. Sequenase sequence profiles used for HLA-DPB1 sequencing-based typing. Tissue Antigens 1996;47:72–79.
- **34.** Claas FH. Predictive parameters for in vivo alloreactivity. Transpl Immunol 2002;10:137–142.
- **35.** Taylor CJ, Dyer PA. Maximizing the benefits of HLA matching for renal transplantation: alleles, specificities, cregs, epitopes, or residues. Transplantation 1999;68:1093–1094.
- **36.** Schiffer CA. Diagnosis and management of refractoriness to platelet transfusion. Blood Rev 2001;15:175–180.
- **37.** Petersdorf E, Anasetti C, Martin PJ, Woolfrey A, Smith A, Mickelson E, Malkki M, Lin MT, Hansen JA. Genomics of unrelated-donor hematopoietic cell transplantation. Curr Opin Immunol 2001;13: 582–589.
- **38.** Charron D. Immunogenomics of hematopoietic stem cell transplantation. Transfus Clin Biol 2003;10:156–158.