Sequence similarity matching: proposal of a structure-based rating system for bone marrow transplantation

H.-A. Elsner & R. Blasczyk

Summary

Recent advances in DNA-based typing have led to the detection of a continuously growing number of HLA alleles. For this reason, HLA matching in transplantation of hematopoietic stem cells from unrelated donors has become increasingly complicated. When there is no genotypically identical sibling and there are several alternative potential donors that all have a mismatch at a relevant HLA locus, until now no rating system has existed indicating different levels of allogenicity. In order to find a theoretical approach to this problem we propose a rating system ('dissimilarity index') based on structural data of HLA class I molecules, and on published data about frequencies of naturally occurring amino acid exchanges. For demonstration we employ our rating system for the comparison of the HLA-A*23 and A*24 groups, both of which allelic products are subdivisions of the serological HLA-A9 family. Remarkable differences between the subtypes were revealed, which were superior to a simple sequence comparison. More surprisingly, it was uncovered that some alleles of the A*24 group showed fewer differences to A*2301 than to alleles within their own subtype group. Sequence similarity matching may serve as a starting point for the clinical evaluation of acceptable mismatches within the HLA-A9 family and serve as a model for other HLA class I groups.

Introduction

Bone marrow transplantation from an HLA-matched sibling has been demonstrated to be a potentially curative therapy for a variety of diseases, mainly acute leukaemia, chronic myeloid leukaemia (CML) and other haematologic malignancies (Madrigal *et al*., 1997; Petersdorf *et al*., 1998; Mickelson *et al*., 2000). However, given typical family size in North America and Europe, only approximately 30% of the patients to be considered for this therapy will have a HLA genotypically identical sibling. For this reason haematopoietic stem cells from HLA-matched unrelated donors are increasingly used for transplantation (Madrigal *et al*., 1997). The major problem in bone marrow transplantation is the development of graft-vs.-host disease (GVHD), whereby donor T lymphocytes recognize allo-HLA determinants in the recipient's tissue, leading to its destruction. For this reason, HLA matching of potential donor and recipient is essential in order to improve the outcome of bone marrow transplantation. However, the requirements with regard to HLA compatibility and the relative importance of matching for individual HLA alleles in bone marrow transplantation have not yet been established (Sasazuki *et al*., 1998). In a recent study, mismatches for HLA-A and -C alleles, as determined by DNA-based typing, were found to be strong risk factors for the development of grade III or IV acute GVHD (Sasazuki *et al*., 1998). In contrast, Petersdorf *et al*. (1998) found that mismatching for a single class I or class II allele had no significant effect on survival. However, in that study mortality was increased by mismatching for more than one class I allele and by simultaneous mismatching for class I and II alleles.

Recent advances in DNA-based typing have led to the detection of a continuously growing number of HLA class I alleles, mainly HLA-A and HLA-B (Marsh *et al*., 2001). The extensive polymorphism of HLA class I molecules results from natural selection, presumably as a consequence of the advantage of T-cell responses against the diverse pathogens to which human populations are exposed. The polymorphism has evolved on the one hand by the gradual accumulation of point mutations and on the other hand by gene conversion events or recombination (Bjorkman & Parham, 1990; Parham *et al*., 1995). Yet, HLA polymorphism is a major barrier for bone marrow and solid organ transplantation. For this reason, HLA matching in transplantation of haematopoietic stem cells from unrelated donors has become increasingly complicated. On account of the high HLA polymorphism, a large donor pool is necessary to increase the chances of finding an HLA-matched unrelated donor. In 1999, 6 196 341 potential bone marrow donors and cord blood units were registered worldwide, of whom 1 603 644 were HLA-A, -B and -DR split typed, representing 414 765 different phenotypes (http://www.bmdw.org). Although this number may be sufficient to provide donors

Department of Transfusion Medicine, Hannover Medical School, Hannover, Germany.

Received 3 May 2001; revised 31 August 2001; accepted 26 September 2001

Correspondence: R. Blasczyk, Abteilung für Transfusionsmedizin, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, D-30625 Hannover, Germany. Tel: 49 511532 6700; Fax: 49 511532 2079; E-mail: blasczyk.rainer@mh-hannover.de

for the majority of patients bearing the most common HLA antigens, it may be impossible to find an HLA-compatible donor for carriers of rare antigens. When there is no genotypically identical sibling and there are several alternative potential donors that all have a mismatch at a relevant HLA locus, no ranking system exists that indicates different levels of allogenicity. To date, the impact of HLA mismatches on allogeneic T-cell activation has not been evaluated systematically because of the large number of $[n(n-1)]/2$ possible mismatches when there are n alleles to be compared with each other. The same limitation applies to clinical studies, which would have to include huge numbers of patients in order to cover all allelic mismatch combinations. Even *in vitro* studies will not solve this problem since they cannot cover the possible number of mismatched allele combinations, especially if a perfect allelic match at the other genes is desired. In order to find a theoretical approach to this problem and not to leave mismatching to chance, we propose a scoring system that is based on structural data of HLA class I molecules, and on data on naturally occurring frequencies of amino acid exchanges. This system considers not only the number of different amino acid residues but also their relevance with regard to antigen presentation, and the influence of their different chemical and physical properties. For demonstration, we employ our rating system for the comparison of the HLA-A*23 and A*24 groups, both of which allelic products are subdivisions of the serological HLA-A9 family, which is frequent in Caucasian (11.2%), African-American (13.2%), Asian-American (22.5%), Latin-American (16.2%) and Native-American (14.7%) populations (Mori *et al*., 1997).

Materials and methods

In the alleles A*2301–A*2303 and A*2402–A*2423, all positions assigned to the α1 and α2 domains (Bjorkman *et al*., 1987a,b, 1990) were examined for sequence differences by comparing all alleles to each other. The null alleles A*2409 and A*2411 N were excluded from the study. The allele A*2419 was also not considered, as the sequence has not yet been released. From crystallographic data on the structure of HLA class I molecules it was noted whether positions with polymorphism within the HLA-A9 family were likely to be important for peptide binding (Saper *et al*., 1991) or for recognition by the T-cell receptor (TCR) (Garboczi *et al*., 1996).

The possible influence of amino acid substitutions on functional properties of HLA-A molecules was rated using the distance matrix by Risler *et al*. (1988) based on amino acid substitutions in structurally related proteins. The basic idea of Risler's score was that two distinct amino acids are less dissimilar the more often they substitute each other in different proteins. Frequent substitutions represent high functional similarity, and rare substitutions represent low functional similarity. Accordingly, the less often members of a pair of amino acids substitute each other, thus representing functional dissimilarity, the higher is the score obtained, with a maximum value of 100. For rating structure differences within the HLA-A9 family a new weighted score ('dissimilarity index') was established, which considers both position and quality of substitution: to each amino acid disparity the value of Risler's distance matrix divided by 100 was assigned. When the disparity is located at a position that is important for peptide binding and/or recognition by the TCR, the value 1 was added in order to obtain a weighted score for this position. Residues that contribute to both functional properties were rated as if they possessed only one function, because there are no data indicating that their role is of major importance compared to positions that are important for only one function. The described algorithm results in higher scores even for low disparity at functionally relevant positions compared to any disparity at positions that are probably of functionally minor importance. In order to obtain the dissimilarity index, the scores representing the single amino acid substitutions were added up.

Results

Polymorphic positions in the α1 and α2 domains and assignment to the peptide binding region and the region important for recognition by the TCR are shown in Table 1. Table 2 (lower left section) shows the number of amino acid disparities within the HLA-A9 family. The extensive polymorphism within this group is revealed, with up to 17 residue disparities, 11 of them located at positions important for peptide binding or TCR recognition (A*2404 vs. A*2416). All expressed A9 variants differ from each other by at least one residue disparity in the α1 or α2 domain. Overall dissimilarity indices of amino acid comparisons are shown in Table 2 (upper right section).

The use of the dissimilarity index is illustrated in the following. For example, the alleles HLA-A*2402 and A*2405 differ by one residue (K vs. Q) at position 144, which is important neither for peptide binding nor for recognition by the TCR (Tables 1 and 2, lower left section). Lysine (K) and glutamine (Q) have been previously reported to substitute one another comparatively often, which has resulted in the score of 13 from Risler's matrix. In order to obtain the dissimilarity index, Risler's score is divided by 100.

In contrast, the alleles A*2402 and A*2406 differ by one residue (Q vs. W) located at position 158, which is likely to be important for recognition by the TCR. Glutamine and tryptophane (W) have been previously reported to substitute one another comparatively rarely, which has resulted in the score of 80 from Risler's matrix. In this case, the dissimilarity index is 1.80, as Risler's score is first divided by 100, and then 1 is added.

Besides the extensive polymorphism, Table 2 (upper right section) also demonstrates that most A*24 alleles show a greater difference to at least one other A*24 variant than to A*2301. Surprisingly, matching of A*2413 with A*2301 give a dissimilarity index of 1.77, which is lower than the indices yielded for comparisons between

Table 1. Variations in amino acid sequences of the α1 and α2 domains

Allele	1	3	9	62	65	70	76	79	80	81	82	83	95	97	99	105	107	114	116	127	144	151	152	156	158	161	163	166	167
A*2301	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	**Q**	**R**	V	**L**	A	E	T	D	G
A*2302	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	**Q**	**R**	V	**W**	A	E	T	D	G
A*2303	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	**N**	**Q**	**R**	V	**L**	A	E	T	D	G
A*2402	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	Q	A	E	T	D	G
A*2403	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	Q	A	E	T	**E**	**W**
A*2404	G	H	S	E	G	H	**A**	**G**	**T**	**L**	**R**	**G**	L	M	F	S	G	H	Y	K	K	H	V	Q	A	E	T	D	G
A*2405	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	**Q**	H	V	Q	A	E	T	D	G
A*2406	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	**W**	A	E	T	D	G
A*2407	G	H	S	E	G	**Q**	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	Q	A	E	T	D	G
A*2408	G	**Q**	S	**G**	**R**	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	Q	A	E	T	D	G
A*2410	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	Q	A	E	**R**	**E**	**W**
A*2413	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	**L**	A	E	T	D	G
A*2414	G	H	S	E	G	H	E	R	I	A	L	R	**V**	**R**	**Y**	S	**W**	H	Y	K	K	H	V	Q	A	E	T	D	G
A*2415	G	H	S	E	G	H	E	R	I	A	L	R	L	M	**Y**	**P**	G	H	Y	K	K	H	V	Q	A	E	T	D	G
A*2416	G	H	**T**	E	G	H	E	R	I	A	L	R	**I**	M	**Y**	S	G	**Q**	**D**	**N**	**Q**	**R**	V	**L**	A	E	T	**E**	**W**
A*2417	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	**R**	**D**	K	K	H	V	Q	A	E	T	D	G
A*2418	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	**E**	**L**	A	**D**	T	**E**	**W**
A*2420	G	**Q**	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	Q	A	E	T	D	G
A*2422	G	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	**W**	**V**	E	T	**E**	**W**
A*2423	**S**	H	S	E	G	H	E	R	I	A	L	R	L	M	F	S	G	H	Y	K	K	H	V	Q	A	E	T	D	**W**
Peptide binding	–	–	+	–	–	+	+	–	+	+	–	–	–	+	+	–	–	+	+	–	–	–	+	+	–	–	+	–	+
TCR binding	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	+	–	+	+	+

Numbers in the first line refer to the position within the HLA-A molecule. Amino acid residues are represented by the one-letter code. Bold text represents those residues that differ from the residue that is present at this position in the majority of HLA-A9 variants. + represents involvement of residues in TCR or peptide binding; – represents no involvement.

© 2002 Blackwell Science Ltd,

Table 2. Number of amino acid disparities (lower left section, under diagonal) and dissimilarity indices (upper right section, above diagonal)

Alleles			**2302**	**2303**	**2402**	**2403**	**2404**	**2405**	**2406**	**2407**	**2408**	**2410**	**2413**	**2414**	**2415**	**2416**	**2417**	**2418**	**2420**	**2422**	**2423**	
2302	1 (1)		1.76	0.31	3.04	6.21	7.82	2.91	3.53	4.72	5.71	7.29	1.77	6.28	4.70	8.91	6.34	6.39	3.72	7.76	5.28	**2301**
2303	1 (0)	2 (1)		2.07	3.57	6.74	8.35	3.44	1.77	5.25	6.24	7.82	3.53	6.81	5.23	10.67	6.87	8.15	4.25	6.00	5.81	**2302**
2402	3 (2)	3 (2)	4 (2)		3.35	6.52	8.13	3.22	3.84	5.03	6.02	7.60	2.08	6.59	5.01	8.60	6.65	6.70	4.03	8.07	5.59	**2303**
2403	5 (4)	5 (4)	6 (4)	2 (2)		3.17	4.78	0.13	1.80	1.68	2.67	4.25	1.27	3.24	1.66	11.95	3.30	5.89	0.68	6.03	2.24	2402
2404	9 (5)	9 (5)	10 (5)	6 (3)	8 (5)		7.95	3.30	4.97	4.85	5.84	1.08	4.44	6.41	4.83	8.78	6.47	2.72	3.85	2.86	1.67	**2403**
2405	2 (2)	2 (2)	3 (2)	1 (0)	3 (2)	7 (3)		4.91	6.58	6.46	7.45	9.03	6.05	8.02	6.44	16.73	8.08	10.67	5.46	10.81	7.02	**2404**
2406	3 (2)	2 (1)	4 (2)	1 (1)	3 (3)	7 (4)	2 (1)		1.93	1.81	2.80	4.38	1.40	3.37	1.79	11.82	3.43	6.02	0.81	6.16	2.37	**2405**
2407	4 (3)	4 (3)	5 (3)	1 (1)	3 (3)	7 (4)	2 (1)	2 (2)		3.48	4.47	6.05	1.76	5.04	3.46	12.44	5.10	6.38	2.48	4.23	4.04	**2406**
2408	6 (3)	6 (3)	7 (3)	3 (1)	5 (3)	9 (4)	4 (1)	4 (2)	4 (2)		4.35	5.93	2.95	4.92	3.34	13.63	4.98	7.57	2.36	7.71	3.92	**2407**
2410	6 (5)	6 (5)	7 (5)	3 (3)	1 (1)	9 (6)	4 (3)	4 (4)	4 (4)	6 (4)		6.92	3.94	5.91	4.33	14.62	5.97	8.56	1.99	8.70	4.91	**2408**
2413	2 (1)	3 (2)	3 (1)	1 (1)	3 (3)	7 (4)	2 (1)	1 (1)	2 (2)	4 (2)	4 (4)		5.52	7.49	5.91	9.86	7.55	3.80	4.93	3.94	2.75	**2410**
2414	7 (4)	7 (4)	8 (4)	4 (2)	6 (4)	10 (5)	5 (2)	5 (3)	5 (3)	7 (3)	7 (5)	5 (3)		4.51	2.93	10.68	4.57	4.62	1.95	5.99	3.51	**2413**
2415	5 (3)	5 (3)	6 (3)	2 (1)	4 (3)	8 (4)	3 (1)	3 (2)	3 (2)	5 (2)	5 (4)	3 (2)	4 (1)		2.82	13.04	6.54	9.13	3.92	9.27	5.48	2414
2416	8 (6)	9 (7)	7 (6)	11 (8)	9 (6)	17 (11)	10 (8)	11 (8)	12 (9)	14 (9)	10 (7)	10 (7)	12 (8)	11 (7)		11.53	4.96	7.55	2.34	7.69	3.90	**2415**
2417	5 (4)	5 (4)	6 (4)	2 (2)	4 (4)	8 (5)	3 (2)	3 (3)	3 (3)	5 (3)	5 (5)	3 (3)	6 (4)	4 (3)	10 (7)		9.66	8.96	12.63	10.33	10.45	**2416**
2418	6 (4)	7 (5)	7 (4)	5 (4)	3 (2)	11 (7)	6 (4)	5 (4)	6 (5)	8 (5)	4 (3)	4 (3)	9 (6)	7 (5)	10 (6)	7 (6)		9.19	3.98	9.33	5.54	2417
2420	4 (2)	4 (2)	5 (2)	1 (0)	3 (2)	7 (3)	2 (0)	2 (1)	2 (1)	2 (1)	4 (3)	2 (1)	5 (2)	3 (1)	12 (8)	3 (2)	6 (4)		6.57	4.27	4.39	**2418**
2422	6 (5)	5 (4)	7 (5)	4 (4)	2 (2)	10 (7)	5 (4)	3 (3)	5 (5)	7 (5)	3 (3)	4 (4)	8 (6)	6 (5)	10 (7)	6 (6)	4 (3)	5 (4)		6.71	2.92	**2420**
2423	5 (3)	5 (3)	6 (3)	2 (1)	2 (1)	8 (4)	3 (1)	3 (2)	3 (2)	5 (2)	3 (2)	3 (2)	6 (3)	4 (2)	11 (7)	4 (3)	5 (3)	3 (1)	4 (3)		4.53	**2422**
	2301	**2302**	**2303**	**2402**	**2403**	**2404**	**2405**	**2406**	**2407**	**2408**	**2410**	**2413**	**2414**	**2415**	**2416**	**2417**	**2418**	**2420**	**2422**			

In the lower left section, the first number in each square represents the overall number of disparities, and the number in brackets represents differences with regard to residues that are likely to interact with presented peptides or to be recognized by the T-cell receptor.

© 2002 Blackwell Science Ltd,

A*2413 and A*2403, A*2404, A*2407, A*2408, A*2410, A*2414, A*2415, A*2416, A*2417, A*2418, A*2420, A*2422 and A*2423, respectively.

Discussion

HLA class I molecules are heterodimeric cell surface glycoproteins that present antigenic peptides to CD8+ T cells and thus are crucial for the specificity of cellular immunity. The extracellular portion of the heavy chain has been divided into three domains called α1, α2 and α3 that are followed by a membrane-spanning region and a cytoplasmic tail. The α3 domain shows little within-group polymorphism and is important for interaction with the CD8 glycoprotein of the T cell. The peptide-binding site is formed by amino acid residues that belong to the α1 and α2 domains (Bjorkman *et al*., 1987a,b; Saper *et al*., 1991). Residues on the α-helices with side-chains pointing 'up' from the surface are important for direct recognition by the TCR (Garboczi *et al*., 1996). Residues located on the α-helices with side-chains in an intermediate position could interact with either peptides or TCR (Bjorkman *et al*., 1987a,b; Saper *et al*., 1991; Garboczi *et al*., 1996). Differences in the amino acid alignments of HLA class I molecules are predominantly found in the α1 and α2 domains, many of them in residues that line the sides and the bottom of the peptide-binding groove. This finding correlates with the observed specificity in peptide binding to individual HLA class I molecules (Bjorkman *et al*., 1987a,b, 1990; Garrett *et al*., 1989; Saper *et al*., 1991).

Crystallographic data on HLA-A2 and HLA-A68 revealed that the amino acid residues involved in peptide binding can be assigned to six specificity pockets (Garrett *et al*., 1989; Saper *et al*., 1991), of which pockets B and F, which bind P2 and P9 of the presented peptide, have been demonstrated to be crucial for determination of the allelic peptide motif (Falk *et al*., 1991; Parker *et al*., 1992; Carreno *et al*., 1993; Rammensee *et al*., 1993).

Variations in the peptides bound by related HLA class I molecules have been reported (Rötzschke *et al*., 1992). However, peptide-binding specificities of closely related HLA class I molecules can be strikingly similar (Sidney *et al*., 1996). Corresponding to crystallographic data (Bjorkman *et al*., 1987a,b, 1990; Garrett *et al*., 1989; Saper *et al*., 1991) these findings indicate that, rather than genetic relatedness itself, the position and type of amino acid exchanges between HLA molecules are pivotal for the peptides preferentially bound.

Allogeneic recognition by CD8+ T cells is thought to be brought about by different mechanisms. The high determinant density hypothesis (Bevan, 1984) proposes that the receptor of the alloreactive T cell is specific for the foreign HLA molecule itself, irrespective of the peptides bound (if present at all), which may be true in some cases, as experimental data suggest (Smith *et al*., 1997). However, most studies suggest the involvement of the presented peptide. Experimental data suggest that alloreactive T-cell clones recognize unique peptides in the context of allogeneic HLA class I molecule (Heath *et al*., 1991), conformational changes of presented peptides (Herman *et al*., 1999), or peptide-induced conformational changes of the HLA molecule (Bluestone *et al*., 1993), whereby subtle changes in the peptide environment may impact alloreactivity (Rudolph *et al*., 2001).

Yet, only in the most extreme cases might it be possible to show that either peptide or MHC plays a more dominant role in a TCR–allo–MHC interaction (Kranz, 2000).

In order to approach the clinical problem of finding the best-fitting donor in unrelated bone marrow transplantation, we propose a scoring system based on knowledge of allogeneic recognition and the structure of HLA class I molecules. This dissimilarity index considers both position and quality of residue disparities between compared alleles. The possible influence of the exchange of amino acid residues is rated on the basis of Risler's distance matrix, supposing that the more similar two amino acids are rated to be, the smaller is the impact on peptide binding and/or recognition by the TCR. This algorithm results in higher dissimilarity indices even for low disparity at functionally relevant positions compared to any disparity at positions that are probably of functionally minor importance. Using this method, remarkable differences between HLA-A9 subtypes were revealed, as shown in Table 2 (upper right section).

When possible, the donor carrying the allele with the lowest dissimilarity index may be preferred. As stated above, comparison of the alleles A*2402 and A*2405 yields the dissimilarity index 0.13, whereas comparison of A*2402 and A*2406 gives 1.80. Accordingly, when there is a patient carrying A*2402, a potential bone marrow donor carrying HLA-A*2405 who differs from the patient only by this single allele may be preferred to a HLA-A*2406-carrying donor. Although both A*2405 and A*2406 differ from A*2402 by one amino acid each, the dissimilarity index enables a similarity analysis that is superior to simply counting the number of differences.

Table 2 (upper right section) also reveals that most A*24 alleles show a greater difference to at least one other A*24 variant than to A*2301. Particularly, this observation may be of clinical significance in the mismatching of A*2413 with A*2301, which gives a dissimilarity index of 1.77, which is lower than the indices obtained for comparisons between A*2413 and most other A*24 alleles. This difference between functional similarity and serological relatedness may be explained by an overriding influence of distinct epitopes of the HLA class I molecule on serological recognition that only in part reflects the genetic (Hildebrand *et al*., 1992) or functional relationship of alleles. Yet, it has to be considered that names of alleles do not always reflect their relationship since the serological assignment of some alleles is unlikely to be known with certainty.

The dissimilarity index provides a simple tool for comparing similarity between different allelic combinations. However, some limitations have to be considered. In order to distinguish between residues of major (peptide-binding site and/or contact with the TCR) or minor (remaining residues of the α1 and α2 domains) importance, crystallographic

data obtained from HLA-A2 (Saper *et al*., 1991; Garboczi *et al*., 1996) were used, as there are no data on the three-dimensional structure of A9 molecules. Although previous data suggest a great structural similarity between all HLA-A molecules (Bjorkman *et al*., 1987a,b; Saper *et al*., 1991; Garboczi *et al*., 1996), there may be some minor differences from A9 which may not be reflected correctly by the dissimilarity index. Furthermore, in order not to complicate the index, peptide-binding regions and regions that have contact with the TCR were regarded as equally important. This procedure also avoids the problem of assessment of positions that can be allocated both to the peptide-binding regions and the regions important for contact with the TCR. The dissimilarity index does not weight the assignment of residues to pockets B and F as the binding site is a continuous surface, and designating the beginning and end of such pockets is somewhat arbitrary (D. C. Wiley, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MAUSA, personal communication). This suggestion is supported by the finding that single residue exchanges at positions that cannot be assigned to pockets B or F may also lead to strong alloreactivity (Herman *et al*., 1999).

The Bw4/Bw6 epitope has been reported to strongly influence binding of antigenic peptides (Hill *et al*., 1992). Accordingly, the absence of the Bw4 epitope in A*2404 (Kashiwase *et al*., 1995), which results in the exchange of the residues 76, 79, and 80–83 (Table 1), is likely to enhance allogenicity in the case of a mismatch with other A*24 alleles. In order not to complicate the dissimilarity index proposed here, the influence of residues on allogenicity was rated regardless of possible allocation to the Bw4/Bw6 epitope. Since the exchange of this epitope always comprises several residues, some of them involved in peptide binding (Saper *et al*., 1991), comparisons between alleles that differ by Bw4/Bw6 will always yield high dissimilarity indices, as shown here for A*2404 (Table 2, upper right section).

In the algorithm described here, different residues at functionally relevant positions always yield higher scores, even for low disparity, compared to any disparity at positions that are probably of functionally minor importance. In some cases, the relation between functionally important and less important regions of the HLA molecule thus may be not reflected linearly. For this reason, the dissimilarity index may over- or underestimate exchanges in regions important for peptide binding or T-cell attachment compared to changes in other regions of the HLA molecule. Risler's score, which has been used in order to rate the impact of amino acid exchanges on peptide binding, is based on frequencies of naturally occurring amino acid substitutions in structurally related proteins. Similarities between distinct amino acids in this score often, but not always, correlate with chemical or physical properties (Grantham, 1974; Risler *et al*., 1988; Stanfel, 1996). We preferred Risler's score to other matrices as it reflects functional similarity or disparity in the context of complete proteins, whereas other scores only represent classification on the basis of physical and chemical properties of the single components (Grantham, 1974; Stanfel, 1996).

Evidence that sequence similarity matching may be a useful strategy for bone marrow transplantation has come from previous studies on renal allograft transplantation. Takemoto & Terasaki (1996) proposed a matching procedure for kidney transplantation based on the three polymorphic positions of HLA molecules which were shown to be oriented upwards to the TCR (Bjorkman *et al*., 1987a,b). This procedure yielded very similar results compared to conventional HLA matching. The probable correlation between mismatch of functionally important amino acid residues of HLA class II molecules and outcome of renal transplantation was also analysed in the study by Sada *et al*. (1997). They found that rejection occurrence correlated with incompatible DRB1 amino acid residues on the β-pleated sheet.

In addition to our scoring system, a sophisticated mathematical model for comparing HLA alleles has been proposed by Cano *et al*. (1998), in which data on peptide binding by different HLA class I alleles were used to calculate an empirical similarity matrix between HLA alleles. This model may be a very interesting approach for sequence similarity matching in bone marrow transplantation. However, it is based only on peptide binding, and does not consider recognition of superficial residues of the HLA molecule by the TCR which is also of importance for immunotolerance in bone marrow transplantation. Nevertheless, this proposed model deserves further attention and correlation to clinical data may be promising.

By using the dissimilarity index proposed here, remarkable differences between the HLA-A9 subtypes were revealed, which are superior to a simple sequence comparison, as both the function of the residues and the quality of the exchange are considered. Within the limitations discussed above, new scoring matrices can also easily be established for other HLA genes and allelic groups. For HLA-DR and -DQ, a crystallographic analysis of the TCR–peptide–MHC complex still needs to be performed, so that transfer of the algorithm described above to HLA class II will be somewhat limited.

Because of the large number of possible mismatches, only a small fraction can be covered by systematic *in vitro* experiments and by clinical studies. Furthermore, in addition to the HLA match the outcome of unrelated-donor bone marrow transplantation is dependent on other factors such as cell dose, age, gender, cytomegalovirus status of the recipient, and ethnic group (Davies *et al*., 2000). Accordingly, the HLA match is only one player in the donor selection algorithm, and the scoring system proposed here may help to weight the immunogenetic variable in the selection procedure. Studies focusing on the systematic analysis of a reduced number of mismatches may be helpful in order to further improve the theoretical model proposed here. However, as long as such studies are not available, present knowledge of the structure and function of HLA molecules, as in part reflected by the dissimilarity index, may be utilized in order to prevent HLA mismatching being a matter of chance.

Acknowledgements

The critical advice of Don C. Wiley, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA, is gratefully acknowledged.

References

Bevan, M.J. (1984) High determinant density may explain the phenomenon of alloreactivity. *Immunology Today*, **5**, 128.

Bjorkman, P.J. & Parham, P. (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. *Annual Review of Biochemistry*, **59**, 253.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. (1987a) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. (1987b) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*, **329**, 506.

Bluestone, J.A., Kaliyaperumal, A., Jameson, S., Miller, S. & Dick, R.I.I. (1993) Peptide-induced changes in class I heavy chains alter allorecognition. *Journal of Immunology*, **151**, 3943.

Cano, P., Fan, B. & Stass, S. (1998) A geometric study of the amino acid sequence of class I HLA molecules. *Immunogenetics*, **48**, 324.

Carreno, B.M., Winter, C.C., Taurog, J.D., Hansen, T.H. & Biddison, W.E. (1993) Residues in pocket B and F of HLA-B27 are critical in the presentation of an influenza A virus nucleoprotein peptide and influence the stability of peptide–MHC complex. *International Immunology*, **5**, 353.

Davies, S.M., Kollman, C., Anasetti, C., Antin, J.H., Gajewski, J., Casper, J.T., Nademanee, A., Noreen, H., King, R., Confer, D. *et al.* (2000) Engraftment and survival after unrelated-donor bone marrow transplantation: a report from the national marrow donor program. *Blood*, **96**, 4096.

Falk, K., Rötzschke, O., Stevanovic, S., Jung, G. & Rammensee, H.-G. (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature*, **351**, 290.

Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E. & Wiley, D.C. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature*, **384**, 134.

Garrett, T.P.J., Saper, M.A., Bjorkman, P.J., Strominger, J.L. & Wiley, D.C. (1989) Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature*, **342**, 692.

Grantham, R. (1974) Amino acid difference formula to help explain protein evolution. *Science*, **185**, 862.

Heath, W.R., Kane, K.P., Mescher, M.F. & Sherman, L.A. (1991) Alloreactive T cells discriminate among a diverse set of endogenous peptides. *Proceedings of the National Academy of Science USA*, **88**, 5101.

Herman, J., Jongeneel, V., Kuznetsov, D. & Coulie, P.G. (1999) Differences in the recognition by CTL of peptides presented by the HLA-B*4402 and the HLA-B*4403 molecules which differ by a single amino acid. *Tissue Antigens*, **53**, 111.

Hildebrand, W.H., Madrigal, J.A., Belich, M.P., Zemmour, J., Ward, F.E., Williams, R.C. & Parham, P. (1992) Serologic cross-reactivities poorly reflect allelic relationships in the HLA-B12 and HLA-B21 groups. Dominant epitopes of the α_2 helix. *Journal of Immunology*, **149**, 3563.

Hill, A.V., Elvin, J., Willis, A.C., Aidoo, M., Allsopp, C.E., Gotch, F.M., Gao, X.M., Takiguchi, M., Greenwood, B.M., Townsend, A.R. *et al.* (1992) Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature*, **360**, 434.

Kashiwase, K., Tokunaga, K., Ishikawa, Y., Lin, L., Imanishi, T., Akaza, T., Tadokoro, K. & Juji, T. (1995) A new HLA-A9 subtype lacking the Bw4 epitope. Ancestral or revertant allele? *Human Immunology*, **42**, 221.

Kranz, D.M. (2000) Incompatible differences: view of an allogeneic pMHC-TCR complex. *Nature Immunology*, **1**, 277.

Madrigal, J.A., Scott, I., Arguello, R., Szydlo, R., Little, A.-M. & Goldman, J.M. (1997) Factors influencing the outcome of bone marrow transplants using unrelated donors. *Immunological Reviews*, **157**, 153.

Marsh, S.G.E., Bodmer, J.G., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Hansen, J.A., Mach, B., Mayr, W.R. *et al.* (2001) Nomenclature for factors of the HLA system, 2000. *European Journal of Immunogenetics*, **28**, 377.

Mickelson, E.M., Petersdorf, E., Anasetti, C., Martin, P., Woolfrey, A. & Hansen, J.A. (2000) HLA matching in hematopoietic cell transplantation. *Human Immunology*, **61**, 92.

Mori, M., Beatty, P.G., Graves, M., Boucher, K.M. & Milford, E.L. (1997) HLA gene and haplotype frequencies in the North American population. *Transplantation*, **64**, 1017.

Parham, P., Adams, E.J. & Arnett, K.L. (1995) The origins of HLA-A, B, C polymorphism. *Immunological Reviews*, **143**, 141.

Parker, K.C., Bednarek, M.A., Hull, L.K., Utz, U., Cunningham, B., Zweerink, H.J., Biddison, W.E. & Coligan, J.E. (1992) Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. *Journal of Immunology*, **149**, 3580.

Petersdorf, E.W., Gooley, T.A., Anasetti, C., Martin, P.J., Smith, A.G., Mickelson, E.M., Woolfrey, A.E. & Hansen, J.A. (1998) Optimizing outcome after unrelated marrow transplantation by comprehensive matching of HLA class I and class II alleles in the donor and recipient. *Blood*, **92**, 3515.

Rammensee, H.-G., Falk, K. & Rötzschke, O. (1993) Peptides naturally presented by MHC class I molecules. *Annual Review of Immunology*, **11**, 213.

Risler, J.L., Delorme, M.O., Delacroix, H. & Henaut, A. (1988) Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *Journal of Molecular Biology*, **204**, 1019.

Rötzschke, O., Falk, K., Stevanovic, S., Jung, G. & Rammensee, H.-G. (1992) Peptide motifs of closely related HLA class I molecules encompass substantial differences. *European Journal of Immunology*, **22**, 2453.

Rudolph, M.G., Speir, J.A., Brunmark, A., Mattsson, N., Jackson, M.R., Peterson, P.A., Teyton, L. & Wilson, I.A. (2001) The crystal structures of Kbm1 and Kbm8 reveal that subtle changes in the peptide environment impact thermostability and alloreactivity. *Immunity*, **14**, 231.

Sada, M., Takahara, S., Tada, M., Hatori, M., Wang, J.D., Okuyama. A. & Tsuji, T. (1997) Importance of HLA-DRB1 molecular matching based on three-dimensional structure in cadaveric renal transplantation: a trial of new criteria for recipient and donor selection. *Transplantation Proceedings*, **29**, 1440.

Saper, M.A., Bjorkman, P.J. & Wiley, D.C. (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *Journal of Molecular Biology*, **219**, 277.

Sasazuki, T., Juji, T., Morishima, Y., Kinukawa, N., Kashiwabara, H., Inoko, H., Yoshida, T., Kimura, A., Akaza, T., Kamikawaji, N. *et al.* (1998) Effect of matching of class I HLA alleles on clinical

outcome after transplantation of hematopoietic stem cells from an unrelated donor. *New England Journal of Medicine*, **17**, 1177.

Sidney, J., Grey, H.M., Southwood, S., Celis, E., Wentworth, P.A., del Guercio, M.-F., Kubo, R.T., Chesnut, R.W. & Sette, A. (1996) Definition of an HLA-A3-like supermotif demonstrates the overlapping peptide-binding repertoires of common HLA molecules. *Human Immunology*, **45**, 79.

Smith, P.A., Brunmark, A., Jackson, M.R. & Potter, T.A. (1997) Peptide-independent recognition by alloreactive cytotoxic T lymphocytes (CTL). *Journal of Experimental Medicine*, **185**, 1023.

Stanfel, L.E. (1996) A new approach to clustering the amino acids. *Journal of Theoretical Biology*, **183**, 195.

Takemoto, S. & Terasaki, P.I. (1996) HLA compatibility can be predicted by matching only three residues with outward orientated side chains. *Transplantation Proceedings*, **28**, 1264.