Possible Assortment of a1 and a2 Region Gene Segments in Human MHC Class I Molecules

George Johnson and Tai T. Wu

Departments of Biochemistry, Molecular Biology and Cell Biology, and Biomedical Engineering, Northwestern University, Evanston, Illinois 60208

Manuscript received August 5, 1997
Accepted for publication January 26, 1998

ABSTRACT

Using pair-wise comparison of aligned nucleotide sequences of distinct and complete human MHC class I molecules, we have constructed triangular tables to study the similarities and differences of various a1 (exon 2) and a2 (exon 3) region sequences. There are two *HLA-A* (A*6901 and A*6601) and 13 *HLA-B* (B*4201, B*8101, B*4102, B*4801, B*4007, B*4001, B*4802, Dw53, B*4406, B*4402, B*3901, B*1514 and B*3702) sequences that have identical a1 sequences with other known MHC class I molecules, while their a2 sequences are the same as those of different ones. Of these 15, A*6901, B*4001 and B*4802 have previously been suggested as the results of recombination between A*6801 and A*0201, B*4101 and B*8101, and B*4801 and B*3501, respectively. However, many other sequences can also be used to generate them by recombination. Furthermore, their reciprocal products have never been identified. Thus, gene conversion has subsequently been suggested as an alternative. Another possible genetic mechanism for generating these nucleotide sequence similarities can be assortment, or that some gene segments can be duplicated or multiplicated to be used in different human MHC class I molecules. Interestingly, this genetic mechanism is probably absent for the generation of different mouse MHC class I molecules.

THE nucleotide sequences of human major histocompatibility complex (MHC) class I molecules have been extensively collected by Parham and his collaborators (Arnett and Parham 1995). They have noticed that the a1 region nucleotide sequences of A*6901 and A*6801 are identical, and A*6901 and A*0201 have the same a2 region nucleotide sequence. Therefore, they initially suggested that A*6901 might be the result of recombination of A*6801 and A*0201 molecules (Holmes and Parham 1985). Because the reciprocal product has not been found, and because the differences between A*6901 and A*6801 sequences in their a2 regions are localized to a relatively short segment, Hemmi *et al.* (1988) and Parham *et al.* (1995) have subsequently suggested that gene conversion may be a more likely mechanism.

Hughes *et al.* (1993) have found two similar examples for *HLA-B* molecules: B*4001 being the product of B*4101 and B*8101 recombination, and B*4802 being that of B*4801 and B*3501 with one nucleotide difference.

A thorough pairwise comparison of these sequences was initiated by Parham *et al.* (1995), and they noticed the following: "In comparing pairs of *HLA-A,B,C* alleles, only 2 pairs out of a total of 6,400 cannot be distinguished on the basis of sequences in exons 2, 3 and 4. . . . In contrast, when comparison is restricted to either exon 2 or exon 3 alone then the number of ambiguous pairs increases significantly." They also made plots of occurrences *vs.* differences to separate interlocus and intralocus comparisons.

We have been using similar methods to study human and mouse immunoglobulin V-gene nucleotide sequences (Johnson and Wu 1997a) and possible evolutionary differences between T-cell receptor for antigen and immunoglobulin V-genes (Johnson and Wu 1997b). In addition, we have made use of the triangular tables that list the differences of pairwise comparisons.

In this study, triangular tables are constructed for human and mouse MHC class I a1 (exon 2) and a2 (exon 3) region gene segments. The combined tables of a1/a2 region nucleotide sequence differences are then used for detailed analysis.

MATERIALS AND METHODS

The sequence database has been previously described (Johnson and Wu 1997a). Briefly, the aligned sequences of human and mouse MHC class I a1 (exon 2) and a2 (exon 3) regions were obtained through the Kabat database WWW server (Johnson *et al.* 1996; http://immuno.bme.nwu.edu), which has been maintained in our laboratory. This is the only database with aligned nucleotide and amino acid sequences of proteins of immunological interest.

Triangular tables were generated, listing the number of nucleotide differences between any two complete and distinct sequences of human or mouse MHC class I chains for a1 and a2 regions separately. As shown before (Johnson and Wu

Corresponding author: Tai Te Wu, Room E267, Technological Institute, 2145 Sheridan Rd., Northwestern University, Evanston, IL 60208. E-mail: tt@immuno.bme.nwu.edu

TABLE 1

Triangular table listing pairwise differences for a1/a2 region sequences of 12 *HLA-A* molecules

	A*6901	A*6801	A*6602	A*6802	A*0206	A*0211	A*0201	A*0204	A*6601	A*2501	A*2603	A*2601
A*6901	0	0 7	0 16	1 4	9 0	11 0	11 0	11 1	1 16	11 16	3 16	4 16
A*6801		0	0 13	1 5	9 7	11 7	11 7	11 6	1 13	11 13	3 13	4 13
A*6602			0	1 12	9 16	11 16	11 16	11 17	1 2	11 2	3 2	4 2
A*6802				0	10 4	12 4	12 4	12 5	2 12	12 12	4 12	5 12
A*0206					0	4 0	2 0	2 1	10 16	18 16	8 16	11 16
A*0211						0	2 0	2 1	12 16	20 16	12 16	13 16
A*0201							0	0 1	12 16	20 16	10 16	13 16
A*0204								0	12 17	20 17	10 17	13 17
A*6601									0	10 0	2 0	3 0
A*2501										0	10 0	8 0
A*2603											0	3 0
A*2601												0

TABLE 2

Triangular table listing pairwise differences for a1/a2 region sequences of 15 *HLA-B* molecules

	B*4201	B*8101	B*0704	B*0705	B′′′	B*0801	B*4102	B*4801	B*4007	B*4001	B*4101	B*4501	B*5001	B*4505	B*4802
B*4201	0	0 10	0 7	0 9	0 9	10 1	19 0	8 10	19 11	19 11	19 6	19 17	19 20	17 10	8 16
B*8101		0	0 7	0 4	0 6	10 9	19 10	8 0	19 1	19 1	19 16	19 25	19 22	17 8	8 18
B*0704			0	0 4	0 4	10 6	19 7	8 7	19 8	19 8	19 13	19 22	19 23	17 9	8 19
B*0705				0	0 2	10 8	19 9	8 4	19 5	19 5	19 15	19 24	19 20	17 6	8 18
B7′′′					0	10 10	19 9	8 6	19 7	19 7	19 15	19 24	19 20	17 8	8 16
B*0801						0	13 1	9 9	12 10	13 10	13 7	13 18	13 21	15 9	9 17
B*4102							0	11 10	1 11	0 11	0 6	0 17	0 20	2 10	11 16
B*4801								0	12 1	11 1	11 16	11 25	11 22	9 8	0 18
B*4007									0	1 0	1 15	1 24	1 21	3 9	12 17
B*4001										0	0 15	0 24	0 21	2 9	11 17
B*4101											0	0 11	0 14	2 16	11 17
B*4501												0	0 5	2 19	11 13
B*5001													0	2 14	11 10
B*4505														0	9 12
B*4802															0

TABLE 3

Triangular table listing pair-wise differences for a1/a2 region sequences of five *H-2 D* molecules

	D-Bm24	D-Bm14	Dg7	D-Bm13	D-B
D-Bm24	0	7	9	8	9
		0	0	4	1
D-Bm14		0	2	1	2
			0	4	1
Dg7			0	1	0
				4	1
D-Bm13				0	1
					5
D-B					0

1997a), these tables were very large and difficult to print in a journal article. If interested, please write to the authors for copies. Combined triangular tables for a1/a2 regions, listing the difference of a1 region sequences above those of a2 region sequences, were then constructed for three examples: one for 12 human *HLA-A* sequences (Table 1), one for 15 human *HLA-B* sequences (Table 2), and one for 5 mouse *H-2* sequences (Table 3).

RESULTS

Table 1 illustrates the nucleotide differences of some *HLA-A* sequences. A*6901 (Holmes and Parham 1985) has an a1 region sequence identical to those of A*6801 and A*6602, and differing by only one nucleotide from the A*6601 a1 sequence, but its a2 region sequence differs by 7, 16, and 16 nucleotides, respectively. On the other hand, A*6901 has an a2 region sequence identical to those of A*0206, A*0211, and A*0201, and differing by only one nucleotide from the A*0204 a2 sequence, but its a1 region sequence differs by 9, 11, 11, and 11 nucleotides, respectively. This finding is summarized as one item in Table 4, where A*6901 is connected to the first row of identical or nearly identical a1 region sequences and to the second row of identical or nearly identical a2 region sequences. This example was initially reported by Holmes and Parham (1985), with A*6901 being the result of recombination between A*6801 and A*0201. As illustrated in Table 1, A*6901 and A*6801 differ by only 7 nucleotides, most of which are localized in a relatively short segment. Hemmi *et al.* (1988) and Parham *et al.* (1995) have subsequently proposed that gene conversion can be another mechanism for generating new MHC class I molecules.

A similar situation is observed for some of the *HLA-B* sequences (Table 2). B*4201 has an a1 region sequence identical to those of B*8101, B*0704, B*0705, and B7′′′, but its a2 sequence differs by 10, 7, 9, and 9 nucleotides, respectively. On the other hand, B*4201 has an a2 region sequence identical to that of B*4102 and differs from that of B*0801 by 1 nucleotide, but its a1 region sequence differs by 19 and 10 nucleotides, respectively. This is summarized in the third line of Table 4.

TABLE 4

Listing of 2 *HLA-A* and 13 *HLA-B* sequences that share identical or nearly identical a1 sequences with one set of other molecules and a2 sequences with another set

A*6901	A*6801, A*6602, A*6601
	A*0206, A*0211, A*0201, A*0204
A*6601	A*6901
	A*2501
B*4201	B*8101, B*0704, B*0705, B7′′′
	B*0801, B*4102
B*8101	B*4201, B*0704, B7′′′
	B*0801, B*4007, B*4001
B*4102	B*4007, B*4001, B*4101, B*4501, B*5001
	B*4201, B*0801
B*4801	B*4802
	B*8101, B*4007, B*4001
B*4007	B*4101, B*4501, B*5001, B*4102
	B*8101, B*4801
B*4001	B*4102, B*4101, B*4501, B*5001
	B*8101, B*4801
B*4802	B*4801
	B*3501, B*3508, B*3507, Dw53, B*5801
Dw53	B*4406, B*5103
	B*4801, B*5801
B*4406	Dw53, B*5103, B*5104
	B*4402, B*4401
B*4402	B*4404
	B*4406
B*3901	B*3909, B*1401, B*39061, B*1402
	B*3801
B*1514	B*1504
	B*4601
B*3702	B*3701
	B*2701, B*2704, B*2702

As illustrated in Table 4, there are 2 *HLA-A* and 13 *HLA-B* molecules with such shared sequence similarities. The circled ones, namely, A*6901, B*4001, and B*4802, have previously been investigated by Holmes and Parham (1985) and Hughes *et al.* (1993). They have suggested only 1 pair of other sequences to generate each of the above 3 sequences. In fact, for all the 15 sequences listed in Table 4, there are a total of 83 possible pairs of other sequences that may be involved.

For mouse MHC class I sequences (Table 3), however, the nucleotide sequence differences among D-Bm24, D-Bm14 (Hemmi *et al.* 1988), and Dg7 (Girgis *et al.* 1996) can be explained by gene conversion (Hemmi *et al.* 1988; Parham *et al.* 1995). Their a2 region sequences

are identical, and their a1 region sequences differ by two, seven, and nine nucleotides.

DISCUSSION

A*6901 was considered to be the result of intra-allelic reciprocal recombination between A*6801 and A*0201 by Holmes and Parham (1985). They pointed out that this can be a useful mechanism for generating novel MHC class I molecules. As shown in Table 1, A*6901 and A*6801 are identical in their a1 regions and differ by seven nucleotides in their a2 regions, most of which are localized in a short segment. For this reason, these authors subsequently favored the idea of gene conversion as a more likely mechanism of generating such differences (Parham *et al.* 1995). On the other hand, A*6901 can be derived from A*0201 and A*6602 or other pairs (Tables 1 and 4), where the nucleotide differences are distributed over the entire regions of a1 and a2, rather than localized.

The other molecule, consisting of the a1 region of A*0201 and the a2 region of A*6801, expected as a result of reciprocal recombination, is, however, not found. For *HLA-B* molecules, the two examples found by Hughes *et al.* (1993) have similar problems. Indeed, our simple method of using triangular tables to list the pairwise nucleotide differences of a1/a2 regions have uncovered many other examples, as illustrated in Table 4. To account for all of these, a simpler mechanism of gene assortment may, therefore, be needed. Alternatively, the same a1 or a2 region gene segment can be duplicated or multiplicated to be used in different MHC class I molecules. Indeed, gene fragments of human MHC class I molecules have previously been observed (Geraghty *et al.* 1992).

If assortment of a1 and a2 region genes, that is, exons 2 and 3, respectively, occurs in human MHC class I molecules, the logical extension is that this genetic mechanism should be operative in mouse MHC class I molecules also. However, as we have noted, the only similar situation was found for three mouse sequences: D-Bm14, D-Bm24, and Dg7 (Table 3). Their differences are localized in the segment from codons 63 to 80 of the a1 region. As pointed out by Hemmi *et al.* (1988) and by Parham *et al.* (1995), this can be the result of gene conversion.

In conclusion, on the basis of pairwise comparison of a1 (exon 2) and a2 (exon 3) region nucleotide sequences of human MHC class I molecules, in addition to reciprocal recombination and gene conversion, assortment of the entire a1 or a2 region can be a versatile mechanism for generating novel class I molecules in the human system. However, this genetic process does not seem to be important in the mouse system, although gene conversion may play a role.

This research was supported in part by National Institutes of Health grant 5 R01 AI-25616-08.

LITERATURE CITED

Arnett, K. L., and P. Parham, 1995 HLA class I nucleotide sequences. Tissue Antigens **46:** 217–257.

Geraghty, D. E., B. H. Koller, J. A. Hansen and H. T. Orr, 1992 The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. J. Immunol. **149:** 1934–1946.

Girgis, K. R., J. D. Capra and I. Stroynowski, 1996 Nucleotide sequences of H2g7 K and D loci of non-obese diabetic mice. Immunogenetics **41:** 386.

Hemmi, S., J. Geliebter, R. A. Jeff, R. W. Melvold and S. G. Nathenson, 1988 Three spontaneous H-2Db mutants are generated by genetic micro-recombination (gene conversion) events. Impact on the H-2-restricted immune responsivenesses. J. Exp. Med. **168:** 2319–2335.

Holmes, N., and P. Parham, 1985 Exon shuffling *in vivo* can generate novel HLA class I molecules. EMBO J. **4:** 2849–2854.

Hughes, A. L., M. K. Hughes and D. I. Watkins, 1993 Contrasting roles in interallelic recombination at the *HLA-A* and *HLA-B* loci. Genetics **133:** 669–680.

Johnson, G., and T. T. Wu, 1997a A method of estimating the numbers of human and mouse immunoglobulin V-genes. Genetics **145:** 777–786.

Johnson, G., and T. T. Wu, 1997b Profile of numbers of sequence differences among V-genes coding for the variable regions of T cell receptor for antigen alpha and beta chains. J. Mol. Evol. **44:** 253–257.

Johnson, G., E. A. Kabat and T. T. Wu, 1996 Kabat database of proteins of immunological interest, pp. 6.1–6.21 in *Weir's Handbook of Experimental Immunology. I. Immunochemistry and Molecular Immunology*, edited by L. A. Herzenberg, W. M. Weir, L. A. Herzenberg and C. Blackwell. Blackwell Science, Cambridge, MA.

Parham, P., E. J. Adams and K. L. Arnett, 1995 The origin of HLA-A,B,C polymorphism. Immunol. Rev. **143:** 141–180.

Communicating editor: S. L. Allen