Variability and Conformation of HLA Class I Antigens: A Predictive Approach to the Spatial Arrangement of Polymorphic Regions†

Miguel A. Vega, Rafael Bragado, Angel Ezquerra, and José A. López de Castro*

ABSTRACT: Comparison of available sequences of HLA-A and HLA-B antigens shows that variable positions are predominantly localized in four segments spanning residues 63–85, 105–116, 138–156, and 177–194. The fourth segment is unique in that it contains no differences between antigens of the same locus. Secondary folding of HLA heavy chain was estimated by three independent predictive methods and areas of defined structure were correlated with the distribution of local hydrophobicity to outline putative internal and external portions. The three analyses each independently predict a high probability for β structure in the α1, α2, and α3 domains. A single α-helix is predicted within residues 146–160, a segment of likely importance in cytotoxic T cell recognition and graft rejection. Substitutions within this segment are spatially related by the helical turn. Variable residues usually lie in areas of high local hydrophilicity, and therefore they are probably on the surface of the molecule. The model predicts that they are frequently located in β strands, β-turns, or the above-mentioned α-helix, so that most substitutions would be accommodated within rigid frameworks that may impose structural constraints to variability. The secondary structure of α1, α2, and α3 domains presents some analogies that suggest that they might share common features in their tertiary folding. The predicted structure of α3 is strongly reminiscent of that of immunoglobulin constant domains. Possible arrangements of elements of secondary structure are discussed, as an attempt to situating the polymorphic regions of HLA class I antigens in a spatial context.

Human class I histocompatibility antigens are a family of extremely polymorphic cell surface glycoproteins encoded by the polyallelic HLA-A, -B, and -C loci of the major histocompatibility complex (MHC).[1] Polymorphism is thought to be intimately related to their role in restricting the recognition of foreign antigens on the surface of virus-infected (McMichael et al., 1977) or otherwise modified cells (Dickmeiss et al., 1977) by cytotoxic T lymphocytes (CTLs). For this reason, much effort is being devoted to elucidate the nature and location of diversity areas in the molecule that may form the alloantigenic determinants. Following the establishment of the primary structure of HLA class I antigens, evidence accumulated in support of the notion that the extracellular portion of their heavy chain may be organized in three globular domains of about 90 residues each (López de Castro et al., 1979; Orr et al., 1979b) that were designated as α1, α2, and α3. Sequence comparisons among different HLA and H-2 antigens have shown that variable positions are located mainly, although not exclusively, in the two amino-terminal α1 and α2 domains. Furthermore, a number of discrete segments have been made evident where variable positions cluster. The relevance of these variable segments as putative alloantigenic sites is strongly suggested by comparative analysis of the sequences of cross-reactive alloantigens such as HLA-A2 and -A28 (López de Castro et al., 1982) or HLA-B7 and -B40 (López de Castro et al., 1983). Unfortunately, the precise delineation of the variable segments and the definition of putative new ones are biased by the limited number of sequences available. There are a significant number of substitutions in both HLA and H-2 antigens, which lie outside the variability clusters as outlined so far (Kimball & Coligan, 1983).

A limitation of sequence studies is that they do not provide information about the spatial relationship of variable residues that are distant in the linear sequence, and therefore it is generally difficult to assign the contribution of a given substitution to putative antigenic sites. This is especially so because there are a significant number of differences even between closely cross-reactive antigens (López de Castro et al., 1982). In the absence of a crystallographic model of class I molecules, a number of other approaches have proved useful in outlining the relative location of some antigenic determinants. Thus, proteolytic digestion studies have shown that the alloantigenic determinants of H-2K^b antigen are located in the amino-terminal 180 residues (Yokoyama & Nathenson, 1983) as expected from the sequence variability of this portion of the molecule. The use of allospecific monoclonal antibodies in competitive binding studies has established the existence of spatially separated alloantigenic sites in H-2 antigens (Ozato et al., 1981). An indication of the topographic arrangement of these allodeterminants may be obtained by combining such studies with the use of biochemically characterized H-2 mutants (Hämmerling et al., 1982). At least two separate sites are defined in this way that correlate with segments of variability as outlined by sequence comparisons, thus providing information about the relative location of such segments. A new and promising approach makes use of the possibility of constructing hybrid molecules by joining together fragments of various H-2 genes. The corresponding antigens may be expressed on the surface of transformed cells, and the presence or absence of particular determinants may be screened with monoclonal antibodies of defined specificities. Evans et al. (1982) have shown that it is possible in this way to assign serologically defined antigens to discrete portions of the histocompatibility antigen molecule.

In this paper, an attempt has been made to locate the polymorphic regions of HLA class I antigens as defined by sequence in a spatial context by means of methods that make use of the knowledge of the primary structure to predict the

† From the Department of Immunology, Fundación Jiménez Díaz. Avenida Reyes Católicos, 2 Madrid-3, Spain. *Received April 28, 1983.* Supported in part by grants from the Spanish Ministerio de Educación y Ciencia (No. 3448-79) and from the Fondo de Investigaciones Sanitarias de la Seguridad Social (No. 10/81 and 12/81). M.A.V. and A.E. are fellows of the F.I.S. and of the C.A.M.P. of Madrid, respectively.

[1] Abbreviations: MHC, major histocompatibility complex; CTL, cytotoxic T lymphocyte; Ig, immunoglobulin.

0006-2960/84/0423-0823$01.50/0

distribution of surface and internal residues and the basic features of polypeptide folding.

Materials and Methods

Variability Analysis. An estimation of variability at individual positions of HLA heavy chains was carried out by the method of Wu & Kabat (1970). Briefly, a variability parameter was computed for each position as the ratio between the number of different residues at a given position and the frequency of the most common residue at that position. This frequency is defined as the number of times the most common residue occurs divided by the total number of proteins being considered. Sequence data for only four individual proteins are available: HLA-B7 (Orr et al., 1979b), HLA-B40 (Lôpez de Castro et al., 1983), and HLA-A2 and HLA-A28 (Lôpez de Castro et al., 1982). In addition, the selection of these proteins is biased by the antigenic cross-reactivity of the pairs -B7/-B40 and -A2/-A28. Thus, the analysis of HLA variability must be considered only as a qualitative estimation with no statistical value.

Hydropathy Analysis. The distribution of hydrophilic and hydrophobic areas of HLA chains was calculated as described by Kyte & Doolittle (1982). A hydropathy value was assigned to each amino acid, and these values were continuously averaged in spans of seven consecutive residues from the N terminus to the C terminus of the chain, with overlapping segments displaced from each other by one position. Average hydropathy was plotted vs. sequence position at the middle point of each segment. The entire procedure was encoded in a Basic program and run in an HP-85 computer equipped with an automatic plotter.

Prediction of Secondary Structure. The occurrence of putative α-helices and β strands was studied by two independent procedures: (A) the directional method of Garnier et al. (1978), in which the conformation of a given residue is predicted taking into account the influence of residues up to eight positions before and after in the sequence. Conformational parameters for each residue are derived from the statistical analysis of 25 proteins of known sequence and conformation. Decision constants for α-helix and extended conformation of 158 and 87.5 centinats, respectively, were incorporated into the algorithm. These values are to be applied for proteins with less than 20% α-helix and more than 20% β structure, which is the case for HLA antigens as established by circular dichroism studies (Lancet et al., 1979; Trägard et al., 1979). This method has been reported to correctly predict about 80% of residues in β pleated sheet conformation for proteins built predominantly from β sheets (Busetta & Hospital, 1982). (B) The second is the theoretical method of Lim (1974b). This method develops algorithms for the prediction of α-helices and β structural regions on the basis of a series of structural principles (Lim, 1974a) that take into account packing interactions of polypeptide chains in globular proteins as well as interactions of proteins with water molecules. Since it is based on the stereochemistry of the Pauling–Corey α-helix and β structure, the method imposes very severe restrictions and may fail in locating regions with significant deviations from the theoretical parameters of α-helix and β strands. Hence, Lim's method was used to reinforce the predictions obtained by that of Garnier et al., concerning α-helical and β pleated sheet residues. The percentage of residues correctly predicted as being in helical or nonhelical conformation and in β or non β structure is 80% and 85%, respectively (Lim, 1974b).

Prediction of β-turns was carried out as described (Chou & Fasman, 1979) by means of an algorithm derived from the statistical analysis of a great number of chain reversal regions in 29 proteins of known three-dimensional structure. This method has been shown to correctly predict the location of β-turns within ± 2 residues in 78% of the cases.

The procedures of both Garnier et al. and Chou and Fasman were encoded in Basic programs and run in a HP-85 computer. Probability plots of β-turn occurrence were generated from the computer output with an automatic plotter. The procedure of Lim was not computerized.

Results

Variability in HLA-A and HLA-B Molecules. The alignment of HLA-B7, -B40, -A28, and -A2 sequences is presented in Figure 1. Overall variability in each domain is plotted in Figure 2. There are 51 positions of nonidentity, 47 of which are in the amino-terminal 194 residues. Differences between allelic products are predominantly located in the $\alpha 1$ domain (13 of 19 substitutions or 68% for the -B7/-B40 pair and 6 of 10 differences or 60% for the -A28/-A2 pair, whereas positions identical in either allelic pair but different when comparing -A vs. -B products (hereon referred to as A/B differences) are predominantly in the $\alpha 2$ domain and at the N-terminal portion of $\alpha 3$ (17 of 24 or 71%). There are four major clusters of variable positions spanning residues 63–80, 105–116, 138–156, and 177–194. These segments together include 35 (69%) nonidentical positions although they encompass only 25% of the polypeptide length. Other substitutions appear scattered or forming small clusters, such as that in segment 41–45. A striking distribution of differences is observed within the first major diversity segment. Substitutions within allelic products are concentrated between residues 66 and 74 in both -A2/-A28 and -B7/-B40 pairs followed by a clustering of A/B differences between residues 76 and 80 (Figure 1). This pattern is not observed in the second cluster (105–116) nor is it clear in the third one (138–156). The fourth major cluster (177–194) is unique in that no -A2/-A28 or -B7/-B40 differences have been detected in it. However, five A/B differences are concentrated between residues 177 and 184.

Hydropathic Properties of the HLA Heavy Chain. The hydropathy of HLA chains through their amino acid sequence was calculated to outline the relative distribution of hydrophobic and hydrophilic areas and to assess the effect of residue substitutions on the local polarity of the protein. Figure 3A shows the hydropathy profile of HLA-B7 and HLA-B40 $\alpha 1$ domains. Areas of maximum hydrophobicity are located in the N-terminal half of the domain, while the C-terminal half possess a remarkably polar character. All 13 differences between -B7 and -B40 in the $\alpha 1$ domain are accommodated in such a way that the local hydropathy remains essentially unchanged. The only exception observed is due to the Gln/Leu substitution at position 32 (Figure 1). This conservation of local polarity is most striking in the variable segment spanning residues 63–74, which includes a cluster of six differences (Figure 1). Obviously, the pattern does not necessarily hold when other antigens are compared. For example, a very conspicuous change in the hydropathy profiles of highly homologous H-2D^b and H-2L^d antigens is observed around residues 60–70 (data not shown). It is due to the presence of Ile at positions 63 and 66 in H-2L^d substituting Glu and Lys, respectively, in H-2D^b (Kimball & Coligan, 1983), which strongly increases the local hydrophobicity of H-2L^d in this segment. The hydropathic profile of the HLA-A28 $\alpha 1$ domain is compared to that of HLA-B7 in Figure 3B. The profiles are very similar, the only significant difference being an increase in the hydropathy upspike centered around residue 79

FIGURE 1: Comparison of the amino acid sequences of papain-solubilized HLA-B7 (Orr et al., 1979b), HLA-B40 (López de Castro et al., 1983), and HLA-A28 and HLA-A2 (López de Castro et al., 1982). Blanks correspond to nonassigned residues. Positions of nonidentity are boxed as to distinguish differences between alleles from those between products of different loci. Asterisks denote variable positions in which there are nonassigned residues. The following one-letter code for amino acids is used: A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; Z, Glx.

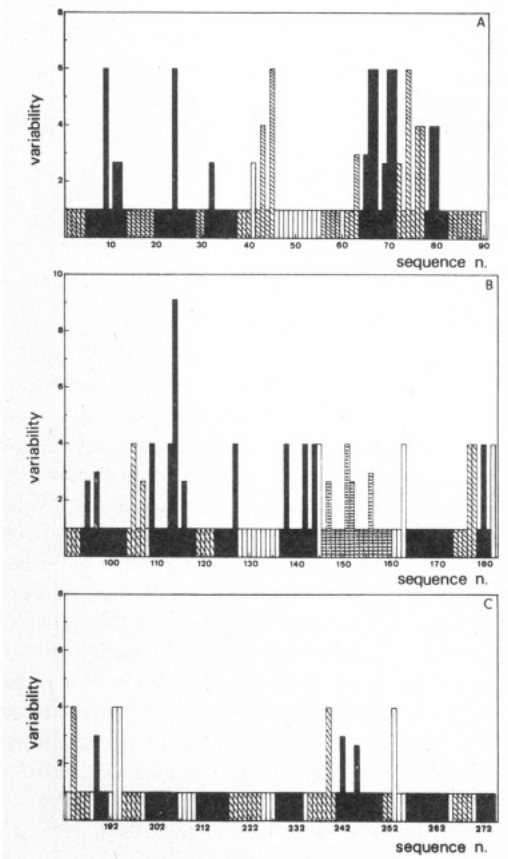

FIGURE 2: Variability of the amino acid residues in (A) $\alpha 1$, (B) $\alpha 2$, and (C) $\alpha 3$ domains of HLA class I antigens. Calculated from data in Figure 1. Superimposed are β strands (solid segments), β-turns (diagonally hatched segments), and α-helix (horizontally hatched segments) as predicted for HLA-B7 according to Garnier et al. (1978).

Table I: β Strand Prediction for the HLA $\alpha 1$ Domain[a]

strand	HLA-B7	HLA-B40	HLA-A28[b]
A	5–13 (5–9)	6–13 (5–9)	5–13 (5–9)
B	20–28 (21–28)	20–28 (21–28)	20–28 (21–28)
C	31–37 (32–36)	31–38 (32–36)	31–37 (32–36)
D	64–71 (65–71)	63–66	64–70
E	78–82	73–83	72–84

[a] Residues predicted as being in β-strand conformation by the method of Garnier et al. (1978) are indicated. In parentheses are the predictions obtained by the procedure of Lim (1974b). [b] A short β strand spanning residues 42–44 is predicted in this protein by the method of Garnier et al. (1978) but not in HLA-B7 or -B40. It has not been tabulated.

for HLA-A28, which is a consequence of the clustering of four A/B differences between residues 76 and 80.

The second domain, $\alpha 2$, shows an alternating pattern of four hydrophobic and four hydrophilic areas (parts C and D of Figure 3). A drastic change in local hydropathy between HLA-B7 and HLA-B40 is observed centered around residue 154. This change is due to the substitution of Glu-152 and Arg-156 in -B7 for Val and Leu, respectively, in -B40 (Figure 1). The hydropathic profile of HLA-A2 in this domain is quite similar to that of HLA-B40 (Figure 3D) in spite of the existence of at least 17 differences between both molecules (Figure 1). HLA-A2 was chosen for comparison instead of HLA-A28 because the sequence of the former in $\alpha 2$ is nearly complete.

The C-terminal domain, $\alpha 3$, presents two major hydrophobic segments centered around residues 207 and 247 and two prominent hydrophilic peaks centered around positions 221 and 254. Other, less conspicuous, peaks of alternating hydropathic character complete the profile of this domain (Figure 3E).

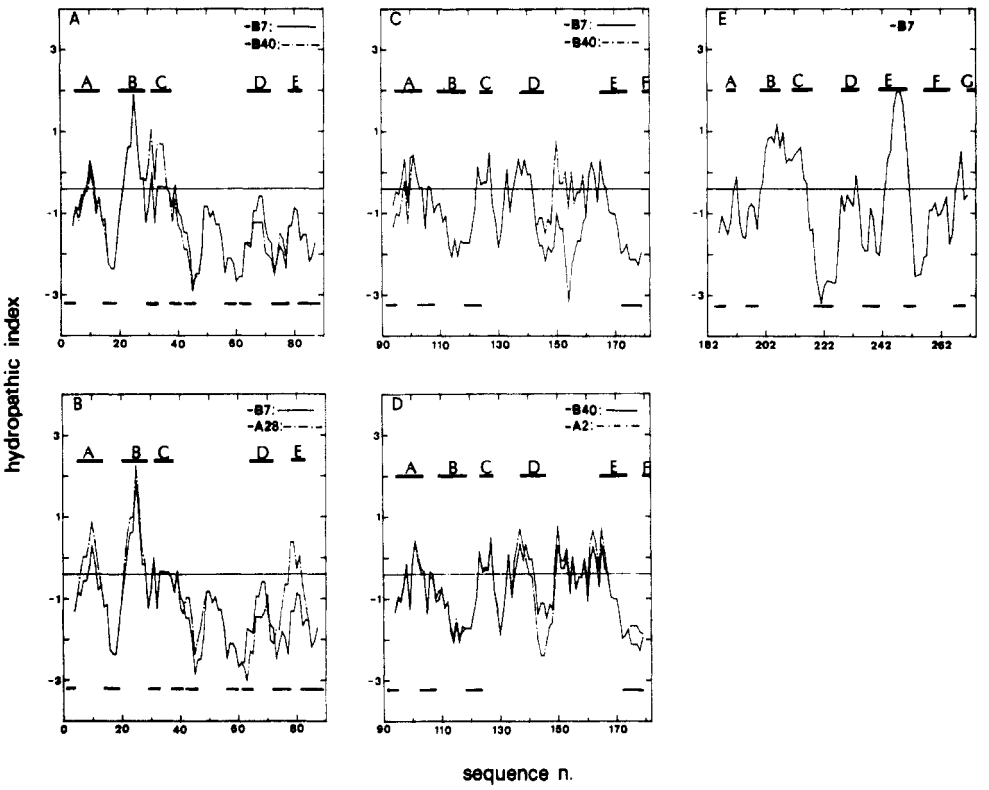

FIGURE 3: Hydropathy profiles of HLA class I antigens, obtained according to Kyte & Doolittle (1982) at a span setting of 7 (see Materials and Methods). The midpoint line of the profiles corresponds to the grand average of the hydropathy of the amino acid compositions found in most proteins of known sequence. β strands and β-turns are represented above and below the midpoint lines, respectively, as predicted for HLA-B7. (A) α1, HLA-B7 vs. HLA-B40; (B) α1, HLA-B7 vs. HLA-A28; (C) α2, HLA-B7 vs. HLA-B40; (D) α2, HLA-B40 vs. HLA-A2; (E) α3, HLA-B7.

Table II: α-Helix and β Strand Prediction for the HLA α2 Domain[a]

strand	HLA-B7	HLA-B40	HLA-A2
A	94-103 (96-103)	94-103 (96-103)	94-100 (96-103)
B	109-118 (108-111)	109-118 (108-111)	109-118 (107-111)
C	123-127 (121-128)	123-127 (121-128)	123-127 (121-128)
D	137-144	137-145	136-141
E	164-173 (163-172)	164-173 (163-172)	164-173 (164-172)
F	179-181	179-181	179-182
α-helix	146-160	146-160 (148-157)	146-158 (148-157)

[a] See footnote *a* in Table I.

Table III: β Strand Prediction for the HLA α3 Domain[a]

strand	HLA-B7	strand	HLA-B7
A	189-191	E	241-250 (241-250)
B	200-206 (199-208)	F	256-264 (257-261)
C	211-217 (211-217)	G	271-274 (270-274)
D	228-233 (227-233)		

[a] See footnote *a* in Table I.

Prediction of Secondary Structure. All β strands predicted by Lim's method were also predicted by that of Garnier et al. in all three α1, α2, and α3 domains. However, strands predicted by the latter method were not always confirmed by Lim's procedure or were sometimes of greater length (Tables I–III). When this was the case, the predictions obtained with the method of Garnier et al. were adopted because Lim's algorithm is known to lead to incomplete prediction of β strands, especially for those composed mainly of hydrophilic residues (Lim, 1974b).

(1) α1 Domain. No α-helices are predicted in the α1 domain of HLA-B7, -B40, or -A28. The span of predicted β strands in this domain is presented in Table I for these three proteins. The boundaries of the β strands and the structure of residues between strands were established, taking into account the prediction of β-turns (Figure 4). The N-terminal portion of α1 (residues 1–40) shows a well-defined secondary structure consisting of three β strands spanning residues 5/6–13, 20–28, and 31–37 bounded by β-turns at residues 1–4, 14–17, 16–19, 29–32, and 37–40. These strands involve the most hydrophobic segments of α1 (Figure 3), and for this reason it is conceivable that they may contribute to the domain core. A G1 β-bulge (Richardson et al., 1978) could exist at positions 26 and 27, in the second strand. This is suggested by the disruption of alternation in the polarity of residues 23–29 at the constant Gly residue at position 26 (see Figure 1). In addition, the detection of a putative β-turn at position 24–27 (Figure 4) is consistent with the known preference of G1 β-bulges to participate in interlocking structures in which Gly at position 1 of the β-bulge is simultaneously the favored residue at position 3 of the β-turn (Richardson et al., 1978).

Residues 41–50 do not show a consistent pattern of secondary structure. This area consists of a conserved hydrophobic sequence (residues 49–52) flanked by two very hydrophilic segments spanning residues 42–48 and 53–55 (Fig-

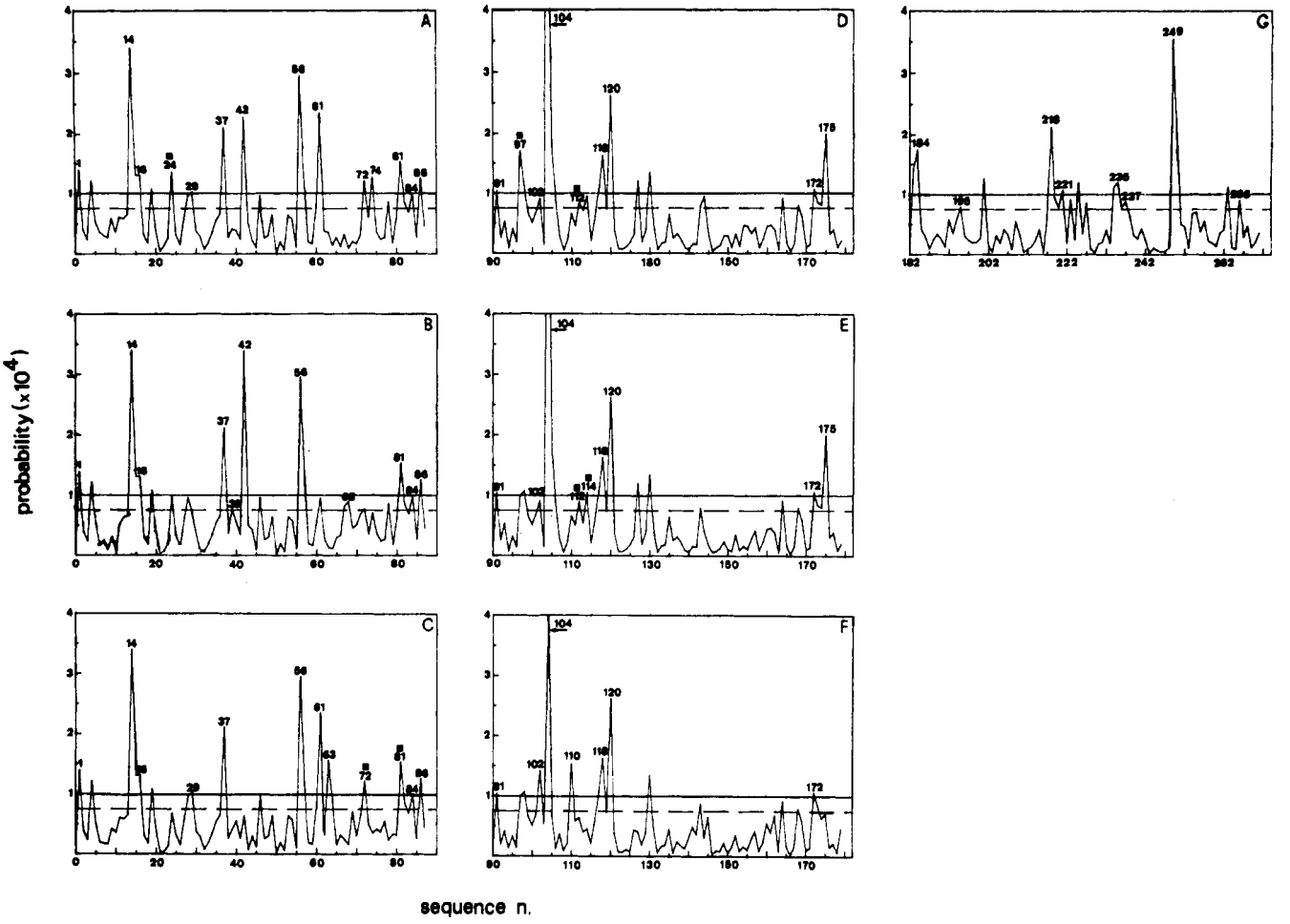

FIGURE 4: Probability of tetrapeptide β-turns in HLA class I antigens. β-Turn conformation is assigned to those tetrapeptides that fulfill the two following requirements: (1) Their probability of β-turn occurrence is greater than 0.75×10^{-4} (dashed line). (2) Their β-turn potential is greater than both their α-helix and β strand potentials (Chou & Fasman, 1979). Predicted β-turns are denoted by the position number of their first residue over their corresponding probability peaks. Those marked with a solid square fall into β strands as predicted according to Garnier et al. (1978) and are not assigned. (A) HLA-B7, α1; (B) HLA-B40, α1; (C) HLA-A28, α1; (D) HLA-B7, α2; (E) HLA-B40, α2; (F) HLA-A2, α2; (G) HLA-B7, α3.

ures 1 and 3). The first of these segments may include a β-turn involving residues 42–45 at least in HLA-B7 and -B40 (Figure 4). This β-turn is also conspicuously predicted in most H-2 antigens (data not shown) but not in HLA-A28 or -A2. The C-terminal portion of α1 (residues 56–90) appears to be folded in two β strands flanked by β-turns. The beginning of this area is marked by a β-turn at residues 56–59. These residues are constant in all HLA and H-2 molecules investigated so far, which suggests that this β-turn may be a structurally important feature of the molecule. The prediction of the detailed structure of residues 61–80 is somewhat obscured by the exceptional variability of this segment. A β-turn is predicted at residues 61–64 in HLA-B7 and -A28 (Figure 4). Residues 64–71 include a β strand whose precise span is predicted differently in different molecules (Table I). A β-turn involving residues 72–75 is predicted in -B7 and -A28 (Figure 4), which may mark the end of the strand. This β-turn does not appear in HLA-B40, but for this protein a corresponding β-turn is predicted at residues 68–71, which is also present in H-2 antigens (data not shown). A final strand starts immediately after these turns and extends up to residue 81, being bounded by three consecutive β-turns at the end of the domain (Table I and Figure 4). The whole variable area spanning residues 60–80 is remarkably polar. However, the less hydrophilic portions of this segment correlate with the predicted β strands. Conversely, the flanking β-turns lie in zones of maximal local hydrophilicity.

(*2*) *α2 Domain.* Prediction of secondary structure in this domain was carried out for HLA-B7, -B40, and -A2. As in the hydropathy analysis, the latter protein was chosen instead of -A28 because its sequence is almost complete (Figure 1). Those positions that are not yet assigned (residue 143 in HLA-B40 and residues 108, 131, 132, and 137 in HLA-A2) were considered to be identical with the most frequent residue at each position in other HLA antigens. Overall definition of secondary structure is good. There are six β strands (Table II). The C-terminal boundary of the first one (94/96–103) is defined by a conspicuous β-turn at residues 104–107. This β strand includes the first cysteine (position 101) of the HLA molecule, and there are some indications that this residue might be forming a β-bulge with Gly-100. This is suggested by the reversal of alternative polarity between residues 97 and 103 at Gly-100 and by the relatively high β-turn probability at residues 98–101 (Figure 4) that would include Gly-100 as the residue in its third position (this β-turn is not predicted because the β strand potential of the corresponding residues is bigger). As it was argued above, these two features are frequent in G1 β-bulges. The second β strand (residues 108–118) includes a number of variable positions. This strand is composed mainly of polar residues (Figures 1 and 3), which explains why the span predicted by Lim's method is significantly smaller than that predicted by the procedure of Garnier et al. (Table II). Two β-turns included in residues 118–124 (Figure 4) mark the C-terminal boundary of the second strand,

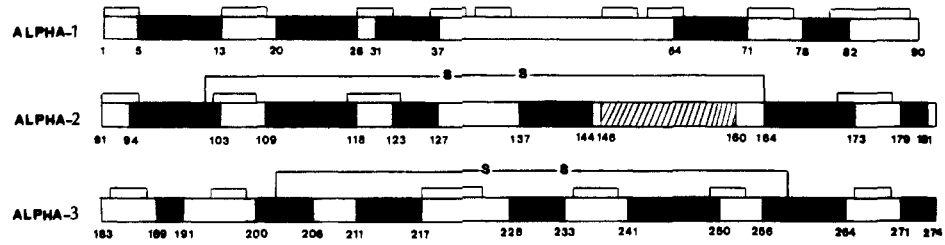

FIGURE 5: Diagrammatic scale drawing showing the prediction of secondary structure for HLA-B7 heavy chain domains. Areas in β strand and α-helix conformation according to Garnier et al. (1978) are represented as solid and diagonally hatched segments, respectively. The β-turns are marked as lines over the bar diagram. Overlapping β-turns are drawn together.

which is immediately followed by a third one (residues 121/123–127/128) that has the features of internal strands (Lim, 1974a) since it includes a cluster of hydrophobic residues (123–126) flanked at both ends by bulky polar side chains (residues 121, 122, and 127–129). A short segment spanning residues 129–136 separates the third and fourth β strands. The structure of this segment is not well defined although it might include a putative β-turn at residues 130–134 (Figure 4) that cannot be assigned because α-helix and β strand potentials for these residues are slightly bigger.

A α-helix of 11 or 14 residues spans the segment between the fourth and the fifth β strands within residues 146–160 (Table II). The substitution of Val-152 and Leu-156 in -B40 and -A2 by the polar residues Glu and Arg in HLA-B7 (Figure 1) precludes the fulfillment of the criteria applied by Lim's method for the prediction of α-helices. However, these substitutions could be accommodated without disturbing the helical structure. The fifth β strand (Table II) is flanked at the C-terminal end by one or two β-turns at residues 172–178 (Figure 4). A very short β strand is predicted at the C-terminal end of the domain, spanning residues 179–182 (Table II).

There is a good correlation between the distribution of β strands and β-turns and the features of the hydropathy profiles of $\alpha 2$ domain. Strands A, C, D, and E lie into the main hydrophobic portions of the chain. Strands B and F and all β-turns are located in hydrophilic areas (Figure 3).

(3) $\alpha 3$ Domain. Seven β strands are predicted for this domain (Table III) that are bounded by β-turns as shown in Figure 4 (a β-turn spanning residues 207–210, between the second and third strands, is not predicted by the Chou and Fasman method but it is clearly defined in the predictions based in that of Garnier et al.). Strands B, C, and E correspond to segments of high local hydrophobicity, the remaining strands being more hydrophilic (Figure 3E). β-Turns lie in areas of high local hydrophilicity, as it is the case in $\alpha 1$ and $\alpha 2$ domains.

A remarkable feature of the secondary structure predictions in all three $\alpha 1$, $\alpha 2$, and $\alpha 3$ domains is that essentially all predicted β-turns lie in interstrand regions as outlined by the prediction of β strands. Since both types of structure are derived from independent procedures, their lack of overlapping confers mutual consistency to the assignments. A schematic diagram of the secondary structure predictions is presented in Figure 5.

Discussion

A comparison of available HLA sequences clearly illustrates the point that structural polymorphism of these molecules is mainly restricted to the N-terminal two-thirds of the extracellular portion of HLA heavy chain, which encompasses $\alpha 1$ and $\alpha 2$ domains. As in other multidomain proteins, structural diferentiation within the polypeptide chain may be related to a corresponding division of function. Hence, the conserved C-terminal $\alpha 3$ domain probably interacts with β_2-microglobulin, as suggested by their Ig-like structure (Petersen et al., 1972; Orr et al., 1979) and by recent experiments on H-2 antigens (Yokoyama & Nathenson, 1983), constituting an essentially invariant region that might function as a recognition unit for cell interactions (Jensenius & Williams, 1982). By contrast, $\alpha 1$ and $\alpha 2$ domains must be responsible for most of the antigenic polymorphism and they presumably carry the structural elements that determine cytotoxic T cell restriction. The unbalanced distribution of allele-related vs. A/B differences in $\alpha 1$ and $\alpha 2$ domains that is apparent in Figure 1 must be interpreted with caution because only two allelic products from each locus are being considered and because they have not been randomly selected since both HLA-B7 and -B40 and HLA-A2 and -A28 are strongly cross-reactive alloantigens. Nonetheless, their comparison suggests that the $\alpha 1$ domain could play a predominant, although nonexclusive (Hämmerling et al., 1982), role in determining alloantigenic polymorphism. From this it does not follow, however, that the diversification rate of both domains should be different. Indeed, the number of substitutions between HLA-B40 and HLA-A2 or -A28 is nearly the same in $\alpha 1$ and $\alpha 2$ (Figure 1).

The search for locus-specific structural features in class I histocompatibility antigens has found little success (Kimball & Coligan, 1983), perhaps due to the limited number of known sequences. Locus-specific antigenic markers are difficult to find because they would not elicit a response upon allogenic immunization, unless they were themselves polymorphic. This is probably the case with the supertypic diallelic specificities HLA-Bw4 and -Bw6, which are associated with all HLA-B antigens (Ayres & Cresswell, 1976). Indeed, locus-specific antigens may be revealed upon xenogeneic immunization with purified HLA-A or HLA-B proteins followed by absorption of the resulting antisera with soluble HLA-B or HLA-A products, respectively (Cresswell & Ayres, 1976). Thus, it is conceivable that some of the A/B differences observed in Figure 1 could be related to locus-specific structural features. In general, A/B differences are close in sequence to positions of allelic variability, which obscures the identification of putative locus-specific areas, the only exception being the cluster at position 177–184. This segment has been predicted to constitute an antigenic determinant in HLA-B7 on the basis of its high local hydrophilicity (Vega et al., 1982). Its sequence in H-2 antigens is significantly different from those of HLA-A and HLA-B proteins. In spite of its apparent uniqueness, it is clear that establishing the putative relevance of residues in this segment as locus-specific structural markers of HLA class I antigens must await additional information.

The scattering of substitutions through the sequence makes it difficult to understand the native structure of polymorphic areas. In the absence of a three-dimensional model it is possible to address this question by evaluating the information contained in the sequence regarding the folding of the polypeptide chain by means of methods of high predictive accuracy

as estimated by their correlation with proteins of known X-ray structure. Local hydropathy of polypeptide chains has been shown to remarkably reflect the distribution of surface and internal residues in proteins, so that hydrophilic segments are generally external whereas hydrophobic ones usually correspond to interior portions (Kyte & Doolittle, 1982). Thus, the structural elements of the HLA chain that should be external, such as β-turns, and those internal β strands or other structures that contribute to the core of the domains may be outlined by comparing the predicted secondary structure with hydropathy profiles as shown in Figure 3. If the distribution of variability is considered in relation to these structural features (Figure 2), some interesting points emerge. Thus, in the α1 domain variability within the N-terminal 40 residues is restricted to a few scattered positions. The segment includes A, B, and C β strands, which correspond to major hydrophobic peaks of the α1 hydropathy profile and may therefore contain many buried residues. By contrast, D and E β strands encompass almost completely the main variability cluster of the domain (residues 63–80). These β strands are predominantly composed by hydrophilic residues and possess the features of peripheral β strands (Lim, 1974a) that are found at the protein surface, shielding the hydrophobic core. This type of strand is not analyzed in Lim's algorithm (Lim, 1974b) so that bands D and E of α1 are not predicted by this method. Whether the different span of these β strands in different specificities and the exact location of the interstrand segment reflect real structural differences or are just an obscuring effect of the sequence variability over the predictions cannot be decided. The coincidence of predicted β-turns with hydrophilic peaks of the hydropathy profile in the α1 domain is in agreement with the strong preference of this type of structure to be at the protein surface. Although some variability occurs in β-turns, most notably in that at residues 42–45, it is not especially conspicuous. Such conservation may be attributed to their strict steric requirement as well as to their precise role in the reversal of the polypeptide chain (Chou & Fasman, 1977). Strikingly, residues 46–55 show no predicted structure. Still, they are conserved in HLA (Figure 1) and H-2 antigens (Kimball & Coligan, 1983), which suggests that they may play an important structural role. Possibly the hydrophobic residues in this segment (49–52) are buried and shielded by the flanking polar ones at the surface (48 and 53–55). The predicted secondary structure in the α2 domain along its 37 N-terminal residues (91–127) is very similar to that in the corresponding region of α1 regarding the positioning and length of A, B, and C strands and their intermediary β-turns (Figure 5). Strands A and C correlate with hydrophobic upspikes of the α2 hydropathy profile, suggesting that at least some residues in these strands may be buried. This is particularly clear in the internal strand C, for the stretch of four consecutive apolar residues 123–126. This stretch is invariant in all HLA (Figure 1) and H-2 sequences known (Kimball & Coligan, 1983). Significantly, the corresponding C β strand in the α1 domain is also of the internal type, as mentioned above, which adds to the analogy of α1 and α2 domains in this region. By contrast, there is a conspicuous difference between strands B of both domains regarding their hydropathy. While the former corresponds to the peak of maximum hydrophobicity of the α1 domain, the B strand of α2 is quite hydrophilic and presumably most of its side chains are exposed at the surface. This strand includes most of the variability cluster at residues 105–116.

At the C-terminal part of α2, the variable segment 138–156 includes the D strand and the only predicted α-helix of the molecule. The segment that spans this α-helix (146/148–157/160) has been pointed out as being a region of critical importance in determining graft rejection (Nairn et al., 1980) and as a site for CTL recognition (Krangel et al., 1983) and serologic allodeterminants (Hämmerling et al., 1982; Lõpez de Castro et al., 1983). The pattern of restricted variability in this segment adjusts well to the steric restrictions of α-helices. For instance, the variable residues at positions 152 and 156 in HLA-B7 and -B40 would be brought into proximity by the helical turn. The same is true for the three substitutions in HLA-A2 variant DK1, at positions 149, 152, and 156 (Krangel et al., 1983). In both cases, the nature of the substitutions involve concomitant changes of hydrophobic side chains by hydrophilic ones, thus inducing a strong change in local polarity that could influence the binding of antibodies or T cell receptors without disturbing the conformation of the site. In the mouse, a concomitant variation at positions 152, 155, and 156 is observed when comparing H-2K^b, H-2L^d, and H-2D^b antigens (Kimball & Coligan, 1983). The H-2K^b mutant bm1, which vigorously rejects exchanged grafts with the parental H-2K^b strain (Nairn et al., 1980), differs from the H-2K^b molecule in all three positions and is identical in all of them to H-2L^d, a fact that has been interpreted as originating from a genetic event analogous to gene conversion (Pease et al., 1983).

Most of the variability cluster spanning residues 177–184 is included in the hydrophilic F strand and in its preceding β-turn. This β-turn is the attachment site of the second carbohydrate moiety of H-2 antigens at position 176 (Kimball & Coligan, 1983). Thus, the carbohydrate attachment sites in both α1 and α2 domains are predicted to be in surface reversal points of the polypeptide chain. It is noteworthy that many variable positions in α1 and α2 are included in the predicted β strands, β-turns, or α-helix (Figure 2) which may conceivably restrict the extent of allowed variability of those residues in defined conformational states.

The Ig-like nature of α3 domain (Orr et al., 1979a) provides a model for ordering the predicted β strands following the pattern of Ig constant domains. The prediction of secondary structure for α3 adjusts well to that expected from its sequence similarity with immunoglobulins, in agreement with previous reports (Cohen et al., 1980). This is illustrated in Figure 6 for the distribution of strands, which also includes a prediction of the C_γ3 Ig domain to assess the reliability of the assignments. If the Ig folding for α3 is assumed (Figure 7A), most of the variability in this domain is located in the connecting loops between strands or at the edges of the strands, the strands themselves being invariant except for the limited variability in strand E. The diversity area spanning residues 193–198 in H-2 antigens, which has been predicted to form an antigenic determinant (Vega et al., 1982), is located in the hydrophilic loop connecting the contiguous antiparallel A and B strands, and therefore it is likely to constitute a defined spatial epitope. It is conceivable that the α3 domain and β_2-microglobulin interact in a similar way to that of homologous Ig constant domains. In immunoglobulins this interaction involves a contact face composed by A, B, D, and E strands (Deisenhofer, 1981). Therefore, the equivalent strands in α3 could be interacting with β_2-microglobulin, which is in agreement with the higher hydrophobicity of the β sheet integrated by these four strands (Figure 3E). In contrast, C, F, and G are typical surface-type β strands that would expose a hydrophilic external face.

A point of critical importance is outlining the spatial relationship between the elements of secondary structure predicted from the sequence in α1 and α2 domains. Unfortunately, reliable algorithms for the prediction of tertiary folding of polypeptides are not sufficiently developed. Furthermore, it

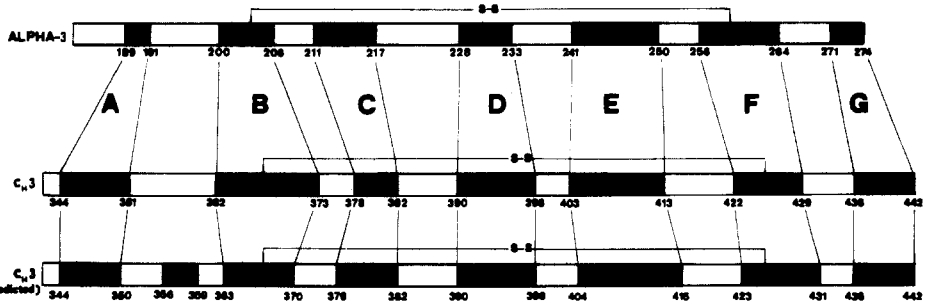

FIGURE 6: Diagrammatic comparison of predicted β strands for the HLA-B7 $\alpha 3$ domain with those of immunoglobulin $C_{\gamma 1}3$ as determined by X-ray crystallography (Deisenhofer, 1981) or predicted by the method of Garnier et al. (1978).

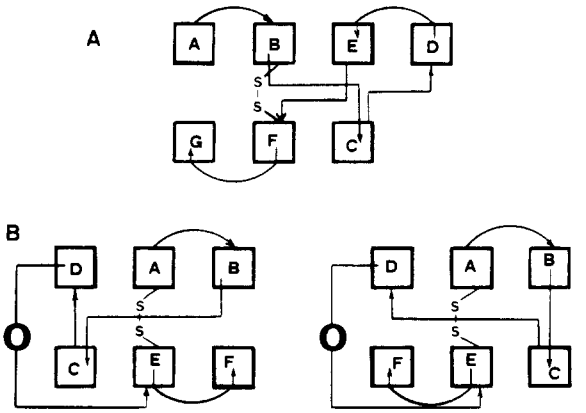

FIGURE 7: (A) Diagrammatic representation of the predicted folding for the HLA class I $\alpha 3$ domain. The convention of Levitt & Chothia (1976) was used. β strands and α-helix are represented by boxes and circles, respectively. Strand connections are represented by arrows. (B) Diagrams representing putative folding models for the HLA class I $\alpha 2$ domain. Note that in both models the β strand topology of the *y* face of Ig constant domains (Beale & Feinstein, 1976) is preserved in the CEF β sheet. In addition, changes in the β strand topology of the *x* face of immunoglobulins are minimized. The type of strand connections in the DAB/CEF model is much more similar to that of Ig constant domains than that of the DAB/FEC model.

would require an unambiguous assessment of the conformational state of every residue (Burgess & Sheraga, 1975), which is not possible due to the error inherent in all predictive methods for secondary structure. Nevertheless, analysis of the folding observed in proteins of known structure provides a series of criteria that severely restrict the number of possibilities in which a given set of elements of secondary structure may be ordered (Levitt & Chothia, 1976; Richardson, 1977; Sternberg & Thornton, 1977a,b). These criteria were used for investigating possible folding models for the $\alpha 2$ domain on the basis of its predicted secondary structure. A systematic application of such criteria (see paragraph at end of paper regarding supplementary material) reduced the plausible ordering of β strands to two favored arrangements (Figure 7B), both consisting of two β sheets composed by strands DAB and CEF (or FEC), respectively. In both models most variable positions in the β strands belong to the DAB sheet. Strands A and B, which display much variability in HLA and H-2 antigens, are adjacent, forming a spatial cluster. By contrast, the CEF sheet is much more conserved. Both alternatives would have different implications as to the spatial relationship between the variability clusters at residues 105–116 and 177–184. In the orientation DAB/CEF both of them would be at the same side of the domain, whereas in the alternative arrangement DAB/FEC both clusters would be at opposite sides. The α-helix connecting strands D and E would presumably be protecting hydrophobic side chains from solvent, while exposing a hydrophilic side. Its topology is not determined. However, if $\alpha 2$ would form a β-barrel, as is the case in Ig domains, the crossover connection that forms the helix should be right-handed (Richardson, 1977). This would bring the helix close to strand C in the DAB/CEF model or to strand F in the DAB/FEC alternative. In either case the hydrophilic side of the helix would probably not be close to the variability cluster formed by strands A and B. It is clear that the selected models for $\alpha 2$ folding are no more than favored possibilities that are compatible with a number of frequently observed constraints in proteins. Nevertheless, they predict a restricted number of arrangements for the variable positions, which may be tested by means of competition studies with monoclonal antibodies (Ozato et al., 1981; Hämmerling et al., 1982) or by further sequence comparisons.

The lack of β strand definition in the middle part of the $\alpha 1$ domain prevents making a reliable outline of putative tertiary folding models. However, the weak sequence homology with the $\alpha 2$ domain (Orr et al., 1979b) and the similarities in their respective A, B, and C strands discussed above suggest that both domains could share common features in their folding. For instance, A and B strands could form a pair of antiparallel adjacent strands, as it is predicted in $\alpha 2$ and $\alpha 3$. This would bring into proximity the variable residues at positions 9–12 and those close to position 24. In the same way, the highly variable D and E strands of $\alpha 1$ could conceivably form a loop of contiguous antiparallel strands, as it is predicted for the C-terminal strands of $\alpha 2$ and $\alpha 3$, thus constituting a spatially defined epitope.

In summary, there is a good correlation between the local distribution of hydrophobicity through the HLA chain and the predicted features of secondary structure. The proposed model predicts that variable residues are frequently located in segments of defined structure, so that most substitutions appear to be accommodated within rigid structural frameworks that would impose important constraints to variability. This is a major difference with the hypervariable regions of immunoglobulins, which are located in bends without a defined secondary structure (Saul et al., 1978). It is striking that the single α-helix of the molecule encompasses a region of restricted diversity that has been pointed out as being a putative CTL recognition site (Krangel et al., 1983). This poses intriguing questions regarding the pathways of structural diversification of HLA domains in relation to their functional differentiation. The validity of the arguments discussed above hinges on the reliability of the methods employed for the prediction of secondary structure. However, the results presented in this study must be considered only as a starting point for the interpretation of the wealth of structural and serological data that are relevant to the mapping of HLA antigenic determinants and of putative functional sites.

Supplementary Material Available

Tertiary folding assessment of the HLA class I $\alpha 2$ domain (3 pages). Ordering information is given on any current masthead page.

References

Ayres, J., & Cresswell, P. (1976) *Eur. J. Immunol. 6*, 794–799.

Beale, D., & Feinstein, A. (1976) *Q. Rev. Biophys. 9*, 135–180.

Burgess, A. W., & Sheraga, H. A. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1221–1225.

Busetta, B., & Hospital, M. (1982) *Biochim. Biophys. Acta 701*, 111–118.

Chothia, C., & Janin, J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4146–4150.

Chou, P. Y., & Fasman, G. D. (1977) *J. Mol. Biol. 115*, 135–175.

Chou, P. Y., & Fasman, G. D. (1979) *Biophys. J. 26*, 367–383.

Cohen, F. E., Sternberg, M. J. E., & Taylor (1980) *Nature (London) 285*, 378–382.

Deisenhofer, J. (1981) *Biochemistry 20*, 2361–2370.

Dickmeiss, E., Soeberg, B., & Svejgaard, A. (1977) *Nature (London) 270*, 526–528.

Evans, G. A., Margulies, D. H., Shykind, B., & Ozato, K. (1982) *Nature (London) 300*, 755–757.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Hämmerling, G. J., Rüsch, E., Tada, N., Kumura, S., & Hämmerling, U. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4737–4741.

Jensenius, J. C., & Williams, A. F. (1982) *Nature (London) 300*, 583–588.

Kimball, E. S., & Coligan, J. E. (1983) *Contemp. Top. Mol. Immunol.* (in press).

Krangel, M. S., Biddison, W. E., & Strominger, J. L. (1983) *J. Immunol. 130*, 1856–1862.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Lancet, D., Parham, P., & Strominger, J. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3844–3848.

Levitt, M., & Chothia, C. (1976) *Nature (London) 261*, 552–557.

Lim, V. I. (1974a) *J. Mol. Biol. 88*, 857–872.

Lim, V. I. (1974b) *J. Mol. Biol. 88*, 873–894.

López de Castro, J. A., Orr, H. T., Robb, R. J., Kostyk, T. G., Mann, D. L., & Strominger, J. L. (1979) *Biochemistry 18*, 5704–5711.

López de Castro, J. A., Strominger, J. L., Strong, D. M., & Orr, H. T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3813–3817.

López de Castro, J. A., Bragado, R., Strong, D. M., & Strominger, J. L. (1983) *Biochemistry 22*, 3961–3969.

McMichael, A. J., Ting, A., Zweerink, H. J., & Askonas, B. A. (1977) *Nature (London) 270*, 524–526.

Nairn, R., Yamaga, K., & Nathenson, S. G. (1980) *Annu. Rev. Genet. 14*, 241–277.

Orr, H. T., Lancet, D., Robb, R. J., López de Castro, J. A., & Strominger, J. L. (1979a) *Nature (London) 282*, 266–270.

Orr, H. T., López de Castro, J. A., Lancet, D., & Strominger, J. L. (1979b) *Biochemistry 18*, 5711–5720.

Ozato, K., Henkart, P., Jensen, C., & Sachs, D. (1981) *J. Immunol. 126*, 1780–1785.

Pease, L. R., Schultze, D. H., Pfaffenbach, G. M., & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 242–246.

Peterson, P. A., Cunningham, B. A., Berggard, I., & Edelman, G. M. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1697–1701.

Richardson, J. S. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2619–2623.

Richardson, J. S. (1977) *Nature (London) 268*, 495–500.

Richardson, J. S., Getzoff, E. D., & Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2574–2578.

Saul, F. A., Amzel, L. M., & Poljak, R. J. (1978) *J. Biol. Chem. 253*, 585–597.

Sternberg, M. J. E., & Thornton, J. M. (1977a) *J. Mol. Biol. 110*, 285–296.

Sternberg, M. J. E., & Thornton, J. M. (1977b) *J. Mol. Biol. 115*, 1–17.

Trägård, L., Curman, B., Wiman, K., Rask, L., & Peterson, P. A. (1979) *Biochemistry 18*, 2218–2225.

Vega, M. A., Ezquerra, A., & López de Castro, J. A. (1982) *Biochem. Biophys. Res. Commun. 107*, 1545–1550.

Wu, T. T., & Kabat, E. A. (1970) *J. Exp. Med. 132*, 211–250.

Yokoyama, K., & Nathenson, S. G. (1983) *J. Immunol. 130*, 1419–1425.