Volume 12 Number 1 1984 Nucleic Acids Research

A program for prediction of protein secondary structure from nucleotide sequence data: application to histocompatibility antigens

Jiří Novotný and Charles Auffray*

Molecular and Cellular Laboratory, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, and *Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138, USA

Received 5 August 1983

ABSTRACT

A computer program is described which, given a nucleotide or an amino acid sequence, outputs protein secondary structure prediction curves as well as hydrophobicity and charged-residue profiles. The program allows for cumulative averaging of properties (secondary structure propensities, hydrophobicity and charge profiles) from several homologous primary structures, a novel concept shown to improve the predictive accuracy. The use of the program is demonstrated on a set of nucleotide and amino acid sequences from human and murine histocompatibility antigens of class I and II. The last extracellular domains of both class I and II antigens (α3 of class I, α2 and β2 of class II) and the β2-microglobulin domain are predicted to consist of seven anti-parallel β-strands, in accord with previous claims of homology between these domains and the constant domains of immunoglobulin chains. The remaining extracellular domains are all proposed to form an anti-parallel, four-stranded β-sheet with one of its faces being covered by α-helices and/or structureless segments ("open face sandwiches").

INTRODUCTION

The availability of techniques for DNA cloning, nucleotide sequencing and site-directed mutagenesis substantially broadened the scope of problems a molecular biologists is able to contemplate and solve in a short time. It is not unusual nowadays that "molecular engineering" experiments are designed with the purpose to rationally alter protein function by changing its structure, or that an antibody is sought against a structural determinant of a putative protein known to exist only as an open reading frame within a section of a gene sequence. Since secondary and tertiary structure considerations are crucial to these types of experiments and since the X-ray crystallographic data are still rather rare, there is a keenly felt need for a quick and reliable procedure which would allow to deduce the folding pattern from a given amino acid sequence. In this article we describe a computer program which provides a comprehensive secondary structure information in the form of curves (amino acid sequence profiles) computed from several different primary-structural parameters. Because the current knowledge of protein folding is far from being complete, the interpretation of the amino acid sequence profiles requires an experience and structural insight, yet often remains ambiguous. We therefore present an example of secon-

dary structure analysis which illustrates the types of reasoning and deductions likely to be relevant in this context.

Prediction of secondary structure from amino acid sequences dates back to early sixties (1,2). Several predictive algorithms are currently available (3-9), based on tabulated frequencies with which individual amino acid residues occur in polypeptide chain segments of defined secondary structure. The predictive power of all these methods, however, is limited by the inherent probabilistic nature of amino acid occurrences in α-helices, β-strands and reverse turns and has never exceeded 75%. One way to improve the accuracy of statistical predictions is to combine them with algorithms which employ different principles [*cf. e.g.* (10)]. There are three such methods currently avaliable (11-13), all based on the key role hydrophobicity plays in protein folding. Rose (14) and Rose and Roy (11) showed that "hydrophobicity profiles" [side chain hydrophobicities (15) are smoothed over short polypeptide chain segments and resulting numbers plotted against the amino acid sequence] accurately identify reverse turns and secondary structure elements. α-Helices and β-strands correspond to positions of hydrophobicity maxima whereas turns occur where the minima are (16). Since α-helices and β-sheets usually have polar surfaces which face the solvent and non-polar surfaces which close-pack against each other, and since the periodicities of side chains in helices [n+3 or n+4] and sheets [n+1] differ (17), it is conceivable to distinguish these two types of structures by examining the periodicity of non-polar residues. Schiffer and Edmundson (12) showed how "helical wheels" can identify α-helical segments: circular projections of non-α-helical segments display a random distribution of hydrophobic residues while arc-shaped clusters of these residues can be seen in true helices. Cid *et al.* (13) used characteristic periodicity of "bulk hydrophobic character" profiles (18) to identify helices and sheets, but it is evident that there are too many irregularities in secondary structures which prevent a routine use of their method.

Another phenomenon which can potentially be exploited for prediction purposes is that of amino acid sequence homology. The "stereochemical code" which relates primary structure to tertiary structure is degenerate (19) in the sense that many distantly related amino acid sequences [*e.g.*, myoglobins from various species] share the same fold. Here we demonstrate that the accuracy of secondary structure prediction is indeed improved if α-helical, β-strand and reverse turn propensities are computed from several homologous amino acid sequences and then averaged together. The predictive performance is improved even further when the above method is supplemented by profiles of hydrophobic and charged residues.

METHOD

A computer program, written in FORTRAN, combines the following features: *(i)*

input of nucleotide sequences [together with their reading frames] or amino acid sequences, in one-letter code, from terminal-formatted files (20). The program allows for reading either the whole string of letters as contained in the file or any substring thereof and is compatible with Dayhoff's Nucleic Acid and Protein Sequence Database (21). If several homologous sequences of different length are to be input, they have to be aligned first by inserting 'X' at the places of deletions. *(ii)* computation, averaging, optional smoothing (seven-point moving-window smoothing acording to the formula

$$N_i = \{7N_i + 3[2(N_{i-1}+N_{i+1}) + N_{i-2} + N_{i+2}] - 2(N_{i-3}+N_{i+3})\} / 21)$$

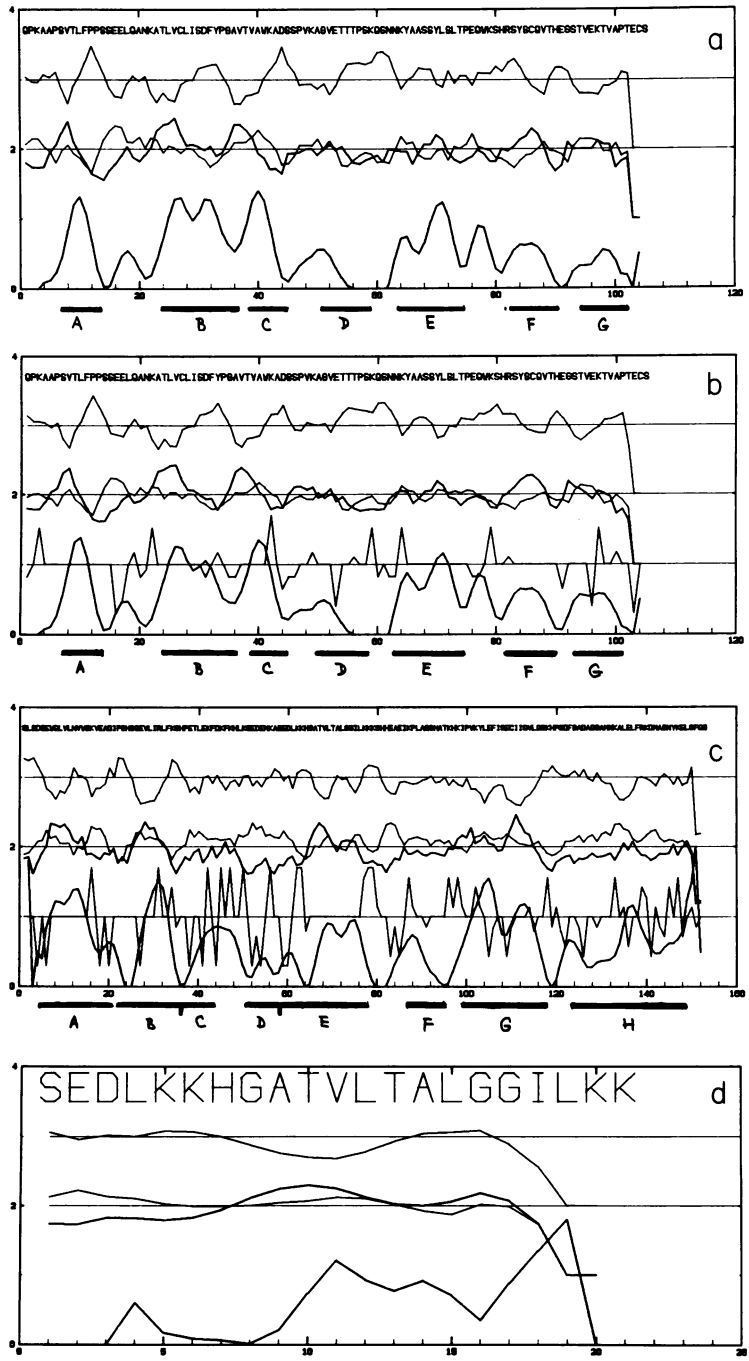
and display of α-helix, β-sheet and reverse turn propensities. The computation is based on Chou & Fasman (22) parameters and algorithm. *(iii)* computation, averaging and optional smoothing of hydrophobicity profiles. The method and parameters of Rose and Roy (11) are used with small modifications. *(iv)* construction and optional display of the profile of charged residues [lysine, arginine, aspartic and glutamic acid] along the amino acid sequence. *(v)* Display of the amino acid sequence, in one-letter code [IUPAC-IUB Tentative Rules 1968 (23)]. The program currently runs on a Digital VAX 11/780 computer and is interfaced with the publicly available United Graphics System of R.C. Beach (Computation Research Group, Stanford Linear Accelerator Center, Stanford CA 94305); this makes it possible to direct the output either to several graphics terminals (VT 125, Tektronix 4010, GIGI) or to an electrostatic plotter Versatec.

RESULTS AND DISCUSSION.

predictive procedure.

Our decision to use the Chou and Fasman (22) parameters is based on the fact that their algorithm is clearly and simply formulated and has proven itself as being one of the best available (24). In pilot experiments, parameters by Levitt (25) gave results virtually identical to those obtained with the Chou and Fasman propensities. Since the Chou and Fasman method often does not distinguish between an α-helix and a β-sheet, we included charge profiles into the program output. The charged side chains, and particularly the negatively charged acids, rarely occur in β-strands, especially in those located in the middle of pleated sheets. Also, charged side chains are reliable markers of solvent-exposed segments.

The user of the program is free to specify the number of curve-smoothing cycles desired. Thus, if only the number and relative location of major secondary structure elements is sought, one performes 3 to 6 hydrophobicity smoothing cycles [*cf.* Figures 1b,c]. If, on the contrary, one is interested in the periodicity of hydrophobicity within a single secondary structure, one smoothes the hydrophobicity profile only once [Figure 1d] or not al all. β-Strands, being shorter then α-helices and having a shorter periodi-


a
b
c
d
SEDLKKHGATVLTALGGILKK
A
B
C
D
E
F
G
H

city of hydrophobic side chain distribution, usually yield a single hydrophobicity maximum. α-Helices are often characterized by a more complicated pattern.

Figure 1 shows program outputs for standard all-β and all-α proteins, namely, constant domains of human, mouse, pig and rabbit immunoglobulin light chains and myoglobins from different species. The overall curve shapes in these two examples are different and typical for these two types of secondary structures: structures composed of antiparallel β-sheets display sharp spikes of reverse-turn and hydrophobicity profiles [Figure 1a,b] while α-helical myoglobins are characterized by relatively featureless turn profiles and more varied hydrophobicity profiles [Figure 1c]. Comparison of Figures 1a and 1b shows that averaging of amino acid sequences eliminates some ambiguitites. Note that Figure 1a would lead to incorrect assignment of the β-strands D,E,G as α-helices whereas the profiles computed as an average of 4 different primary structures [Figure 1b] give a correct result in all three cases although the assignment of the strand D remains rather ambiguous. Note also a conspicuous difference in distribution of charged residues [particularly the negative charges] relative to positions of α-helical segments and β-strands.

In attempting a secondary structure prediction, one starts by tentatively associating maxima of hydrophobicity profiles with centers of secondary structure segments. When a peak in hydrophobicity profile coincides both with *(i)* a maximum of α-helix and/or β-sheet propensity and *(ii)* a minimum of reverse-turn propensity, an α-helix or a β-strand is strongly suggested. Ambiguous secondary structure assignments [either sheets or helices] can be resolved by an *ad hoc* application of Chou and Fasman (22) rule 3 [helix and sheet boundary analysis], the helical wheel procedure (12) and a closer examination of simple-smoothed hydrophobicity profiles [*cf.* an example in Figure 1d]. The fundamentally different charge profiles of the two different types of struc-

FIGURE 1. Secondary structure prediction profiles. Curves, from top to bottom: [*1*] reverse turn propensity; [*2*] α-helix [light line] and β-sheet [heavy line] propensity; [*3*] electric charges (positive charges point up, negative charges point down); [*4*] hydrophobicity smoothed by 3 cycles of the moving-window algorithm [see text for details]. Elements of secondary structure, as determined by X-ray crystallography (41,42) are indicated at the bottom. **(a)**: constant domain of human myeloma λ chain NEW (43). Note that the charge profile was omitted from the plot. **(b)**: composite profiles obtained by averaging amino acid sequences of 4 different light chain constant domains, namely, pig λ chains (44), human myeloma λ chain NEW, mouse myeloma λ chain M315 (45) and rabbit κ chain of b9 allotype (46). **(c)**: composite profiles obtained by averaging amino acid sequences of human (47), bovine (48) opossum (49), dolphin (50) and chicken (51) myoglobins. The amino acid sequence displayed at the top is that of human myoglobin. **(d)**: detailed hydrophobicity pattern in α-helix E of human myoglobin. The amino acid residues 58-78 from the human myoglobin primary structure file (21) were read into the program and all the profiles were smoothed by 1 cycle of the moving-window algorithm. Note a succession of two-point hydrophobicity maxima [bottom line], characteristic for α-helical segments.

tures [*i.e.*, β-sheet and α-helical proteins] could also be helpful. It should be remembered, however, that secondary structures with "exceptional" features will always be present. An example is the strand "D" in immunoglobulin domains [Figure 1a,b] which occurs on the edge of the β-sheet and gives rise to unsignificant hydrophobicity and β-sheet propensity peaks.

When dealing with homologous proteins likely to share the same fold, idiosyncrasies like the one mentioned above can often be turned into an advantage. One can assume that exceptional features which occur at identical positions in different amino acid sequences further strengthen the starting assumption of structural similarity.

In summary, the program output provides the worker with a concise overview of a number of amino acid sequence properties (profiles) thought to be important in determining the secondary structure. A rational combination of these and the final decision about positions of secondary structure elements will always involve a certain amount of critical judgement and, in this sense, will remain partial and subjective. This point is essential. Examples demonstrate that current predictive methods are inherently limited and that this shortcoming can only be overcome by a thoughtful application of additional structural knowledge, for example, the general rules of protein topology and anatomy [*e.g.*, right-handedness of crossover connections (26,27), types of connections associated with parallel, as opposed to anti-parallel sheets (28) *etc.*]. Accordingly, we do not present a quantitative assesment of the predictive performance of the program (*cf.* 22) although it is obvious that, being a combination of different procedures, the program has a potential of scoring better that any of its components.

histocompatibility antigens.

As an illustration of program application, we present an analysis of amino acid sequences from human and murine histocompatibility antigens. These antigens are molecules associated with cell surfaces and their genetically-determined heterogeneity plays an important role in immune response [reviewed in (29,30)]. The major histocompatibility gene complex [known as HLA in man and H-2 in mouse] encodes a variety of molecules. The class I molecules [HLA-A,B,C and H-2-K,D,L] consist of an α chain [44,000 dalton] and a β2-microglobulin, a 12,000 dalton polypeptide chain encoded outside the MHC locus. The class I antigens are expressed on virtually all nucleated cells. The class II antigens [products of the immune response genes HLA-DR, DC, SB in man and I-A, I-E in mouse] are expressed primarily on B lymphocytes and are composed of two chains, α and β, of molecular masses 34,000 and 29,000, respectively.

Primary structure of several class I antigens and β2-microglobulins has been determined suggesting that the polypeptide chains are organized into domains approximately 90 amino acid residues long. Three such domains compose the extracellular

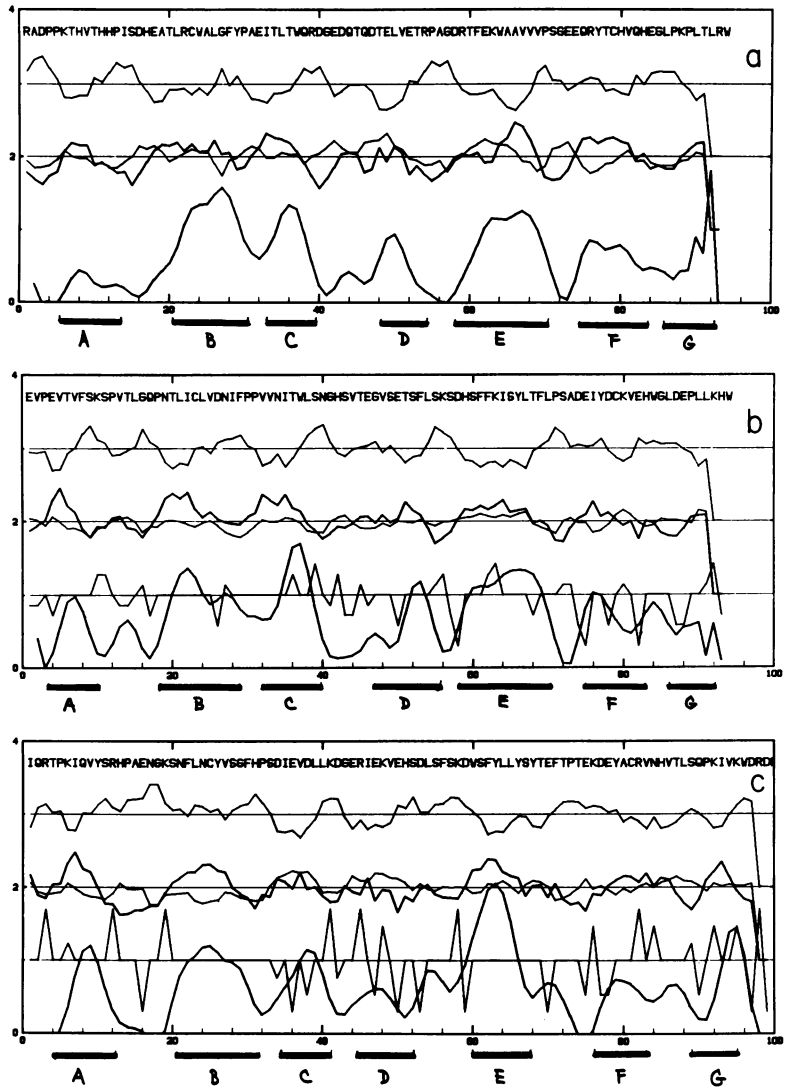
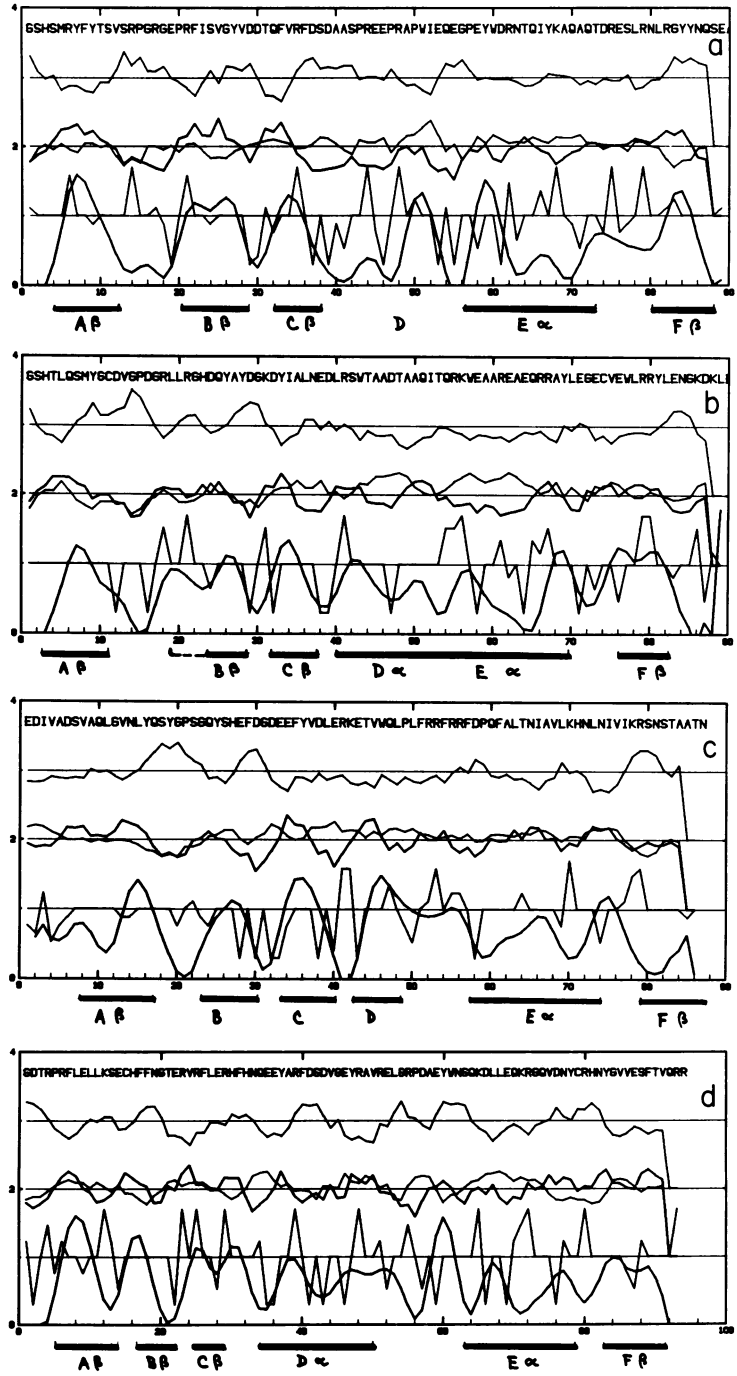


FIGURE 2. Cumulative secondary structure profiles of β2-microglobulins and immunoglobulin-like domains from class I and class II histocompatibility antigens. The seven β-strands identified are labeled sequentially A to G in accordance with the immunoglobulin domain nomenclature of Lesk & Chothia (37). **(a)**: The α3 domains of following class I heavy chains were averaged: human HLA B7 (52), murine H-2 Kb (53) and murine Qa genomic clone 27.1 (54). The amino acid sequence displayed at the top is that of human HLA B7. **(b)**: The domains of following class II heavy and light chains were averaged: human HLA-DR α2 (35,55), human HLA-DC α2 (36), murine H-2 I-E α2 (56), human HLA-DR β2 (57) and human HLA-DR-like β2 (34). The amino acid sequence displayed at the top is that of human class II HLA-DC α2 domain. **(c)**: The following β2-microglobulins were averaged: human (31), murine (58), rabbit (59) and guinea pig (60). The amino acid sequence displayed at the top is the human β2-microglobulin.


GSHSMRYFYTSVSRPGRGEPRFISVGYVDDTQFVRFDSDAASPREEPRAPWIEQEGPEYWDRNTQIYKAQAQTDRESLRNLRGYYNQSE
a
A β
B β
C β
D
E α
F β
GSHTLQSHYGCDVGPDGRLLRGHDQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGECVEWLRRYLENGKDKL
b
A β
B β
C β
D α
E α
F β
EDIVADSVAQLGVNLYQSYGPSGQYSHEFDGDEEFYVDLERKETVWQLPLFRRFRRFDPQFALTNIAVLKHNLNIVIKRSNSTAATN
c
A β
B
C
D
E α
F β
GDTRPRFLELLKSECHFFNGTERVRFLERNFHNQEEYARFDSDVGEYRAVRELGRPDAEYWNSQKDLLEQKRGQVDNYCRHNYGVVESFTVQRR
d
A β
B β
C β
D α
E α
F β

part of class I α chains, whereas the β2-microglobulin constitutes a single domain. Amino acid sequences of homologous domains [*e.g.*, α1 of human and mouse] typically share 60% to 80% identical residues, strongly suggesting that they have the same three-dimensional fold. Rapid advances in recombinant DNA technology resulted in accumulation of amino acid sequence data of class II α and β chains. Similar to the class I antigens, the class II polypeptide chains are organized into two extracellular domains. Since a full understanding of the three-dimensional structure of the MHC antigens would be extremely valuable both from the functional and from the evolutionary point of view, we attempted to deduce secondary structures of all the eight domain types. When analyzed separately, the individual primary structures yielded rather confusing patterns but profiles obtained from averaged properties of homologous domains provided us with considerably better results.

The three domain types termed "last" [α3 of class I, α2 and β2 of class II], as well as β2-microglobulin domains, were previously recognized by several authors as homologous to immunoglobulins (31-36). Comparison of averaged secondary structure propensities of these domains [Figure 2] with those of immunoglobulin light chain constant domains [Figure 1b] strongly suggests that the HLA and H-2 "last" domains, as well as the β2-microglobulin, consist of seven antiparallel β-strands [A through G in the nomenclature of Lesk and Chothia (37)] organized into two antiparallel sheets ABED and CFG. The overall similarity of the averaged profiles from immunoglobulin and histocompatibility antigen domains [Figures 1b, 2a, 2b and 2c] is striking and leaves little doubt about the close structural homology. Note for example that the β-strand D, which in immunoglobulin domains occurs on the edge of the sheet and therefore gives an atypical β-propensity pattern, is unusual in the histocompatibility domains as well. Lesk and Chothia (37) identified 17 β-sheet positions occupied by typical, conserved residues in all the constant-type immunoglobulin domains. Those residues often appear in homologous positions of the "last" domains as well. Cohen et al. (38) previously delineated the seven β-strands of the α3 domain of HLA B7 heavy chain and β2-

FIGURE 3. Cumulative secondary structure profiles of various class I and class II histocompatibility antigen domains. **(a)**: The α1 domains of the following class I antigens were averaged: human HLA B7 (52), human HLA 12.4 (61), murine H-2 Kb (53), murine H-2 Db (62), murine H-2 Kd (63) and murine Qa genomic clone 27.1 (54). The amino acid sequence at the top is that of human HLA B7. **(b)** The α2 domains of the following class I antigens were averaged: human HLA B7, human HLA 12.4, murine H-2 Kb and murine genomic clone Qa 27.1. The amino acid sequence at the top is that of HLA B7. **(c)** The α1 domains of the following class II antigens were averaged: human HLA-DR (35), human HLA-DC (36,64), murine H-2 I-E (65) and I-A (56). The amino acid sequence at the top is that of human HLA-DC. **(d)** The β1 domains of the following class II antigens were averaged: human HLA-DR (57), human HLA-DC (34) and human HLA-SB-like (66). The amino acid sequence at the top is that of human HLA-DR. β-strands and α-helices are labeled sequentially from A to E.

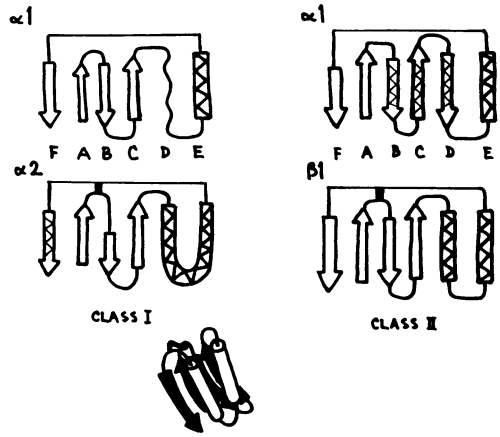


FIGURE 4. Connectivity diagrams of putative secondary structures for the histocompatibility antigen domains. Arrows denote β-strands, rectangles with zig-zag lines α-helices; arrows with zig-zag lines denote secondary structures the assignment of which remains ambiguous [*i.e.*, can be either β-strands or α helices]. The final assignments of α-helices were done after examining the putative α-helical segments and some of the β-strands by the helical wheel method (12). Secondary structure elements are labeled sequentially A to F from the N-terminus. Note that the segment D in class I α1 domains might consist from a series of short, irregular helices [*cf.* Figure 3a]. Disulfide bridges are indicated as short heavy bars. Bottom: it is proposed that all the domains fold into "open-face sandwiches" (28) with a single layer of predominantly antiparallel β-sheet, one side of which is covered by α-helices and/or structureless segments. The proposition is based on the following facts and presumptions: *(i)* the two cysteine residues of class I α2 and class II β1 domains form a disulfide bond, *(ii)* the β-strands form a single antiparallel sheet, *(iii)* all the crossover connections among secondary structure elements must be right-handed (26,27) and *(iv)* α-helices pack to β-strands in an aligned manner (67,68).

microglobulin with use of their combinatorial algorithm. The secondary structure proposed here is identical to that proposed by them, except for differences of one or two residues at the N- and C-termini of β-strands.

The remaining domains [α1 and α2 of class I; α1 and β1 of class II] do not show significant homology to any of the immunoglobulin domains, although some of them [α2 of class I and β2 of class II] possess a disulfide loop reminiscent of that which typifies the immunoglobulin fold. Amino acid alignments suggested homology among both the class I domains and β1 of class II [of about 20% identical amino acid residues, 40% chemically and conformationally homologous residues]. At the same time, the α1 domains of class II stood apart from the others [only 12% identical amino acid residues when compared to class II β1].

The secondary structural patterns of the three mutually homologous domain types [α1 and α2 of class I, β1 of class II; *cf.* Figure 3] are similar. The N-terminal parts seem

to consist of three β-strands [there are three distinct β-sheet maxima coinciding with strong hydrophobicity peaks and pronounced reverse turn minima] whereas the middle parts are likely to fold into one or two α-helical segments [note the persistence of unusually high α-helical propensity over 30-40 residues combined with relative absence of sharp reverse turn maxima]. This arrangement is suggestive of "open-face sandwiches", *i.e.* structures which "have a single, more or less twisted β-sheet, either pure anti-parallel or predominantly so, but not closing around to form a barrel (with) a layer of helices and loops which covers only one side of the sheet, so that they are two-layer structures" (28). It is quite unlikely that the relatively small histocompatibility domains would fold in parallel α/β fashion: parallel sheets are known to be multistranded, surrounded by many helices and occuring in large domains of an average size of 150 amino acid residues [135 residues being the minimum length, Richardson (28)]. Open-face sandwiches, on the other hand, "range from 3 to 15 strands with a wide assortment of topologies, curvatures, and placement of helices and loops" (28). Indeed, connectivity diagrams of the histocompatibility domains [Figure 4] indicate variability in α-helix and β-strand lengths. Very similar folding themes are encoutered in open-face sandwiches of subtilisin inhibitor from *Streptomyces* or in the third domain of glutathione reductase (28). It is tempting to speculate on possible relationship among the suggested structural variability of the domains, the proneness of open face sandwiches to vary, and biological significance of these variations. Until the X-ray crystallography of class I antigens [P.Bjorkman and D.Wiley, Harvard University, work in progress] either confirms or rejects the folds proposed here, however, any such speculation would be premature. So far, it seems that the percentage of secondary structure predicted [40% to 45% of β-sheet and 10% to 15% α-helix in all the polypeptide chains] does not contradict data obtained by circular dichroism measurements on class I antigens (39,40).

Class II α1 domain is more difficult to classify than the other three, mostly due to an uncertainty of its second N-terminal secondary structure element [either α-helix or a β-strand] and the two which follow. Nonetheless, the overall appearance of its profile is similar to those of the other domains [*i.e.*, three hydrophobicity maxima at the N-terminus coincident with peaks of β-sheet propensity and with minima of reverse turns; followed by a long region of elevated α-helix propensity associated with a flat reverse turn profile] and it seems reasonable to assign it tentatively to the same tertiary structure type as the other three [*cf.* Figure 3].

ACKNOWLEDGEMENT

We thank John Newell, Head, Cardiac Computer Center, for generous support and

Professors M.Karplus and J.Strominger (Harvard University) for valuable comments. The computer program is available at request from one of the authors (J.N.).

REFERENCES.

1. Blout,E.R., de Lozé,C., Bloom,S.M. & Fasman,G.D. (1960). *J.Am.Chem. Soc.* **82**, 3787-3789.
2. Davies,D.R. (1964). *J.Mol.Biol.* **9**, 605-609.
3. Finkelstein,A.V. & Ptitsyn,O.B. (1971). *J.Mol.Biol.* **62**, 613-624.
4. Dirkx, J. (1972). *Arch. Int. Physiol. Biochim.* **80**, 185-187.
5. Nagano,K. (1972). *J.Mol.Biol.* **75**, 401-420.
6. Nagano,K. (1974). *J.Mol.Biol.* **84**, 337-372.
7. Robson,B. (1974) *Biochem.J.* **141**, 853-867.
8. Chou,P.Y. & Fasman,G.D. (1974). *Biochemistry* **13**, 222-244.
9. Burgess,A.W. & Scheraga,H.A. (1975).*Proc.Nat.Acad.Sci.* **72**, 1221-1225.
10. Zull,J.E. & Lev,N.B. (1980). *Proc.Nat.Acad.Sci.* **77**, 3791-3795.
11. Rose,G.D. & Roy,S. (1980). *Proc. Nat. Acad.Sci.* **77**, 4643-4647.
12. Schiffer,M. & Edmundson,A.B. (1968). *Biophys.J.* **8**, 29-39.
13. Cid,H., Bunster,M., Arriagada,E. & Campos,M. (1982) *FEBS Letters* **150**, 247-254.
14. Rose,G.D. (1978). *Nature* **272**, 586-590.
15. Nozaki,T. & Tanford,C. (1971). *J.Biol.Chem.* **246**, 2211-2217.
16. Kuntz,I.D. (1972). *J.Am.Chem.Soc.* **94**, 4009-4012.
17. Lim,V.I. (1974). *J.Mol.Biol.* **88**, 873-894.
18. Ponnuswamy,P.K., Phrabhakaram,M. & Manavalan,P. (1980). *Biochim.Biophys.Acta* **623**, 301-316.
19. Schulz,G.E. & Schirmer,R.H. (1979). *in "Principles of Protein Structure"*, Springer-Verlag, Berlin, p.166.
20. Staden,R. (1977). *Nucleic Acid Res.* **4**, 4037-4051.
21. Dayhoff,M.O.(1972). *Atlas of Protein Sequence and Structure*, National Biomedical Science Foundation, Washington.
22. Chou,P.Y. & Fasman,G.D. (1978). *Adv. Enzymol.* **47**, 45-148. **150**, 247-254.
23. IUPAC-IUB Tentative Rules (1968). *Eur.J.Biochem.* **5**, 151-153.
24. Schulz,G.E., Barry,C.D., Friedman,J., Chou,P.Y., Fasman,G.D., Finkelstein, A.V., Lim,V.I., Ptitsyn,O.B., Kabat,E.A., Wu,T.T., Levitt,M., Robson,B. & Nagano,K. (1974). *Nature* **250**, 442-451.
25. Levitt,M. (1978). *Biochemistry* **16**, 4277-4285.
26. Richardson,J.S. (1976). *Proc.Nat.Acad.Sci.* **73**, 2619-2623.
27. Sternberg,M.J.E. & Thornton,J.M. (1976). *J.Mol.Biol.* **110**, 269-283.
28. Richardson,J.S. (1981). *Adv.Protein Chem.* **34**, 167-339.
29. Klein,J. (1979). *Science* **203**, 516-521.
30. Shackleford,D.A., Kaufman,J.S., Korman,A.J. & Strominger,J.L. (1982). *Immunol.Rev.* **66**, 133-187
31. Cunningham,B.A., Wang,J.L., Berggård,I. & Peterson,P.A. (1973). *Biochemistry* **12**, 4811-4816.
32. Orr,H.T., Lancet,D., Robb,R.J., López de Castro,J. & Strominger,J.L. (1979). *Nature* **282**, 266-270.
33. Feinsten,A. (1979). *Nature* **282**, 230.
34. Larhammar,D., Schenning,L., Gustafsson,K., Wiman,K., Claesson,L., Rask,L. & Peterson,P.A. (1982). *Proc.Nat.Acad.Sci.* **79**, 3687-3691.
35. Korman,A.J., Auffray,C., Schamboeck,A. & Strominger,J.L. (1982). *Proc.Nat.Acad.Sci.* **79**, 6013-6017.
36. Auffray,C., Korman,A.J., Roux-Rosetto,M., Bono,R. & Strominger,J.L. (1982) *Proc.Nat.Acad.Sci.* **77**, 6337-6341.
37. Lesk,A. & Chothia,C. (1982). *J.Mol.Biol.* **160**, 325-342.
38. Cohen,F.E., Sternberg,M.J.E. & Taylor,W.R. (1980). *Nature*

285, 378-382.
39. Lancet,D., Parham,P & Strominger,J.L. (1979). *Proc.Nat.Acad.Sci.* **76**, 3844-3848.
40. Trägårdh,L., Curman,B., Wiman,K., Rask,L. & Peterson,P. (1979). *Biochemistry* **18**, 2218-2226.
41. Saul,F.A., Amzel,L.M. & Poljak,R.J. (1978). *J.Biol.Chem.* **253**, 585-597.
42. Kendrew,J.C., Watson,H.C., Strandberg,B.E., Dickerson,R.E., Phillips,D.C. & Shore,V.C. (1961). *Nature* **190**, 666-670.
43. Chen,B.L. & Poljak,R.J. (1974). *Biochemistry* **13**, 1295-1302.
44. Novotný,J., Franěk,F., Margolies,M.N. & Haber,E. (1977). *Biochemistry* **16**, 3765-3772.
45. Dugan,E.S., Bradshaw,R.A., Simms,E.S. & Eisen,H.N. (1973). *Biochemistry* **12**, 5400-5416.
46. Farnsworth,V., Goodfliesh,R., Rodkey,S. & Hood,L. (1976). *Proc.Nat.Acad.Sci.* **73**, 1293-1296.
47. Romero-Herrera,A.E. & Lehmann,H. (1971). *Biochim.Biophys.Acta* **251**, 482-488.
48. Han,K., Dautreveaux,M., Chaila,X. & Bisêtre,G. (1970). *Eur.J. Biochem.* **16**, 465-471.
49. Romero-Herrera,A.E. & Lehmann,H. (1975). *Biochim.Biophys.Acta* **400**, 387-398.
50. Jones,B.N., Vigna,R.A., Dwulet,F.E., Bogardt,R.A., Lehman,L.D. & Gurd,F.R.N. (1976). *Biochemistry* **15**, 4418-4422.
51. Deconinck,M., Peiffer,S., Depreter,J., Paul,C., Schnek,A.G. & Leonis,J. (1975). *Biochim.Biophys.Acta* **386**, 567-575.
52. Orr,H.T., López de Castro,J., Lancet,D. & Strominger,J.L. (1979). *Biochemistry* **18** 5711-5720.
53. Coligan,J.E., Kindt,T.J., Uehara,H., Martinko,J & Nathenson,S.G. (1981). *Nature* **291**, 35-39.
54. Steinmetz,M., Moore,K.W., Frelinger,J.G., Sher,B.T., Shen,F.W., Boyse,E.A. & Hood,L. (1981). *Cell* **25**, 683-692.
55. Lee,J.S., Trowsdale,J., Travers,P.J., Carey,J., Grosveld,F., Jenkins,J. & Bodmer,W.F. (1982). *Nature* **299**, 750-752.
56. Benoist,C.O., Mathis,D.J., Kanter,M.R., Williams,V.E. & McDevitt,H.O. (1983). *Proc.Nat.Acad.Sci.* **80**, 534-538.
57. Kratzin,H., Yang,C.Y., Götz,H., Pauly,E., Kölbel,S., Egert,G., Thinnes,F.P., Wernet,P., Altevogt,P. & Hilschmann,N. (1981). *Hoppe-Seyler's Z.Physiol.Chem.* **362**, 1665-1669.
58. Gates,F.T., Coligan,J.E. & Kindt,T.J. (1981). *Proc.Nat.Acad.Sci.* **78**, 554-558.
59. Gates,F.T., Coligan,J.E. & Kindt,T.J. (1979). *Biochemistry* **18**, 2267-2272.
60. Wolfe,P.B. & Cebra,J.J. (1980). *Mol.Immunol.* **17**, 1493-1505.
61. Malissen,M., Malissen,B. & Jordan,B.R. (1982). *Proc.Nat.Acad.Sci.* **79**, 893-897.
62. Maloy,W.L., Nathenson,S.G., & Coligan,J.E. (1981). *J.Biol.Chem.* **256**, 2863-2872.
63. Kimball,E.S., Nathenson,S.G. & Coligan,J.E. (1981). *Biochemistry* **20**, 3301-3308.
64. Auffray,C., Lillie,J.W., Arnot,D., Grossberger,D., Kappas,D & Strominger,J. (1983). *Manuscript submitted for publication.*
65. McNicholas,J.M., Steinmetz,M., Hunkapillar,T., Jones,P. & Hood,L. (1982). *Science* **218**, 1229-1232.
66. Roux-Dosseto,M., Auffray,C., Lillie,J.W., Boss,J., Cohen,D., DeMars,R., Mawas,C., Seidman,J. & Strominger,J.L. (1983). *Proc.Nat.Acad.Sci.*, in press.
67. Janin,J. & Chothia,C. (1980). *J.Mol.Biol.* **143**, 95-128.
68. Cohen,F.C., Sternberg,M.J.E. & Taylor,W.R. (1982). *J.Mol.Biol.* **156**, 821-862.