The EMBO Journal vol.5 no.2 pp.409–413, 1986

Location of 'continuous' antigenic determinants in the protruding regions of proteins

J.M.Thornton, M.S.Edwards, W.R.Taylor and D.J.Barlow

Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

Communicated by D.C.Phillips

A simple method is described to locate 'antigenic' peptides from the α-carbon co-ordinates of a protein, based on protrusion from the protein's globular surface. A good correlation is found between those parts of a protein which protrude and the experimentally determined antigenic peptides in myoglobin, lysozyme and myohemerythrin. A comparison is made between the use of protrusion index, mobility, solvent accessibility and hydrophilicity for predicting the most likely antigenic peptides.

Key words: antigenicity/peptides/prediction/protrusion index/continuous determinants

Introduction

Although the immune response to a foreign protein is complex and incompletely understood, it has been shown that short peptides taken from a protein will cross-react with antibodies raised against the complete protein (Atassi, 1984; Milton and Van Regenmortel, 1979). Furthermore it has been proposed that some peptides can be used to raise anti-peptide antibodies, which recognize the complete protein and could therefore be used to generate synthetic vaccines (Lerner, 1982). Two recent publications have emphasized the importance of peptide flexibility (Westof *et al.*, 1984; Tainer *et al.*, 1984) and have shown that in the protein the regions corresponding to the antigenic peptides are usually highly mobile. An alternative viewpoint is that these regions (which we will term 'peptide determinants') are antigenic because they are the most accessible to the large antibody molecule. In this paper we show that for lysozyme, myoglobin and myohemerythrin, the experimentally determined peptide determinants correspond predominantly to those parts of the structure which protrude, as defined by generating an equi-momental ellipsoid to fit the protein and calculating those segments which 'stick out'. The correlation with the experimental data is superior to that obtained using hydrophilicity (Hopp and Woods, 1981) and comparable with the results presented on mobility (Westof *et al.*, 1984; Tainer *et al.*, 1984, 1985).

To test the hypothesis that peptide determinants protrude from the surface of the protein, it was necessary to construct an approximation to that surface. Since most proteins are distinctly ellipsoidal in shape (Taylor *et al.*, 1983; Prabhakaran and Ponnuswamy, 1982), the ellipsoid was calculated which has the same moments of inertia as the protein structure (the equi-momental ellipsoid), using a standard analytical method involving Cauchy's Momental ellipsoid (Taylor *et al.*, 1983). This method determines the ratios between the lengths of the principal axes *a, b, c* and their directions, but the absolute size of the ellipsoid is arbitrarily chosen to include a specified percentage of atoms. For example, the 90% ellipsoid will include 90% of the atoms, with 10% lying outside or protruding from the globular shape. Therefore we can simply assign to each residue a 'protrusion index' or PI specifying the % ellipsoid at which that residue first becomes external. For example, all residues which are outside the 90% ellipsoid are assigned a PI=9; those which are outside the 80% ellipsoid (but not the 90% ellipsoid) have PI=8, etc. Thus for each protein the ratios *b/a* and *c/a*, and the directions of the principal axis, are calculated from the co-ordinates (all atoms or just α-carbon co-ordinates can be included). Then the 90%, 80%, 70%, etc. ellipsoids are generated and the PI assigned to each residue. Ellipsoids can be generated for each domain of a multi-domain protein or for the whole structure.

The method was applied to three proteins for which the α-carbon co-ordinates were available as well as some evidence to

Table I. Averaged values of PI, accessibility, mobility and hydrophilicity calculated over the length of each peptide

Protein and peptide	Protrusion index		Accessibility		Mobility		Hydrophilicity	
	PI	%	ACC	%	B	%	H	%
Myoglobin								
1–6	6.8	18	77	12	17	5	0.5	35
15–22	5.5	32	56	42	12	23	0.6	30
56–62	6.9	16	65	28	8	73[a]	1.4	3
94–99	4.2	36	75	14	12	19	0.5	38
113–119	3.6	64[a]	55	46	14	12	−0.2	76[a]
121–127	7.4	8	50	52[a]	15	8	0.4	42
145–151	8.3	2	85	5	13	14	0	68[a]
Lysozyme								
38–54	4.6	43	52	46	15	82[a]	0.19	25
64–80	5.6	15	64	10	25	9	0.17	30
Myohemerythrin								
High[b]								
3–16	5.6	12			31	12	0.16	52[a]
7–16	4.8	40			29.5	34	−0.19	78[a]
37–46	5.8	18			29.5	33	0.35	35
57–66	6.5	9			29	43	0.74	11
Medium[b]								
63–72	7.2	1			32	7	0.3	38
69–82	5.3	21			30	26	0.6	11
73–82	4.1	58			28.6	48	0.84	8
Low[b]								
22–35	2.1	98			22	95	0.78	4
26–35	1.7	98			21.7	92	0.24	44
42–51	3.3	73			29	39	−0.5	92
96–109	2.6	90			22	93	−0.11	86
100–109	2.7	80			21	98	0	65

See legend to Figure 1 for data sources.
The % columns represent the percentage of peptides, of equal length to the epitope, with higher average values than the epitope. For example, 18% of hexapeptides in myoglobin have PI ≥ 6.8
[a]Active peptides where percentile is >50%.
[b]For definition of high, medium and low activity see legend to Figure 1c.

define the location of the peptide determinants (Atassi, 1984; Milton and Van Regenmortel, 1979; Westof *et al.*, 1984; Tainer *et al.*, 1984, 1985) (see Figure 1). A graphics algorithm (ELLIPSE) was also used to display and rotate the protein and its equi-momental ellipsoid (Figure 2a). The peptides are identified either by successful competition with protein-antibody complex formation, or by raising antipeptide antibodies, whose interaction with the complete protein was determined (see legend to Figure 1).

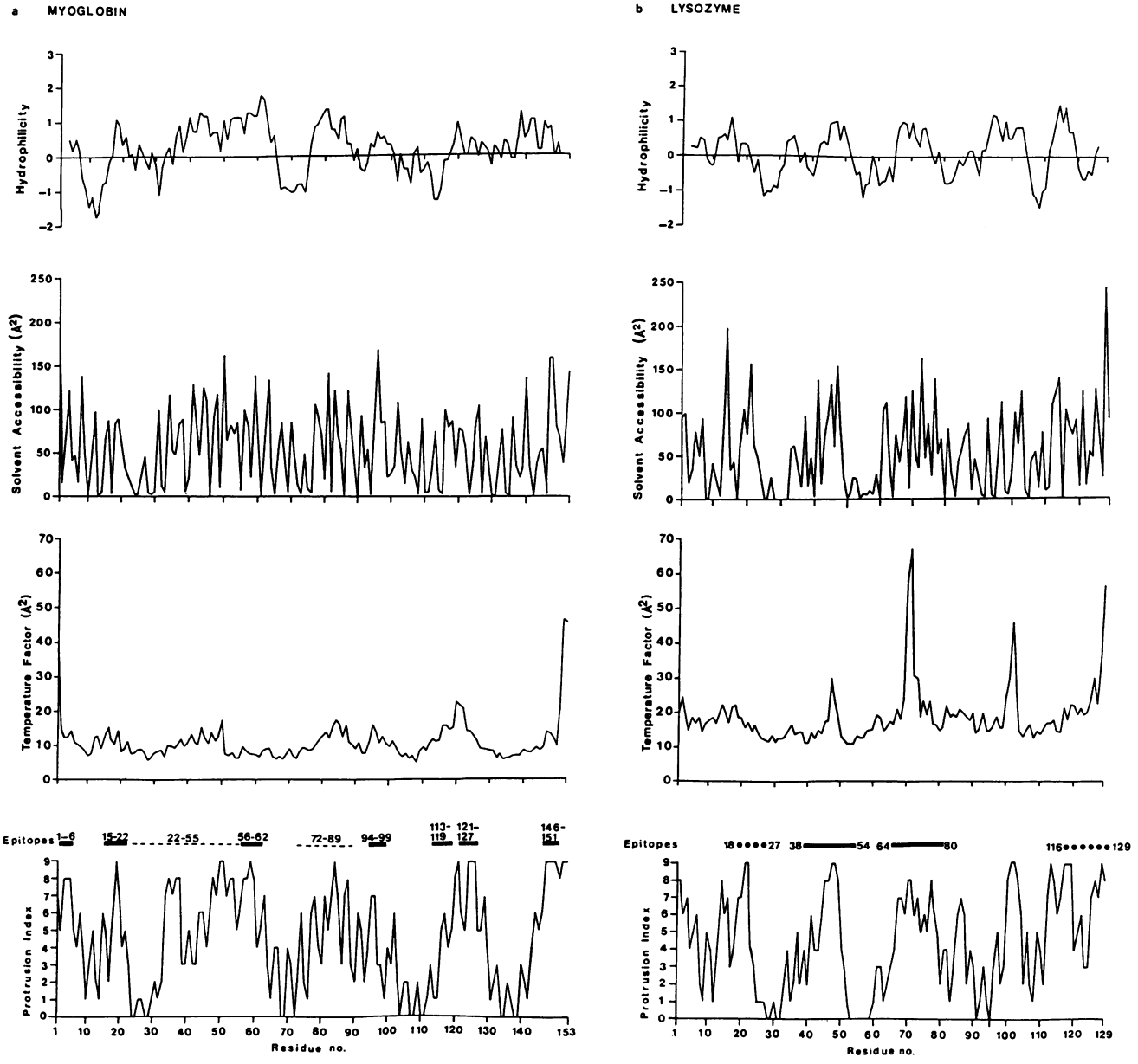
Results

Correlation between protrusion index and the location of the peptide determinants

For each protein a plot is made of the 'protrusion index' (PI) against the residue number, as shown in Figure 1a–c. Also shown are the variations along the chain of solvent accessibility (Lee and Richards, 1971), hydrophilicity (Hopps and Woods, 1981) and mobility (B-values) (Westof *et al.*, 1984; Tainer *et al.*, 1984). Table I gives the averages of these four variables and a ranked percentile (see below) calculated for each peptide determinant.

For all the proteins there is a marked correlation between high PI values and the location of the 'antigenic peptides'. The plots suggest that most of these peptides have PI >5, and often correspond to a peak in the PI plot (see Table I). For myoglobin, all the antigenic peptides, except perhaps residues 113–119, correspond to a peak in the PI plots (Figure 1a). All include at least one residue with PI >7, and four of the seven peptides include the most protrusive residues (PI = 9). For lysozyme (Figure 1b) two of the protruding segments correspond almost exactly to the two identified continuous antigenic regions (residues 38–54 and 64–80). Even more interestingly the low resolution crystal structure of the lysozyme-antibody complex shows that residues 18–27 and 116–128, two of the other external segments, form the non-continuous determinant which interacts with the antibody (Amit *et al.*, 1985).

For myohemerythrin the data define which peptides are effective in raising antibodies which interact with the whole protein. There is a clear correlation between the active peptides and high



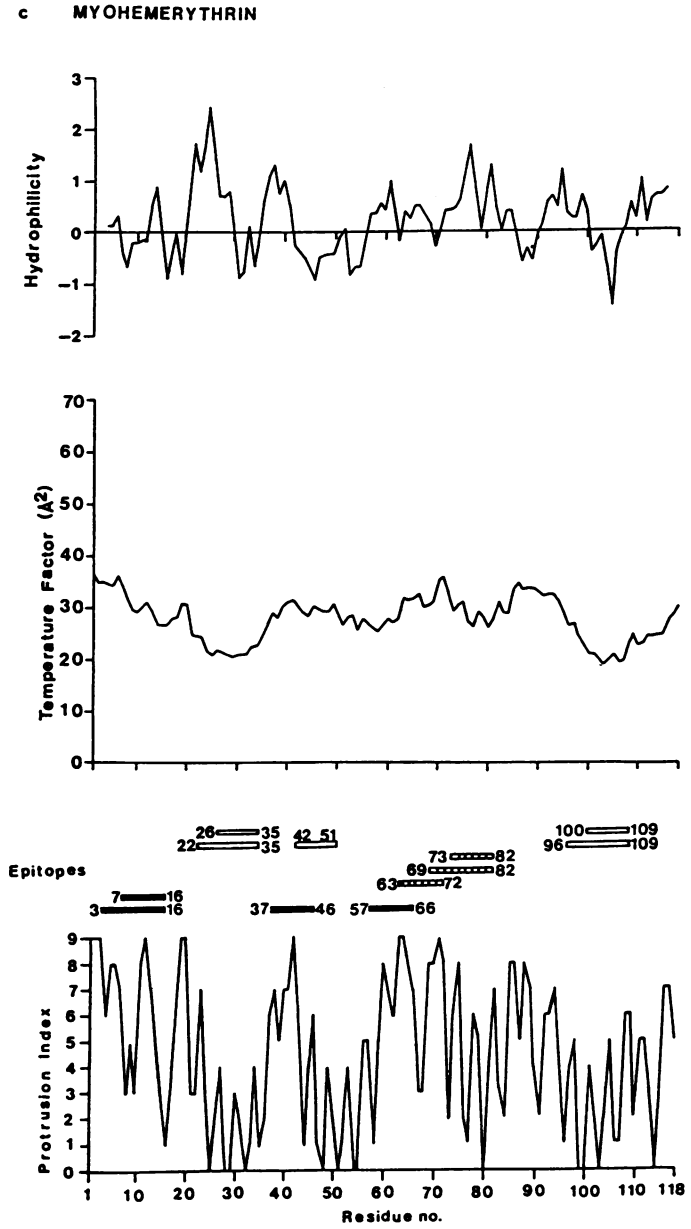


Fig. 1. Variation of the hydrophilicity, solvent accessibility, mobility and protrusion index along the sequence for myoglobin, lysozyme and myohemerythrin. For each protein the equimomental ellipsoid is generated as described in Taylor *et al.* (1983) and the PI is calculated as described in the text. The residue accessibility is calculated using the method of Kabsch and Sander (1983) with their Pascal program which is available on the Protein Data Bank (Bernstein *et al.*, 1977). The hydrophilicity values are calculated using the hexapeptide averaging method of Hopp and Woods (1981). This averaging smooths out the plot, compared with the other variables, which are not averaged. The data for the location of the 'antigenic' peptides are taken from Young *et al.* (1983); Leach (1983), Ibrahimi *et al.* (1980); Atassi (1984) and Milton and Van Regenmortel (1979) and they are indicated by solid bars above the PI plot in **a** and **b**. **(a)** Myoglobin. Co-ordinate data and B-values from the Protein Data Bank file 1MBD. The PI variation is calculated from α-carbon atom co-ordinates. The dashed line indicates an epitope of uncertain location. Peptides 1–6 and 121–127 do not compete with whole protein–antibody complex formation, but are effective in raising antibodies which recognize the whole protein (Young *et al.*, 1983). **(b)** Hen egg white lysozyme. Co-ordinate data from Protein Data Bank file 2LYZ. The PI plot is calculated from α-carbon co-ordinates. The dotted line indicates portions of a discontinuous epitope identified by crystallography (Amit *et al.*, 1985). B-values are from Berthou *et al.* (1983). **(c)** Myohemerythrin. Co-ordinate data from Protein Data Bank file 1MHR. The B-values are taken from Sheriff *et al.* (1985) and are corrected for crystal contacts. The PI plot is calculated from α-carbon co-ordinates. The data for all the atoms were not available and therefore the solvent accessibility plot could not be calculated. As described in Tainer *et al.* (1984) 12 peptides were synthesized and their reactivities, defined from immunoprecipitation results, are indicated as follows: ▬ the most active peptides (>4000 c.p.m.); ▭ medium-activity peptides (200–4000 c.p.m.); ▭ low-activity peptides (<200 c.p.m.).

PI values. The four most active peptides (3–16, 7–16, 37–46 and 57–66), as determined from the immunoprecipitation results, are also the regions which protrude farthest from the molecular surface, including the residues with PI = 9. The 'cold' non-active peptides (22–35, 26–35, 96–109 and 100–109) are the least protrusive ($PI_{AV} < 3$) despite being partially accessible to solvent as helical regions. This plot shows that the N-terminal peptide (1–6) is very protrusive, as has been found for many proteins (Thornton and Sibanda, 1983), and therefore predicts that it should be effective in raising antibodies to the whole protein.

The PI plots derived using all the atoms, instead of just the α-carbons, are broadly comparable (correlation coefficient =

Table II. Correlations[a] between protrusion index, accessibility, mobility and hydrophilicity plots for myoglobin

	PI[b]	Accessibility[c]	Mobility[d]	Hydrophilicity[e]	Antigenicity[f] (χ^2)
PI	1.0	0.63	0.46	0.37	28.5
Accessibility		1.0	0.40	0.64	8.6
Mobility			1.0	0.07	33.2
Hydrophilicity				1.0	0.76

[a]The correlation coefficient is calculated using the standard formulation. Its value varies from −1 (anti-correlation) to +1 (complete correlation).
[b]The PI is calculated from α-carbon co-ordinates as described in the text.
[c]The accessibility is the accessible surface area for each residue, calculated from all-atom co-ordinates using the method of Kabsch and Sander (1983).
[d]Mobility is defined by the crystallographic B-values, obtained from highly refined protein structures (see text).
[e]Hydrophilicity is calculated from the sequence using the method of Hopp and Woods (1981).
[f]The χ^2 values are from 2 × 2 contingency tables of each parameter against the observed location of antigenic residues. Each residue is scored 1 or 0 for its antigenicity (antigenic = 1; non-antigenic = 0). Similarly, each residue is scored 1 or 0 according to whether its parameter value is greater than or less than a threshold, set at the overall average for that parameter (so that equal to or greater than = 1; less than = 0). The larger the value of χ^2 the more the parameter differentiates between antigenic and non-antigenic residues.

0.84) but less sensitive. The advantage of the α-carbon plots is that they highlight segments of chain which protrude, regardless of side-chain conformation, which will almost certainly be flexible in solution. Nevertheless the ellipsoid is clearly a gross approximation to the complex surface of the protein and the rather good agreement suggests that the peptide determinants must often be located on gross protrusions.

Comparison of protrusion index with other parameters

We have attempted to make a rigorous comparison between the use of PI, mobility, solvent accessibility and hydrophilicity for prediction by calculating for each epitope its rank position on the different scales (see Table I). For a given epitope, the average PI is calculated for each residue segment of the same length in the sequence and then the % of segments with PI higher than that calculated for the epitope is given. This procedure is repeated for accessibility, mobility and hydrophilicity. For example, for residues 1−6 in myoglobin, 12% of 6-residue segments have a higher accessibility, 5% have higher mobility, 35% higher hydrophilicity and 18% higher PI. Although this is sensitive to the length of the peptides, which are not well defined, it provides a quantitative comparison of the parameters.

Inspection of Tables I and II indicates that PI, mobility and averaged accessibility have broadly comparable success rates,

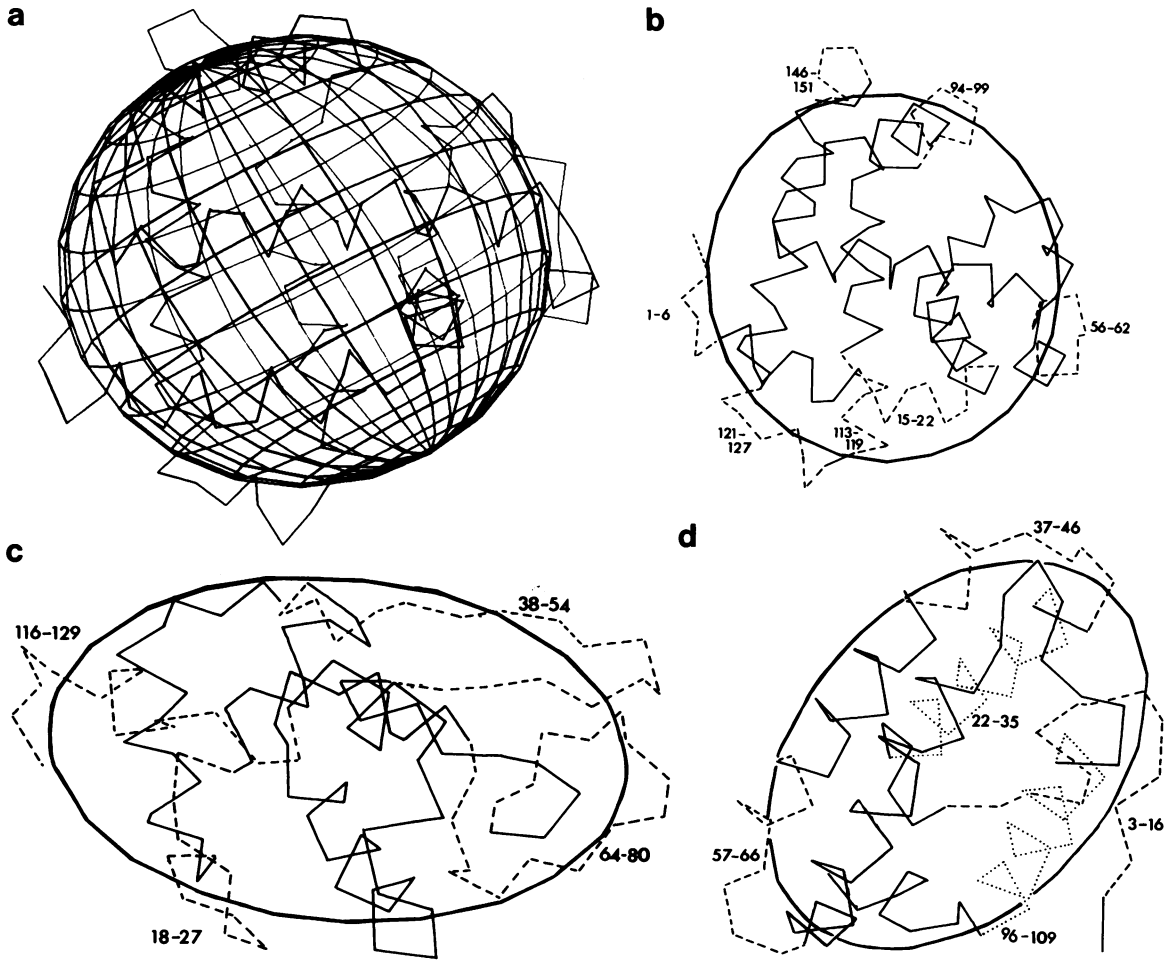


Fig. 2. Plots of the protein backbone showing location of the continuous antigen peptides (dashed) and the ellipsoid with ~50% atoms inside, derived from α-carbon co-ordinates. These figures are plotted using output from the program ELLIPSE which allows the structure and its ellipsoid to be rotated in three dimensions using the tracker ball. For each protein the ratios of the ellipsoid axis lengths *a*, *b*, *c* are given below. **(a)** Myoglobin and its three-dimensional ellipsoid as seen on the Evans and Sutherland PS2 graphics system, where colours (not shown here) are used to differentiate epitopes. **(b)** Myoglobin. b/a=0.9, c/a=0.5. **(c)** Lysozyme. b/a=0.6, c/a=0.6. The discontinuous epitope (18−27; 116−129) is also shown. **(d)** Myohemerythrin. b/a=0.6, c/a=0.5. The cold 'non-active' peptides (residues 22−35 and 96−109) are shown dotted.

and whilst all are better than the sequence-derived hydrophilicity index, none is definitive. For each measure there is at least one epitope which is not recognized (indicated by superscript a in Table I). However, the broad agreement underlines the correlation between these variables (see Table II). For myoglobin the correlation coefficients between the PI, accessibility and mobility range from 0.4 to 0.63. Surprisingly, there is little correlation between hydrophilicity and mobility. In general, though not always, the peptide determinants which protrude are hydrophilic with higher than average accessibilities, and are mobile. For many of the peptides the percentile values are quite high using any of the measures. This suggests that the specific sequence must be important as well as the protrusion, hydrophilicity, etc. The relative importance of flexibility versus antibody accessibility is not resolved and needs more well-defined immunological and crystallographic data.

Discussion

Although the whole surface of a protein is probably immunogenic (Benjamin *et al.*, 1984), most of the peptide determinants protrude from the globular surface. This observation can be most easily explained in terms of continuous and discontinuous determinants, rather than a generalized shape of the antibody combining site (Davies and Metzger, 1983). Apart from the protruding loop regions, the majority of the surface patches on a protein will comprise amino acids from residues distant in the linear sequence. Given the large area of contact found in most protein–protein interactions and expected for antibody–antigen contacts (Amit *et al.*, 1985), all the epitopes will probably be discontinuous to some extent. However, most of the experimental methods used to define peptide determinants inevitably locate only those epitopes where the primary site of interaction, essential for recognition, derives from sequential amino acids. These segments will almost always be protruding loops, since elsewhere amino acids distant in the sequence will be in close proximity.

In conclusion, the calculation of PI provides a simple method for locating the protruding parts of a protein from the α-carbon co-ordinates. Most of the peptide determinants correspond to protrusions, and can be easily located from the three-dimensional structure of the antigen. We are currently developing a simple method, similar to the hydrophilicity plots (Hopp and Woods, 1981), to predict protrusion from sequence. With regard to discontinuous determinants, these protruding loops are highly available for interaction with an antibody and are the sites of maximum sequence variation. Therefore one might expect them to make a major contribution to the antibody recognition response. [See for example the recent study of the rhinovirus (Rossmann *et al.*, 1985).] However, it may be that they are too flexible or too polar to make the best recognition sites for an antibody. Data to answer these questions should be available in the near future.

Acknowledgements

We thank Mr Freddie Elliott for excellent photography, and the SERC for financial support. M.S.E. holds a CASE Studentship in collaboration with Dr Peter Murray-Rust at Glaxo.

References

Amit,A.G., Mariuzza,R.A., Philips,S.E.U. and Poljak,R.J. (1985) *Nature*, **313**, 156–158.

Atassi,M.Z. (1984) *Eur. J. Biochem.*, **145**, 1–20.

Benjamin,D.C., Berzofsky,J.A., East,I.J., Gurd,F.R.N., Hannum,C., Leach,S.J., Margoliash,E., Michael,J.G., Smith-Gill,S.J., Todd,P.E. and Wilson,A.C. (1984) *Annu. Rev. Immunol.*, **2**, 67–101.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F. Jr., Brice,M.D., Rogers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Berthou,J., Lifchitz,A., Artymiuk,P. and Jolles,P. (1983) *Proc. R. Soc.*, **B217**, 471–489.

Davies,D.R. and Metzger,H. (1983) *Annu. Rev. Immunol.*, **1**, 87–177.

Hopp,T.P. and Woods,K.R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3824–3828.

Ibrahimi,I.M., Eder,J., Prager,E.M., Wilson,A.C. and Arnon,R. (1980) *Mol. Immunol.*, **17**, 37–46.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.

Leach,S.J. (1983) *Biopolymers*, **22**, 425–440.

Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.

Lerner,R.A. (1982) *Nature*, **299**, 592.

Milton,L.R.C. and Van Regenmortel,M.H.V. (1979) *Mol. Immunol.*, **16**, 179–184.

Prabhakaran,M. and Ponnuswamy,P. (1982) *Macromolecules*, **15**, 314–320.

Rossmann,M., Arnold,E., Erickson,J.W., Frankenberger,E.A., Griffith,J.P., Hecht,H-J., Johnson,J.E., Kamer,G., Luo,M., Mosser,A.G., Rueckert,R.R., Sherry,B. and Vriend,G. (1985) *Nature*, **317**, 145–153.

Sheriff,S., Hendrickson,W.A., Stenkamp,R.E., Seiker,L.C. and Jensen,L.H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1104–1107.

Tainer,J.A., Getzoff,E.D., Alexander,H., Houghton,R.A., Olsen,A.J., Lerner, R.A. and Hendrickson,W.A. (1984) *Nature*, **312**, 127–134.

Tainer,J.A., Getzoff,E.D., Paterson,Y., Olsen,A.J. and Lerner,R.A. (1985) *Annu. Rev. Immunol.*, **3**, 501–535.

Taylor,W.R., Thornton,J.M. and Turnell,W.G. (1983) *J. Mol. Graphics*, **1**, 30–38.

Thornton,J.M. and Sibanda,B.L. (1983) *J. Mol. Biol.*, **167**, 443–440.

Westof,E., Altschuh,D., Moras,D., Bloomer,A.C., Mondragon,A., Klug,A. and Van Regenmortel,M.H.V. (1984) *Nature*, **311**, 123–126.

Young,C.R., Schmitz,H.E. and Atassi,M.Z. (1983) *Mol. Immunol.*, **20**, 567–570.

Received on 28 October 1985; revised on 17 December 1985