Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains)

(flexibility/contact surface/contour maps/crystal contacts/Debye–Waller temperature factor)


*Molecular and Cellular Research Laboratory and ‡Departments of Molecular Biology and Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114; and †Department of Biological Chemistry, Hershey Medical Center, Pennsylvania State University, Hershey, PA 17033

Communicated by Frederic M. Richards, August 30, 1985

ABSTRACT We evaluated surface areas on proteins that would be accessible to contacts with large (1-nm radius) spherical probes. Such spheres are comparable in size to antibody domains that contain antigen-combining sites. We found that all the reported antigenic sites correspond to segments particularly accessible to a large sphere. The antigenic sites were also evident as the most prominently exposed regions (hills and ridges) in contour maps of the solvent-accessible (small-probe) surface. In myoglobin and cytochrome c, virtually all of the van der Waals surface is accessible to the large probe and therefore potentially antigenic; in myohemerythrin, distinct large-probe-inaccessible, and nonantigenic, surface regions are apparent. The correlation between large-sphere-accessibility and antigenicity in myoglobin, lysozyme, and cytochrome c appears to be better than that reported to exist between antigenicity and segmental flexibility; that is, surface regions that are rigid often constitute antigenic epitopes, whereas some of the flexible parts of the molecules do not appear antigenic. We propose that the primary reason why certain polypeptide-chain segments are antigenic is their exceptional surface exposure, making them readily available for contacts with antigen-combining sites. Exposure of these segments frequently results in high mobility and, in consequence, to the reported correlation between antigenicity and segmental flexibility.

One of the issues currently discussed by immunologists is the nature of protein antigenicity (1–4). Early experimental results (5, 6) were interpreted as suggesting that discrete antigenic epitopes exist, implying that certain regions of protein surface are more antigenic than others. A more recent interpretation (3) of a larger experimental data base seems to indicate, however, that a number of mutually overlapping epitopes exist on protein surfaces and that perhaps the whole surface can be antigenic. It has also been noted (1, 2, 4) that antigenic epitopes composed of contiguous polypeptide-chain segments usually have higher-than-average backbone flexibility, as evidenced by Debye–Waller temperature factors (7). The temperature factor, or the B value, is given by

\[ B = \frac{8}{3} \pi^2 (\sigma^2) \]

where \( \sigma = \sqrt{r} \) is the root-mean-square atomic displacement from the crystal equilibrium position. Thus, it has been hypothesized that flexibility is an essential component of antigenicity (1, 2), perhaps because structural adjustment is a prerequisite for antibody complementarity. The hypothesis has recently been reconsidered by one of these authors (8), however.

These alternative, and often conflicting, interpretations of antigenicity bespeak conceptual difficulties confronted by workers in the field. In principle, the antigenic data is never complete, being dependent on immunogenic potential of selected amino-acid replacements in a set of homologous proteins (3) or on availability of apparently randomly obtained monoclonal antibodies. Moreover, identification of epitopes can only be based on indirect experimental procedures. In this situation, theoretical models of antigenicity are of particular value, but even here difficulties arise. For example, the proposed correlation between antigenicity and segmental flexibility is based on data that are influenced significantly by conditions such as crystal packing (9) and static crystal disorder (10) and whose exact molecular meaning is currently being debated (11–13). Hopp and Woods (14) suggested that antigenic epitopes can be located as those segments of primary structure that are markedly hydrophilic. However, a recent test (15) of their predictive algorithm reports a success rate of only 56%.

As protein antigenicity is clearly a surface property, we decided to examine the static accessibility (16–19) of selected molecules to probes of various radii. In what follows, we discuss the possible relationship of the large-probe accessibility and the location of antigenic epitopes.

Crystallographic Data and Calculations

Myoglobin, lysozyme, and cytochrome c coordinates and B-factor values were obtained from the Brookhaven Data Bank (20), those for myohemerythrin were a gift from S. Sheriff and W. Hendrickson (Columbia University, New York). The van der Waals surface of the proteins was constructed and its areas of contact to spheres of varying radii were computed by use of the Lee and Richards (16) algorithm, as implemented in the program CHARMM version 16 (21). Alternatively, the contact surfaces were computed using a procedure (to be published elsewhere) that involves calculations of the contact area on the van der Waals surface of an atom to a series of probes ordered smallest to largest. A water-sized sphere with radius \( r = 0.14 \text{ nm} \) and four other spheres with radii 0.25 nm, 0.5 nm, 0.75 nm, and 1.0 nm were used in the computations. The last sphere, with diameter 2 nm, is comparable in size to an antibody-domain dimer (Fv fragment) (22).

The contact areas (17) used in this study should not be confused with solvent accessibilities (16), defined as envelopes of protein structures obtained by the path of the center of the probe as it rolls over the protein surface. The contact surface (17) is a set of disconnected patches representing those portions of the atomic surfaces that are in contact with the surface of the probe. As the probe size increases to infinity, the contact surface of a protein converges to a small value. On the other hand, protein accessibility (16) values converge to infinity with increasing probe radii.

Residue sums were obtained from the contact-surface values of individual atoms. The sums were smoothed by the
seven-point moving-window procedure (23) and plotted against residue numbers to obtain the smoothed contact profiles.

Close intermolecular contacts in crystals were examined by use of the program CHARMM (21). Coordinates of symmetry-related molecules were generated from the starting sets of atomic coordinates by applying the rotational and translational operations of the particular space groups P2₁ for hen-egg lysozyme (24) and sperm whale myoglobin (25), P2₁2₁2₁ for cytochrome c (26), and P2₁2₁2₁ for human lysozyme (27). All possible pairs of symmetry-related molecules were then generated in the explicit hydrogen-atom representation (21) and a list of atom pairs falling within the cutoff distance of 4 Å was prepared.

Contour maps of myoglobin and lysozyme surfaces were computed by a method described in greater detail elsewhere (41). Briefly, the molecular surfaces (17) have been determined using a 0.14-nm probe (19) (small sphere), and then contoured³ in increments of 1 Å (i.e., 0.1 nm). The list of points, whose coordinates define the molecular surface, is obtained using an algorithm by Connolly (19). The starting elevation is taken at the surface of a triaxial ellipsoid with axes scaled to be 0.5 times the principal axes of the protein (28); points below the starting elevation are eliminated. The maps are displayed as Mollweide projections, a representation that preserves the relative areas of all features.

Results and Discussion

Correlation of Contact-Area Peaks with Locations of Antigenic Epitopes. Fig. 1 A–C shows contact-surface profiles computed with the largest sphere size (r = 1 nm) for myoglobin, lysozyme, and cytochrome c and smoothed over 7-residue segments. The smoothing procedure was chosen in order to account for the fact that the antibody binding site can accommodate 6–8 amino acid residues (36). Also shown are the 111 amino acid positions that were implicated in antigenic epitopes in all three proteins (3, 4). An overwhelming majority of them coincide with prominent peaks shown on contact-surface profiles. Only 8 of the antigenic residues (140 and 113–116 in myoglobin; 33, 34, and 93 in lysozyme) are located on minor local maxima, and 5 of the antigenic positions (residues 12 and 144 in myoglobin; 94 in cytochrome c; and 5 and 25 in lysozyme) occur in regions not covered by any of the contact-profile peaks. However, these 5 positions are parts of larger epitopes and contain other residues that are prominently exposed. Residue 12 in myoglobin belongs to the epitope that also includes residues 4 and 79, both located in contact-area peaks, whereas residue 144 forms an epitope with residue 83, which is located at a maximum. Similarly, residues 89 (highly exposed) and 92 in cytochrome c are parts of the same epitope; residue 5 in lysozyme belongs to the same epitope as the peak-associated residues 7, 13, 14, and 125; and residue 25 of lysozyme is included in the epitope defined by the neighboring residues 20–23, all prominently exposed. The antigenic residues located on minor local maxima are likewise parts of larger epitopes that also include major contact-profile peaks, with the sole exception of residue 140 in myoglobin. Consequently, the reported antigenic regions can be said to correspond, with a single exception, to surface regions particularly well accessible to the large probe.

The reported antigenic epitopes cover virtually all the prominent maxima of the contact-area profiles in lysozyme and most of those in myoglobin. Only a small part of the accessible surface of cytochrome c is known to be antigenic (Fig. 1C), but evolutionary constancy of the primary structure of this molecule makes the delineation of antigenic sites particularly difficult. The two contact profiles displayed in Fig. 1C, those of tuna and bonito cytochrome c, respectively, suggest that contact-surface profiles of homologous proteins are likely to be very similar.

Antigenic Epitopes and Average Backbone B-Factor Values. The average-backside maxima of B factors of backbone atoms N, Cα, C, and O are plotted in Fig. 1 A–C by light lines. Comparison of the two types of data—namely, the contact profiles and the backbone B factors—indicates that segments of high flexibility constitute a proper subset of regions exceptionally accessible to the large-size probe. That is, of the 46 major peaks of the B factors, all but 1 (lysozyme residue 107) coincide with some of the 54 distinct or partially overlapping maxima of large-probe contact areas. It seems significant that there are several antigenic epitopes associated with highly accessible regions of no exceptional flexibility (residues 56–62 and 140 in myoglobin; 1,41,84 or 19,21 in lysozyme; and 1–4 and 60,62 in cytochrome c), whereas there are no epitopes that are flexible but inaccessible to the large probe.

Next, the possibility was considered that the above-mentioned "rigid" epitopes are artificially constrained by crystal contacts and might be more flexible in solution. An exhaustive search for intermolecular crystal contacts showed no close appositions in the human (27) and hen-egg-white (24) lysozyme crystals that would involve the epitope 1–41–84, although the side-chain terminal nitrogen of Arg-41 (human structure) is only 3.2 Å from the Cα atom of Pro-103. This site, then, can be said to be inherently "inflexible." The myoglobin epitopes 56–62 and 140, as well as the cytochrome c epitopes 1–4 and 60,62 and the lysozyme site 19,21 were involved in intermolecular interactions. The significance of this finding remains unclear, however, particularly because our search also revealed multiple close intermolecular contacts involving residues from segments of above-average backbone B-factor values, as, for example, between Gly-153 (backbone B value = 43.3) and Gln-91 (B = 10.0) in myoglobin, or between the side chain of Lys-27 (B = 21.5) and the backbone oxygen of Gly-77 (B = 15.0) in cytochrome c. Clearly, crystal contacts are not incompatible with high average backbone B factors (10) and, conversely, an existence of close crystal contacts should not be taken to imply that the segment in question becomes more flexible in solution. Instead, independent evidence is needed to prove this point. The α-carbon B values computed from normal-mode dynamics simulations on hen-egg-white lysozyme by Levitt et al. (37) do not indicate an above-average backbone flexibility for the two epitopes 1–41–84 and 19–21, nor do they reproduce the B-value maximum at position 107 (not coinciding with any of the contact-profile peaks). Likewise, dynamical simulations of Northrup et al. (38) do not suggest an above-average flexibility of cytochrome c residues 60–62, although NMR measurements (39) show the region around 60,62 to be flexible.

Is the Whole Protein Surface Antigenic? Comparison of contact-area profiles computed with small- and large-probe radii (e.g., 0.25 nm and 1 nm) revealed an interesting fact: the magnitudes of the contact areas obtained with the two probe radii differed, yet the positions of the maxima of the small-probe and large-probe profiles were virtually identical in all the three cases (cf. Fig. 1D for myohemerythrin). Thus, residues that are accessible to contacts with large, protein-sized spheres correspond to surface regions with an exceptional exposure to small, water-sized spheres. Fig. 1D shows that amino acid sequences of the two "cold" myohemerythrin peptides, whose antibodies fail to crossreact with native myohemerythrin (2), are located in regions that are only marginally accessible to a small-size probe and are inacces-

Biochemistry: Novotný et al.


Fig. 1. Profile of contact areas computed with a spherical probe of radius 1.0 nm. Contact areas (1 Å² equals 0.01 nm²) represent those parts of the van der Waals surface of a protein that come in direct contact with the sphere, when the sphere is rolled over the surface (16). The larger the spherical probe used, the smaller the contact area. Residue-contact areas shown were obtained as sums of computed atomic-contact areas. The values were smoothed by a 7-point moving-window algorithm (23) and plotted against the sequence numbers. The average-contact backbone B factors, as computed from the atomic B factors contained in the Brookhaven Protein Data Bank (20), are shown in light lines. Residues identified as parts of antigenic epitopes are indicated by symbols, with symbols belonging to the same epitope drawn on the same ordinate. (A) Sperm whale myoglobin. Six contiguous epitopes were identified experimentally (residues 15–22, 56–62, 94–99, 113–119, 146–151 (5) and 140 (29)), in addition to two discontinuous epitopes [4, 12, and 79; 83, 144, and 145 (29)]. The average backbone B factor of deoxyxgenated myoglobin (25) is 10.2 (B range from 5.0 to 46.0). (B) Hen-egg-white lysozyme. One contiguous epitope [64–80 (30)] and seven discontinuous ones [5, 7, 13, 14, and 125 (6); 62, 87, 89, 93, 96, and 97 (6); 33, 34, 113, 114, and 116 (6); 45–48 and 68 (31); 19 and 21 (32); 102 and 103 (32); 1, 41, and 84 (32)] were identified experimentally. The average backbone B factor of human lysozyme (27) is plotted. The average backbone B-factor value is 17.2 (B range from 9.5 to 55.3). (C) Bonito and tuna cytochromes c. The experimentally defined antigenic epitopes are 1–4 (33), 42–46 (34, 35), 50 and 62 (34), 89 and 92 (34), and the carboxyl terminus (5). The average backbone B-factor values of the tuna structure (26) is 15.3 (B range from 10.0 to 23.4). (D) Profile of contact areas to myohemerythrin computed with a spherical probe of radius 0.25 nm. As discussed in the text, the profile of contact areas to a large-sized probe (such as 1.0 nm in radius) can be obtained from the one displayed in the figure by considering only those parts of the profile that show an exceptional exposure (heavy lines). Bars drawn in dashed lines denote positions of peptides known as "cold" (2); antibodies elicited against these peptides fail to react with the intact myohemerythrin molecule. It can be seen that polypeptide-chain segments corresponding to "cold" regions are poorly accessible to large-size probes.
**Concluding Remarks**

We have found that the availability of particular regions of protein surface to contacts with large, spherical probes correlates well with the antigenicity of myoglobin, lysozyme, cytochrome c, and myohemerythrin. Although most of the highly accessible regions of the backbone are also more flexible than the average, there are distinct antigenic epitopes that seem to be rigid, particularly the residues 1, 41, and 84 in lysozyme.

A comparison of individual B-factor values found in myoglobin and lysozyme crystals strengthens this interpretation. Most of the above-average B-value maxima associated with antigenic regions in myoglobin (average backbone B value of 10.2) are significantly smaller than those found in lysozyme and rarely exceed the value $B = 17$ ($\sigma = 0.6$ Å). Conversely, the average backbone B-factor values in hen-egg-white and human lysozymes are 13.9 and 17.2, respectively, whereas the resolution of the human structure (1.5 Å) is comparable to that of myoglobin (1.4 Å). The B-factor differences may be attributable to the fact that the myoglobin molecule is more compact, having more secondary structure (76%) than lysozyme (56%) and acquiring an additional stabilization from its heme ligand; however, myoglobin, although apparently more rigid than lysozyme, is an equally good antigen. The only highly flexible segments of myoglobin, yet to be implicated by experimental studies as major independent epitopes, encompass residues 1-4 and 151-153. The same is true of the flexible region in lysozyme, residues 105-110.

This static-accessibility model of antigenicity incorporates both the contiguous and discontinuous epitopes. Profiles in Fig. 1 A-C suggest that virtually all the molecular surface is potentially antigenic, in accordance with the conjecture of Benjamin et al. (3), whereas Fig. 1D seems to indicate that protein molecules exist for which a significant part of the surface is not accessible to contacts with antibody and therefore is not antigenic (2). Because large-probe accessibility and above-average B factors in any given structure are strongly correlated, it is difficult to assign antigenicity with complete confidence to one factor or the other. It would appear that somewhat greater consistency can be found for the accessibility correlation, but more experimental work is needed to clarify this issue.

We thank Professors Martin Karplus (Harvard University, Cambridge, MA) and Frederic M. Richards (Yale University, New Haven, CT) for comments. We also thank Dr. Wayne Hendrickson (Columbia University, New York, NY) for generously sharing his crystallographic results.
University, New York) for providing us with myoglobintryrin coordinates and for communicating his results to us prior to publication. The idea to explore large-probe contact surfaces was also conceived by Dr. W. Wiley (Harvard University), who suggested to us, to M. H. We note that Dr. Mitchell Lewis (Smith, Kline & Beckman, Philadelphia) has arrived at a conclusion similar to that presented here, by different methods; similarly, Drs. G. Petsko, Robert Campbell, James Mottonen, and Robert Tilton (Massachusetts Institute of Technology, Cambridge) have found that a large fractional hydrophobic accessibility is observed for antigenic regions, thus supporting the view that antigenic sites are more strongly correlated with static accessibility than with residue mobility. G.D.R. wishes to thank Dr. Bill DeGrado for useful discussion and gratefully acknowledges support from the National Institutes of Health (Grant GM29458 and a Research Career Development Award). J.A.S. was supported by a grant from Hoechst A.G. (West Germany) and M.H., by a grant from Upjohn (Kalamazoo, MI).