Antigenicity of Myohemerythrin

Geysen *et al.* (*1*) report information bearing on the antigenicity of the protein myohemerythrin. Data assembled from 113 hexapeptides competing with the parent protein for binding to seven different rabbit polyclonal antisera to myohemerythrin allowed assignment of relative antigenicities to segments of primary structure. Their result is compared with sequence profiles of properties such as backbone mobility, atomic packing density, solvent accessibility, shape-accessibility radius, and electric charge distribution of myohemerythrin. The authors conclude that "the most frequently recognized sites form three-dimensional superassemblies characterized by high local mobility, convex surface shape, and often by negative electrostatic potential." They also take issue with our previous work (*2*) and with the work of others (*3*) which implies that antibody accessibility is the primary intrinsic determinant of antigenicity. We wish to demonstrate that the large probe accessibility model accounts for the antigenicity data of Geysen *et al.* at least as well as, and perhaps even better than, any of the structural properties investigated by these authors. In our model, the 1-nm radius probe is rolled over the protein surface in order to identify, relatively sparsely, its most protruding atoms. Antibody binding sites, with more complementary shapes, are known to contact up to six amino acid residues surrounding these most protruding points; data smoothing has been introduced to take account of the difference between the elementary nature of our model and the convolutions of the contact surface.

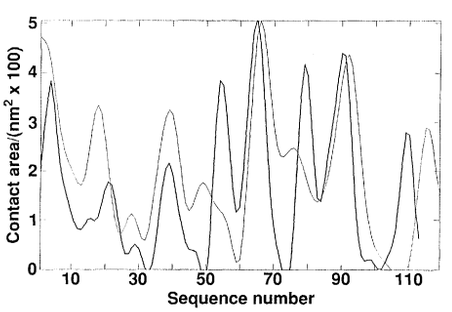

Fig. 1. Antigenicity (heavy line) and areas of contact with a spherical probe (probe radius = 1 nm, light line) of myohemerythrin. The antigenicity data are those given in figure 1C of (*1*), smoothed by three passes of the moving window algorithm (*2*). Nine passes of the same smoothing protocol were applied to the computed large probe contact areas [see (*2*) for details of contact area computations].

In Fig. 1, we have reproduced myohemerythrin antigenicity [data taken from figure 1C of (*1*)] as a smooth curve and have compared this curve with a profile of myohemerythrin large probe contact area computed and smoothed as described (*2*). The similarity of the two curves is striking. They both consist of nine maxima, six of which coincide in the two curves (that is, those at residues 4, 20, 29, 39, 64, and 90) and three of which partially overlap (antigenicity peak at residue 54 and contact area peak at 49; antigenicity peak at residue 79 and contact area peak at 77; antigenicity peak at residue 109 and contact area peak at 114). This degree of accord compares favorably with any of the correlations between peptide antigenicity and structural properties discussed by Geysen *et al.*, as depicted in figures 1, D and E, of their article. For example, Geysen *et al.* classify primary structural segments into three broad categories—most, average, and least antigenic (their table 1). The average (and total) large probe contact areas for the most and the least antigenic sites, evaluated according to this classification, are 2.27 Å^2 (109 Å^2 total) and 1.29 Å^2 (40 Å^2 total), respectively. These values, once again, differentiate better between the antigenic and nonantigenic segments than, for example, data for mobility of such segments favored by Geysen *et al.*, namely, 30 Å^2 and 23.2 Å^2.

The relative shift of antigenicity and large probe accessibility maxima at positions 109 and 114 (and, to a lesser extent, at some other positions) are of potential interest with regard to the large probe accessibility model. The phenomenon either indicates a genuine "frameshift" displacement of the two properties along the amino acid sequence, or it may be a by-product of partial cross-inhibition between peptides with similar sequences (such as the sequence Glu-Glu-His at positions 23 to 25 and the sequence His-Glu-Glu at positions 58 to 60); of competition at sites near to, but not coincident with, the antigenic epitope; or of any other similar "noise" necessarily inherent in the experimental method employed. Each of the rabbit polyclonal sera investigated by Geysen *et al.* probably consists of many partially overlapping antibody specificities; and boundaries of antigenic epitopes, as defined by such mixtures, will be necessarily more diffuse than those defined by individual monoclonal antisera. Examination of molar ratios at which the peptides competed with the parent protein might provide valuable hints; but such data are not easily obtained from assays described in (*1*), where peptides used in immunoassays were not purified subsequent to their chemical synthesis. We note that in experimental systems where the most important antigenic residues were identified in a sharp and unequivocal way [for example, single residue mutants of virus coat proteins (*4*)], the correlation between the most antigenic and the largest probe-accessible sites has been virtually absolute.

Structural attributes listed by Geysen *et al.* as characteristic of antigenic sites are properties known (*5*) to be associated with the most protruding, and therefore the most antibody-accessible, parts of protein surfaces. The relative importance of these various attributes for antigenicity is still being debated, but a review of the available experimental data lead us to conclude (*6*) that antigenicity and antibody accessibility are virtually synonymous, while the other properties listed by Geysen *et al.* are dispensable in at least some antigenic sites (*7*).

JIRI NOVOTNY
ROBERT E. BRUCCOLERI
WILLIAM D. CARLSON
MARK HANDSCHUMACHER
EDGAR HABER
Molecular and Cellular Research Laboratory, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114

REFERENCES

1. H. M. Geysen *et al.*, *Science* **235**, 1184 (1987).
2. J. Novotny *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 226 (1986).
3. D. J. Fanning, J. A. Smith, G. D. Rose, *Biopolymers* **25**, 863 (1986); J. M. Thornton, M. S. Edwards, W. R. Taylor, D. J. Barlow *EMBO J.* **5**, 409 (1986).
4. J. M. Hogle, M. Chow, D. J. Filman, *Sci. Am.* **256**, 42 (March 1987).
5. G. D. Rose, I. M. Gierash, J. A. Smith, *Adv. Protein Chem.* **37**, 1 (1985); G. A. Petsko and D. Ringe, *Ann. Rev. Biophys. Bioeng.* **13**, 331 (1984).
6. J. Novotny, M. Handschumacher, R. E. Bruccoleri, *Immunol. Today* **8**, 26 (1987).
7. J. Novotny and E. Haber *Biochemistry* **25**, 6748 (1986).

10 April 1987; accepted 23 July 1987

Response: Using our experimental data, Novotny *et al.* purport to find a better correlation of antigenic sites with their calculated parameter "antibody accessibility" than with the factors we identified (*1*). We address three basic problems in their technical comment and other published papers (*2*, *3*) that deserve clarification both for this work on antigenicity and for the general field of structural analysis.

1) Methodology to correlate structure and activity should be described with informative terminology and, more important, should be compatible with the relevant experimental data. The "antibody accessibility" nomenclature implies that the area actually measured is that exposed to an antibody binding site, rather than that exposed to a large sphere (radius usually 1 nm). More-

over, the application of this large probe accessibility method is not compatible with the experimental data. Our antigenic data were measured with the use of hexapeptides and thus provide discrete values corresponding to each set of six contiguous residues. To properly match this data, we assessed structural properties for six-residue segments. Novotny *et al.* inappropriately apply their algorithm to our database. In their response, data corresponding to single values are smoothed repeatedly with a moving window of seven residues. This method is mismatched with our antigenic data for several reasons: single valued data for large probe accessibility are compared with hexapeptide antigenicity data; the smoothing algorithm chosen accentuates patterns seven residues in length, while the antigenicity was measured with six-residue peptides; large probe accessibility data are smoothed with nine passes of this moving window algorithm and the antigenic data with three passes. Furthermore, Novotny *et al.* do not give reasons for their choice of coefficients for the smoothing algorithm (either here or in the referenced papers) or for the number of smoothing passes applied. Finally, as seen in figure 1 of the comment by Novotny *et al.*, the calculated curves of antigenicity and large probe accessibility are of different lengths and are not aligned with each other. These discrepancies occur because the antigenic data for hexapeptides were originally plotted at the first residue of each hexapeptide [figure 1 in (*1*)], but the large probe accessibility values were plotted in the center of the seven-residue window of the smoothing algorithm; this results in a 2.5-residue offset between the two plots. A revised version of this figure with only the offset corrected (our Fig. 1) indicates good correlation of peaks at antigenic positions 65 to 70, 91 to 96, and 110 to 115. However, antigenic peaks 21 to 26, 54 to 59, and 80 to 85 match large probe accessibility minima, and the matches at positions 4 to 9 and 40 to 45 correspond with inflection points.

2) A predictive algorithm should give consistent, objective, and accurate results that are not tailored to fit the experimental data. Novotny *et al.* use the same method to predict that the area around residue 54 in myohemerythrin is both antigenic and not antigenic. In (*2*, p. 228) they use their large probe analysis to predict that "antibodies elicited by a peptide corresponding to myohemerythrin residues 53–60 will react poorly, or not at all, with the myohemerythrin molecule," while in their comment they conclude that the contact area peak at residue 49 partially overlaps with the antigenicity peak at residue 54. A visual assessment of peaks (which are perceived as occupying the region between minima, rather than between points of inflection) overestimates peak overlap because the proportion of the width of the figure occupied by peaks is subjectively expanded at the expense of the troughs. Overestimation of overlap is aggravated by arbitrary smoothing of the experimental and calculated curves. Finally, smoothing cannot correct for the difference between large probe accessibility and the actual surface that participates in the van der Waals and hydrogen bonding interactions of a macromolecular contact. Large probe accessibility is an inaccurate representation of surface shape because a small change in atomic positions (or probe radius) can make regions accessible or inaccessible. Thus, as part of our study (*1*), Getzoff and Tainer developed and used a precise, analytical algorithm (that does not depend upon a predetermined probe size) to calculate the largest sphere that can touch any atom in a protein and to evaluate shape-accessibility of the molecular surface and its relationship to antigenicity.

3) Structural correlations must stand or fall on the basis of defined criteria, preferably including statistical analysis. Qualitative claims such as "These [large probe accessibility] values, once again, differentiate better between the antigenic and nonantigenic segments than, for example, data for mobility . . ." occur frequently in their comment and elsewhere (*2, 3*) without substantiation. However, statistical analyses show that large probe accessibility is unable to distinguish among the three different categories of immunological reactivity in myohemerythrin. Table 1 gives statistical data for mobility and large probe accessibility as assessed for the whole protein and for residues within each of the three different categories of immunological reactivity. Student's *t* test results and associated probability levels (Table 2) show that mobility significantly distinguishes the least reactive category from both the most and the average ($P < 0.015$, with Bonferroni's correction), whereas large probe accessibility does not distinguish among any of these categories ($P > 0.05$). Smoothing (forcing a seven-residue periodicity on the data) will not remedy this lack of statistical significance. However, our analytical shape algorithm does show a statistically significant correlation of shape accessibility with

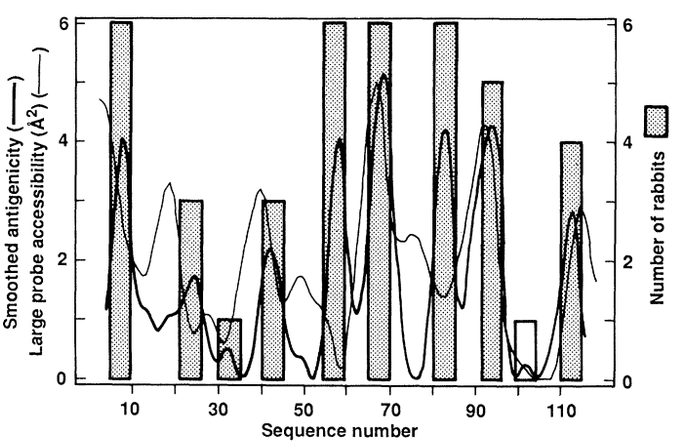

Fig. 1. Antigenicity (heavy line) and large probe (radius = 10 Å) accessibility (light line) of myohemerythrin plotted versus residue number. Curves are taken from figure 1 of the comment of Novotny *et al.*; the antigenicity data have been smoothed by three passes of the moving window algorithm and the large probe accessibility data by nine passes. The antigenicity curve has been offset by +2.5 residues so as to match the alignment of the large probe accessibility curve. For the nine antigenicity maxima used by Novotny *et al.* (shaded bars) and the tenth minor antigenicity peak (unshaded bar), the bar graph shows the frequency of the antigenic response, as defined by the number of rabbit antisera (from a total of seven) that react with each of these hexapeptides.

Table 1. Class of immunoreactivity and number of residues per class defined in Geysen *et al.* (*1*). Mobility is given as the average main-chain temperature factors after correction for crystal contacts (*4*). (The value of 29.0 instead of 30.0 for the most reactive category is based upon a more precise summation of the B-values.) Unsmoothed large probe accessibility data are calculated by the method of Novotny *et al.* (*2*).

Reactivity	No. of residues	Mobility (Å²)*	Large probe accessibility (Å²)*
Most	48	29.0 ± 4.0	2.27 ± 3.52
Average	39	27.6 ± 4.4	2.34 ± 2.77
Least	31	23.2 ± 4.5	1.30 ± 2.59
MHr	118	27.0 ± 4.9	2.04 ± 3.06

*Mean ± standard deviation.

Table 2. Comparison of reactivity categories.

Reactivity pairing	Mobility (Å²)		Large probe accessibility (Å²)	
	t	*P**	*t*	*P**
Most/average	1.52	>0.05	0.09	>0.05
Average/least	4.12	<0.015	1.61	>0.05
Most/least	5.84	<0.015	1.42	>0.05

*With Bonferroni correction.

antigenicity (most/least; $t = 3.87$), but at a lower level than both mobility ($t = 5.84$) and low packing density ($t = 4.66$) (*1*).

In sum, while subjectively appealing as a reflection of the tautology that antibodies must have access to an antigenic site in order to bind, the approach and algorithm applied by Novotny *et al.* here and elsewhere (*2, 3*) do not satisfy the requirements of (i) compatible methodology; (ii) consistent, objective, and accurate results, and (iii) defined criteria for correlation. In contrast, our statistically significant correlation of mobility with the most frequently recognized sites suggests structural shifts in the antigen upon antibody binding. This prediction has since been confirmed in the crystallographic structure of the neuraminidase-antibody complex (*5*). Decoding the structural basis for macromolecular recognition will benefit from multiple approaches, but will require both imagination and rigor.

Note added in proof: Monoclonal antibodies raised against both the mobile C-helix region of myohemerythrin and the whole protein recognize the region 79 to 84 (*6*), which is a large probe accessibility minimum (Fig. 1).

H. MARIO GEYSEN
STUART J. RODDA
TOM J. MASON
Department of Molecular Immunology, Commonwealth Serum Laboratories, Parkville, Victoria 3052, Australia
JOHN A. TAINER
HANNAH ALEXANDER
ELIZABETH D. GETZOFF
RICHARD A. LERNER
Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, CA 92037

REFERENCES

1. H. M. Geysen *et al.*, *Science* **235**, 1184 (1987); E. D. Getzoff *et al.*, *ibid.* p. 1191.
2. J. Novotny *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 226 (1986).
3. J. Novotny and E. Haber, *Biochemistry* **25**, 6748 (1986); ———, M. Handschumacher, R. E. Bruccoleri, *Immunol. Today* **8**, 26 (1987).
4. S. Sheriff, W. A. Hendrickson, R. E. Stenkamp, L. C. Sieker, L. H. Jensen, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1104 (1985).
5. P. M. Colman *et al.*, *Nature (London)* **236**, 358 (1987).
6. T. M. Fieser, J. A. Tainer, H. M. Geysen, R. A. Houghten, R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8568 (1987).

8 June 1987; accepted 23 July 1987

Do 15 Million Cat Neurons Mediate the Memory of a Circle and a Star?

E. R. John *et al.* (*1*) write that about 15 million (*2*) neurons increase activity when the memory of a circle or a star is activated in the cat brain. We believe that ambiguities inherent in John *et al.*'s experimental procedure do not justify this conclusion.

1) Three split-brain cats were trained to go down a runway and push through one of two doors labeled with a white geometric figure (two concentric circles) for a food reward. The door bearing the negative cue (a star) was locked, and self-correction was not permitted. Training sessions consisted of 40 trials, at 1-minute intervals, held at the same time each day. After the end of each training session, the animals had free access to food in their home cage until evening. Upon reaching criterion (90% correct), after a training period of unknown length, each cat was trained for another 6 weeks—or a minimum of 1680 additional trials. We are not told anything about response latencies or about the nature of "incorrect" responses (were they incorrect choices or a refusal to choose?).

2) The white geometric figures were then replaced by identical green symbols. The cat saw the world through only one eye, which was covered with a green transparent contact lenses. After an unspecified "brief initial period of hesitation" each cat once more performed for food at criterion levels. By using various combinations of opaque and colored lens, the authors showed that each hemisphere in these split-brain cats could perform the visual discrimination and presumably guide the motor response.

3) After the first label, [^{14}C]2DG, was injected into a paw, the green contact lens was placed over one eye and a red transparent contact lens over the other eye. Each door was then labeled with a transparent red triangle in addition to the transparent green circles or star. At this point input about the learned cues was delivered only to the hemisphere that saw through the green contact lens. The other, reference, hemisphere saw only the red triangle on each door and thus was denied a discriminant cue.

4) After the injection of the second (control) label, [^{18}F]2DG, into the other paw, the same lenses remained in place, but the symbols on the doors were changed. The card on each door bore both a green and a red triangle, that is, no discriminant cues were offered to either eye.

The effect of the learned cues on neural activity was obtained by subtracting the brain activity of labeled 2DG during the control tests ([^{18}F]2DG) from that in the tests ([^{14}C]2DG) during which one hemisphere received familiar discriminant cues and the other hemisphere received unfamiliar nondiscriminant cues. The difference images of the two hemispheres were then compared. The authors conclude that areas which showed hemisphere asymmetry in their metabolic maps were involved in the storage or processing of the memories of the discriminant cues.

John *et al.* appear to be saying that the paradigm described above has eliminated all information not specific to the learned task, and, therefore, that their difference images characterize the brain components active during retrieval of the "circles and star" memory. However, they (*1*) present no evidence that all or any of the 15 million neurons that show excess activity were involved in the storage of the discriminant cue memory. Part of the increase in neural activity could be due to general processes of memory retrieval. For example, circuits in visual cortex could recognize a particular visual input as familiar, and this could then trigger a memory search to uncover any other memories associated with the familiar input. Much of the observed increase in neural activity could be related to this search function, which tells us little about the anatomical representation of an engram.

Equally basic, during the first [^{14}C]2DG test the hemisphere that looks through the green contact lens sees a double circle on one door and a star on the other one. Both clues are familiar, and as discriminanda, have a symbolic meaning about the presence or absence of food. The hemisphere that looks through the red lens sees unfamiliar cues—the triangles (at least they are unfamiliar the first time they are seen). In the second [^{18}F]2DG test, the situation is more complex; the triangles are then novel—on first exposure—for the green lens hemisphere, and familiar for the red lens hemisphere, which has seen them during the first experimental condition. During the first experimental condition the cat obtained food by pushing doors labeled with red triangles, so that triangles may not be a neutral cue to the red lens hemisphere. Thus, it cannot be argued that during the second test the two hemispheres are operating under "informationally symmetric conditions." Differences in the activity of the green lens hemisphere during the [^{14}C]2DG and [^{18}F]2DG tests could be related to the presence or absence of discriminant cues on which to base a visual choice or to the presence or absence of familiar cues, or both.