recently characterized human mtDNA primase also requires structural RNA for its activity (*45*), and it is possible that RNA import is involved. Furthermore, all known RNase P activities require an RNA component (*46–51*). In yeast, the tRNA synthesis locus of mtDNA encodes the RNA moiety of mitochondrial RNase P (*52*). However, no candidate gene for the RNA moiety of RNase P exists in mammalian mtDNA (*27–29*), suggesting that the RNA component is nucleus-encoded. Therefore, translocation of RNA across the mitochondrial membrane could be a general aspect of mitochondrial biogenesis. Identifying nucleus-encoded RNA's essential for mitochondrial biogenesis and studying their regulation and transport into mitochondria may provide substantive insight into our understanding of nuclear-mitochondrial interactions.

REFERENCES AND NOTES

1. G. Schatz, E. Halsbrunner, H. Tuppy, *Biochem. Biophys. Res. Commun.* **15**, 127 (1964).
2. M. M. K. Nass, S. Nass, B. A. Afzelius, *Exp. Cell Res.* **37**, 516 (1965).
3. P. Borst and L. A. Grivell, *Cell* **15**, 705 (1978).
4. G. Schatz and T. L. Mason, *Annu. Rev. Biochem.* **43**, 51 (1974).
5. R. Hay, P. Böhni, S. Gasser, *Biochim. Biophys. Acta* **779**, 65 (1984).
6. A. Tzagoloff and A. M. Myers, *Annu. Rev. Biochem.* **55**, 249 (1986).
7. J. W. Chase and I. B. Dawid, *Dev. Biol.* **27**, 504 (1972).
8. J. Heasman, J. Quarmby, C. C. Wylie, *ibid.* **105**, 458 (1984).
9. L. S. Dillon, *Ultrastructure, Macromolecules, and Evolution* (Plenum, New York, 1981), pp. 377–386.
10. K. J. Newton and V. Walbot, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6879 (1985).
11. D. A. Clayton, *Cell* **28**, 693 (1982).
12. ———, *Annu. Rev. Biochem.* **53**, 573 (1984).
13. D. D. Chang and D. A. Clayton, *EMBO J.* **6**, 409 (1987).
14. T. R. Cech and B. L. Bass, *Annu. Rev. Biochem.* **55**, 559 (1986).
15. T. E. England and O. C. Uhlenbeck, *Nature (London)* **275**, 560 (1978).
16. Although there is a periodicity in the sizes of these fragments, the two larger species are not artifactual multimers of the smaller 65-nt RNA generated during the in vitro labeling with T4 RNA ligase since the three RNA species can be separated by subsequent fractionation (see Fig. 3).
17. R. E. Lockard *et al.*, *Nucleic Acids Res.* **5**, 37 (1978).
18. H. Donis-Keller, *ibid.* **8**, 3133 (1980).
19. M. S. Boguski, P. A. Hieter, C. C. Levy, *J. Biol. Chem.* **255**, 2160 (1980).
20. The computer analyses were carried out with the MicroGenie Program (SciSoft, Inc.).
21. A. Fersht, *Enzyme Structure and Mechanism* (Freeman, San Francisco, 1977), pp. 93–96.
22. A. Kramer, W. Keller, B. Appel, R. Lührman, *Cell* **38**, 299 (1984).
23. A. R. Krainer and T. Maniatis, *ibid.* **42**, 725 (1985).
24. D. L. Black, B. Chabot, J. A. Steitz, *ibid.*, p. 737.
25. The melting temperature was determined from the following equation: $T_m(°C) = 16.6 \log[Na^+] + 0.41[\%(G + C)] + 81.5 - (820/L)$, where L is the length of the oligonucleotide.
26. J. C. Alwine *et al.*, *Methods Enzymol.* **68**, 220 (1979).
27. S. Anderson *et al.*, *Nature (London)* **290**, 457 (1981).
28. M. J. Bibb, R. A. Van Etten, C. T. Wright, M. W. Walberg, D. A. Clayton, *Cell* **26**, 167 (1981).
29. S. Anderson *et al.*, *J. Mol. Biol.* **156**, 683 (1982).
30. R. A. Collins, L. L. Stohl, M. D. Cole, A. M. Lambowitz, *Cell* **24**, 443 (1981).
31. D. R. Pring, C. S. Levings, III, W. W. L. Hu, D. H. Timothy, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2904 (1977).
32. P. Borst, *Annu. Rev. Biochem.* **41**, 333 (1972).
33. G. Schatz, *FEBS Lett.* **103**, 203 (1979).
34. G. Schatz and R. A. Butow, *Cell* **32**, 316 (1983).
35. F. Wunderlich, in *The Cell Nucleus: Nuclear Particles*, H. Busch, Ed. (Academic Press, New York, 1981), pp. 249–281.
36. G. Zieve and S. Penman, *Cell* **8**, 19 (1976).
37. S. J. Madore, E. D. Wieben, T. Pederson, *J. Cell Biol.* **98**, 188 (1984).
38. I. W. Mattaj and E. M. DeRobertis, *Cell* **40**, 111 (1985).
39. D. E. Fisher, G. E. Conner, W. H. Reeves, R. Wisniewolski, G. Blobel, *ibid.* **42**, 751 (1985).
40. B. Setyono and J. R. Greenberg, *ibid.* **24**, 775 (1981).
41. M. Schleyer and W. Neupert, *ibid.* **43**, 339 (1985).
42. M. Eilers and G. Schatz, *Nature (London)* **322**, 228 (1986).
43. R. P. Martin, J. M. Schneller, A. J. C. Stahl, G. Dirheimer, *Biochemistry* **18**, 4600 (1979).
44. N. Chiu, A. Chiu, Y. Suyama, *J. Mol. Biol.* **99**, 37 (1975).
45. T. W. Wong and D. A. Clayton, *Cell* **45**, 817 (1986).
46. B. C. Stark, R. Kole, E. J. Bowman, S. Altman, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3717 (1978).
47. K. Gardiner and N. R. Pace, *J. Biol. Chem.* **255**, 7507 (1980).
48. E. Akaboshi, C. Guerrier-Takada, S. Altman, *Biochem. Biophys. Res. Commun.* **96**, 831 (1980).
49. L. Kline, S. Nishikawa, D. Söll, *J. Biol. Chem.* **256**, 5058 (1981).
50. C. J. Doerson, C. Guerrier-Takada, S. Altman, G. Attardi, *ibid.* **260**, 5942 (1985).
51. M. Hollingsworth and N. C. Martin, *Mol. Cell. Biol.* **6**, 1058 (1986).
52. D. L. Miller and N. C. Martin, *Cell* **34**, 911 (1983).
53. V. A. Erdmann and J. Walters, *Nucleic Acids Res.* **14**, r1 (1986).
54. J. D. Dignam, R. M. Lebovitz, R. G. Roeder, *ibid.* **11**, 1475 (1983).
55. E. T. Schenborn and R. C. Mierendorf, Jr., *ibid.* **13**, 6223 (1985).

56. We thank A. Pavian for the operation of the oligonucleotide synthesizer. Oligonucleotide 5.8*S* and plasmid SP64/Ulc$^-$ were provided by T. W. Wong and by D. McKenzie and J. E. Dahlberg, respectively. Supported by grants from the NIH (GM-33088-16) and the American Cancer Society, Inc. (NP-9M).

3 October 1986; accepted 14 January 1987

Chemistry of Antibody Binding to a Protein

H. MARIO GEYSEN, JOHN A. TAINER, STUART J. RODDA, TOM J. MASON, HANNAH ALEXANDER, ELIZABETH D. GETZOFF, RICHARD A. LERNER

The chemistry of antibody recognition was studied by mapping the antigenicity of the protein myohemerythrin with peptide homologs of the protein sequence. The results suggest that the entire protein surface is antigenic, but the probability of there being antibodies to a given site is influenced by local stereochemistry. Although accessible to an antibody binding domain, the least reactive positions cluster in the most tightly packed and least mobile regions and are closely associated with narrow, concave grooves in the molecular surface containing bound water molecules. The most frequently recognized sites form three-dimensional superassemblies characterized by high local mobility, convex surface shape, and often by negative electrostatic potential.

ANTIGENIC SITES IN PROTEINS ARE OFTEN COMPLEX CONformations dependent on the tertiary folding of the protein chain. Unless sequenced mutational or evolutionary variants of the protein are available, the exact amino acid residues to which an antibody binds are difficult to determine. Moreover, the proteins that are best defined structurally and antigenically are often very conserved among species, thus producing limited immunological responses experimentally. In this article we describe our study of the immune response to the protein myohemerythrin (MHr), in which a large number of synthetic peptides were used to map antigenic sites on the three-dimensional crystallographic structure.

H. M. Geysen, S. J. Rodda, and T. J. Mason are in the Department of Molecular Immunology, Commonwealth Serum Laboratories, Parkville, Victoria 3052, Australia. J. A. Tainer, H. Alexander, E. D. Getzoff, and R. A. Lerner are in the Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, CA 92037.

Identification of antigenic sites. MHr is a member of the hemerythrin family of proteins that function as oxygen carriers in four invertebrate phyla. MHr lacks homology to rabbit proteins and therefore complications from immunological self-tolerance are reduced. The structural fold for MHr is an antiparallel bundle of four α-helices (named A, B, C, and D in sequence order) surrounding a two-iron (F_e) center at the active site. The amino terminal residues form a loop of nonrepetitive secondary structure and shorter loops are found between the helices and at the carboxyl terminus. The simple four-helix bundle fold results in most local surface regions forming from one, two, or at most three linear segments of the polypeptide chain. Three-dimensional x-ray crystallographic coordinates have been refined at high resolution for monomeric MHr from the marine worm *Themiste zostericola* and octameric hemerythrin from *Themiste dyscritum*. Analysis of these two structures enabled Sheriff *et al.* (*1*) to determine a consistent set of temperature factors (after correction for crystal contacts) describing the relative mobility of different parts of MHr.

To analyze the complex immune response to MHr, we used selective antigens (peptide homologs of the protein sequence from 6 to 14 residues long) to precisely define individual specificities within polyclonal antisera to the native protein (anti-MHr). Polyclonal antisera reveal the total anti-MHr response of each rabbit and avoid unusual monoclonal specificities not favored in vivo (*2*). Thus, this method should complement other immunological methods where monoclonal antibodies are used, including both crystallographic (*3*) and fine specificity analyses (*4*). Our detection system assays the antigenicity of protein sites (their ability to bind existing antibodies) with peptide antigens. Reactive peptides contain individual residues important for the binding interaction, as well as covalently connected residues forming some of their environmental context. Some sites may be incomplete and only a subset of assembled, conformational determinants may be identified; however, the results presented below suggest that partial conformational sites are identified by this method.

Antibodies to the intact MHr protein were raised in seven outbred rabbits (*5–7*) and reacted with all five- and six-residue overlapping peptide homologs of the MHr sequence, as well as selected longer peptides named by range of residues: 3 to 16, 7 to 16, 22 to 35, 26 to 35, 37 to 46, 42 to 51, 57 to 66, 63 to 72, 69 to 82, 73 to 82, 96 to 109, 100 to 109, and 107 to 116 (*8, 9*). MHr antigenicity assessed against hexapeptides using enzyme-linked immunosorbent assays (ELISA) is shown in Fig. 1, A to C. Tests with pentapeptides gave consistent results, and hexapeptide reactivity patterns have been shown to be consistent with those of longer, seven- to ten-residue peptides (*7*). Competition assays, both in solution and in solid phase (*10*), were used to show that the interaction of the antibody with the peptide resembled that with the intact protein. The native protein absorbed or competed for binding of anti-MHr to peptides covering all parts of the MHr sequence exhibiting significant immunological reactivity.

Variation in the immune response. Collectively, the antibody populations from the seven antisera reacted with hexapeptides encompassing all 118 residues of the MHr sequence, although not every hexapeptide was antigenic. This result, that each residue of MHr appears in an antigenic protein segment, agrees with the idea (*4, 11*) that the entire surface of a protein is antigenic. The probability of obtaining antibodies to any given site, however, varied greatly both among the sites and among different animals (Fig. 1A). The specific reactivities of the seven antisera ranged from a total of 12 (rabbit No. 6) to 50 recognized hexapeptides (rabbit No. 2) of 113. Immunoprecipitation assays (*9*) with anti-MHr from rabbits Nos. 1, 2, and 3 performed on four longer peptides (residues 3 to 16, 63 to 72, 96 to 109, and 107 to 116) matched the pattern of hexapeptide reactivities: peptide 3 to 16 reacted with all three antisera, peptide 63 to 72 reacted with antisera No. 2 and No. 3, peptide 96 to 109 was unreactive, and peptide 107 to 116 was reactive with antiserum No. 2. Thus, positive immunoprecipitation assays corresponded to more hexapeptide reactivities. The highest reactivity measured by immunoprecipitation matched the most corresponding hexapeptide reactivities: seven out of the nine hexapeptides within the longer peptide encompassing residues 3 to 16 reacted with antiserum from rabbit No. 2.

Higher titers (Fig. 1B) and frequent reactivity (Fig. 1C) were often associated in our database. Comparison of the most and least antigenic hexapeptide sequences, based on mean titers and number of responding rabbits, shows agreement on four antigenic minima (approximately centered around hexapeptides starting at residues 30, 48, 73, and 99) and six antigenic maxima (4, 54, 65, 80, 88, and 110). Comparison of the observed distribution of reactivities with that of a randomly generated, but numerically equivalent, set of reactivities by χ^2 analysis confirmed that the observed clustering of reactive hexapeptides (antigenic maxima) and of unreactive hexapeptides (antigenic minima) is significant ($P < 0.001$) (*12*). Where applicable, ELISA assays of anti-MHr reactivity with the 13 longer peptides coupled to protein or polylysine carriers (*9*) gave results consistent with these hexapeptide data.

No antibody specificity was common to all seven rabbits: of the 113 hexapeptide homologs of the MHr sequence, four peptides were recognized by six of the seven antisera, whereas 31 reacted with only a single antiserum, and 35 were unrecognized by any antiserum (Fig. 1C). Such variability within a species highlights the need to evaluate a consensus response from multiple animals. Frequently recognized regions illustrate common aspects of antigenic recognition important for understanding the chemistry of the interaction.

The chemical nature of the most reactive and least reactive sites of MHr. Intrinsic stereochemical features that might distinguish the immunologically most reactive and least reactive sites of a protein were examined. First, chemical and structural properties for each hexapeptide were plotted as a function of the first sequence number of the hexapeptide (Fig. 1, D and E) and aligned with immunological titer and frequency data (Fig. 1, B and C). Plots of mobility, packing density (inverted), shape accessibility, and to a lesser extent surface exposure, show overall patterns that match those for experimentally determined immunological reactivity. This was confirmed statistically by a comparison of less antigenic (those recognized by one or no rabbit antisera) and more antigenic hexapeptides (two or more antisera). These two groups differ significantly (t test, df = 111, $P \leq 0.001$) with regard to structure-associated properties: antigenic frequency correlates best with mobility ($t = 5.75$), followed by packing density ($t = 4.66$), shape accessibility ($t = 3.87$), and surface exposure ($t = 3.23$). Hydrophilicity differences were less significant ($t = 1.86$, $P > 0.05$). All of the plotted stereochemical and antigenic properties do share one feature: the deepest minimum occurs between hexapeptides 99 and 106. This maps to the only truly buried part of the MHr structure, the residues of the D helix located underneath the NH_2-terminal loop. Small shifts in the alignment of critical points between the structural and immunological plots highlight the three-dimensional nature of protein epitopes. Local maxima and minima of chemical, structural, or immunological properties traced along the one-dimensional amino acid sequence will not necessarily identify true maxima and minima for a three-dimensional protein.

To identify correlations of immunological reactivity with structure-associated parameters in a three-dimensional way (Table 1), we divided the residues of MHr into three categories according to the reactivity with hexapeptides: (i) most reactive (41 percent)—occur-

Table 1. Molecular surface chemistry and antigenic reactivity. Individual residues are divided into three classes of antigenic reactivity as defined in the text. All stereochemical parameters listed are functions of the three-dimensional structure, not the linear sequence, and are derived from x-ray crystallographic results (*1*). Mobility is calculated from the average main-chain temperature factors as corrected for crystal contacts (*1*). The percentage of the total exposed surface area is calculated from the molecular surface area (*24, 25*) summed for each category as follows: negative, neutral, or positive areas based upon the electrostatic potential calculated from partial charges on all atoms (*13*); hydrophobic (carbon) or hydrophilic (noncarbon) atoms; and side-chain or main-chain atoms. The total exposed surface area and surface area per residue are calculated from the molecular surface area accessible to a water-sized (1.4 Å radius) probe sphere. Exposed side-chain area is expressed as a percentage for hydrophobic (carbon), hydrophilic (noncarbon), or all side-chain atom surface area exposed in the context of the protein structure, calculated from the total side-chain surface area in the absence of neighboring protein residues (*27*).

Reactivity	No. of residues	Mobility (Å²)	Total exposed surface area (%)							Surface (Å²)		Side-chain exposure (%)		
			Potential			Hydrophobic	Hydrophilic	Side chain	Main chain	Total exposed	Per residue	Per residue	Hydrophobic	Hydrophilic
			−	0	+									
Most	48	30.0	48	34	19	66	34	75	25	2337	49	32	29	41
Average	39	27.6	37	40	23	61	39	73	27	2005	51	34	31	42
Least	31	23.2	23	42	35	65	35	83	17	1209	39	23	17	48
MHr	118	27.0	39	38	24	64	36	76	24	5552	47	30	26	43

ring in at least one hexapeptide recognized by four or more rabbit antisera; (ii) least reactive (26 percent)—not occurring in any hexapeptide recognized by more than one antiserum; (iii) average (33 percent)—all others. The most reactive positions include 4 to 10, 54 to 61, 63 to 72, 79 to 85, 88 to 97, and 110 to 115. The least reactive positions are 16 to 17, 28 to 36, 46 to 52, 75, 99 to 108, and 117 to 118. Ten- and 14-residue peptides encompassing the longer stretches of least reactive positions (26 to 35, 42 to 51, 96 to 109, and 100 to 109) were also unreactive with anti-MHr when coupled to protein carriers and assayed by ELISA. In the inverse experiment, antibodies to peptides consisting of these least reactive sequences reacted poorly with MHr (*6*).

Mobility, electrostatic potential, and hydrophilicity. Qualitatively, the trend across the three categories of immunological reactivity coincides with changes in mobility (Table 1). Decreasing mobility corresponds to decreasing immunological reactivity from most reactive, through average, to least reactive positions. The mobilities of the most reactive and of the least reactive residues differ significantly from the mean (Monte Carlo method, $P < 0.001$) (*12*). The major maxima and minima of the immunological profiles align best with those from plots of mobility and inverted packing density (Fig. 1D). Although the mobility curve is smoother, all five of its peaks overlap with maxima in the frequency (Fig. 1C) and mean titer (Fig. 1B) plots; the two prominent temperature factor minima correspond to reactivities with at most one antiserum and titers of less than 500. The peak in reactivity near hexapeptide 54 falls in a local mobility minimum, but this region remains more mobile than the global minima near hexapeptides 27 and 102. The packing density profile (inverted), although more serrated, resembles the overall mobility profile; the major differences are an additional minimum near hexapeptide 13, and a shift in the COOH-terminal peak.

The most frequently recognized sites in MHr generally correspond to regions of negative electrostatic potential (Table 1). The percentage of the total surface area associated with negative electrostatic potentials is greatest for the most reactive positions (Monte Carlo method, $P = 0.02$) (*12*) and least for least reactive positions ($P = 0.005$) and the converse is true, but less significant, for positive electrostatic potentials. In each case, the percentages of total exposed surface area associated with the most and the least reactive positions differ by about a factor of 2. The percentages of negative and positive surface area associated with regions of average reactivity closely match those found for the whole protein. When examined with the use of computer graphics (*13*), all of the most antigenic positions show electrostatic fields with energies greater than thermal energy (3/2 kT) extending beyond 6 Å from the protein surface. These fields have negative electrostatic potential near the immunologically most reactive positions 4 to 10, 54 to 61, and 79 to 85; and positive potential near the most reactive positions 63 to 72, 88 to 97, and 110 to 115. However, by far the largest fields with positive electrostatic potential are associated with least reactive positions 28 to 36 and 46 to 52.

Although the percentage of the total exposed surface area contributed by hydrophilic (noncarbon) atoms is nominally largest for the residues of average reactivity (Table 1), all three categories of antigenic reactivity have percentages of hydrophobic and hydrophilic surface area similar to that for the whole protein. Furthermore, the hydrophilicity profile (Fig. 1D) does not match the pattern of immunological reactivity as well as the mobility and inverted packing density profiles do. The highly reactive site at hexapeptide 4 is a hydrophilicity minimum, and the highest hydrophilicity peaks at hexapeptides 22 and 74, which correspond to valleys in the mobility and inverted packing density plots, show only average and low reactivity values, respectively, for both titer and frequency.

Shape, surface exposure, and secondary structure. Half (59) of the residues of MHr are accessible to a probe sphere of 15 Å radius and another seven residues are accessible to spheres of 11 to 14 Å radius, suggesting that 56 percent of the residues are directly available to the ends of a fairly flat antibody binding region (*3*). In contrast, the shape accessibility of 34 residues (29 percent) is restricted to spheres of 2 Å radius or smaller. The most and least antigenic positions are not directly distinguished by their sphere-accessibility radius: all contiguous stretches of most antigenic positions include one or more residues only accessible to a solvent-sized sphere of 2 Å or less radius, and all contiguous stretches of least antigenic positions include one or more residues accessible to a sphere of 15 Å or more radius. While the least antigenic positions in MHr are thus accessible to antibody binding, they are also all closely associated with narrow grooves containing tightly bound water molecules identified in the x-ray crystallographic structure (*1*).

The plot of shape accessibility as a function of sequence (Fig. 1E) shows similarities with mean titer and frequency profiles, as well as with plots of average main-chain temperature factors, inverted packing density, and surface exposure. However, compared to the mobility and inverted packing density profiles (Fig. 1D), the shape-accessibility plot shows less ability to distinguish between the peak of antigenicity near hexapeptide 54 and the minimum near hexapeptide 30. Continuing this trend, the plot of solvent-exposed surface area (Fig. 1E) shows more of a minimum near hexapeptide 54 (a reactivity peak) than near hexapeptide 30 (a reactivity minimum).

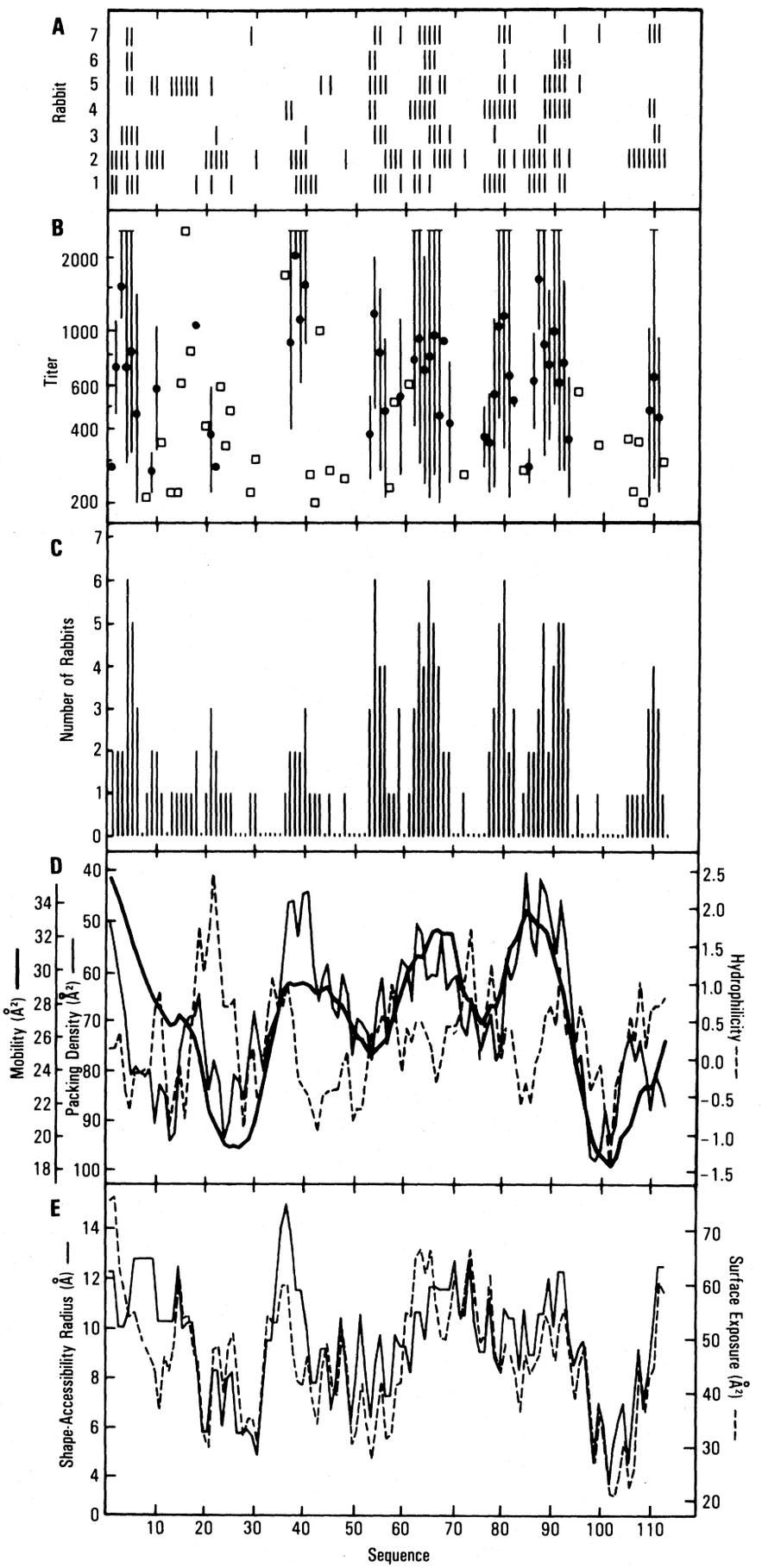

Fig. 1. Profiles of antigenic response (A to C) and stereochemical properties (D to E) as a function of all possible MHr hexapeptides. Antibody recognition of MHr is measured by the reactivity of anti-MHr to hexapeptide analogs of the MHr sequence, which compete for antibody binding to the native protein. Each parameter is plotted at the sequence number corresponding to the first residue of the relevant hexapeptide. (**A**) The reactivity pattern specific to each of the seven rabbit antisera. Responses to any hexapeptide are indicated with a short vertical bar. Immunological reactivity of the first 52 hexapeptides of the sequence was highly variable, with three antisera reacting with only two or three hexapeptides and one antiserum reacting with 20 hexapeptides. Hexapeptides 7 through 52 (as numbered by the first residue) also showed infrequent reactivity, with 18 of 46 hexapeptides unrecognized by any antiserum, 19 recognized by a single antiserum, and none seen by more than three antisera. For hexapeptides 53 through 93, the immune response was more consistent and more frequent: 21 of the 41 hexapeptides react with at least three of the antisera, and each of the seven antisera reacted with between 10 and 22 hexapeptides. The least frequently reactive region of the sequence encompasses hexapeptides 94 through 104; 9 of these 11 hexapeptides were not recognized by any of the seven antisera. (**B**) The individual titer (square) or, where more than one serum reacted, the geometric mean titer (circle) and range of titer (vertical bar) of the antisera. Titers less than 200 (twice the slope of the background) were ignored and those greater than 2000 were not further resolved. (**C**) The frequency of the antigenic response given by the number of rabbit antisera that react with each hexapeptide. High frequency and high titer are often associated. However, peptides corresponding to protein residues 3 to 8, 16 to 21, and 38 to 43 are exceptions for which only one or two antisera reacted, but all had high titers (1120 to >2000). In contrast, titers range widely in regions where 6 of the 7 antisera reacted: the narrowest range of titers is more than fourfold (480 to 1970). (**D**) Mobility [average main-chain temperature factors corrected for crystal contacts (*1*)], inverted packing density (surface area buried in noncovalently bonded close-packing), and hydrophilicity [parameters from Hopp and Woods (*22*)]. Packing density, sphere accessibility radius, and surface exposure calculations were based on individual van der Waals radii (see legend to Fig. 2). (**E**) Shape accessibility [radius of the largest sphere that can access a side chain without collision or interpenetration of any other residue (*23*)], surface exposure [the solvent-exposed molecular surface area (*24, 25*)]. Sphere accessibility radii greater than 15 Å are not distinguished. Surface exposure and packing density were assessed from a water-sized (1.4 Å radius) probe sphere. Except for hydrophilicity, all plotted stereochemical parameters depend on the three-dimensional structure, rather than being defined identically for a given amino acid type. Structural parameters are averaged over six residues so as to correspond with immunological data based on hexapeptides. Hexapeptides are defined from the MHr sequence (*8*) given below:

```
GWEIPEPYVW DESFRVFYEQ LDEEHKKIFK 30
GIFDCIRDNS APNLATLVKV TTNHFTHEEA 60
MMDAAKYSEV VPHKKMHKDF LEKIGGLSAP 90
VDAKNVDYCK EWLVNHIKGT DFKYKGKL   118
```

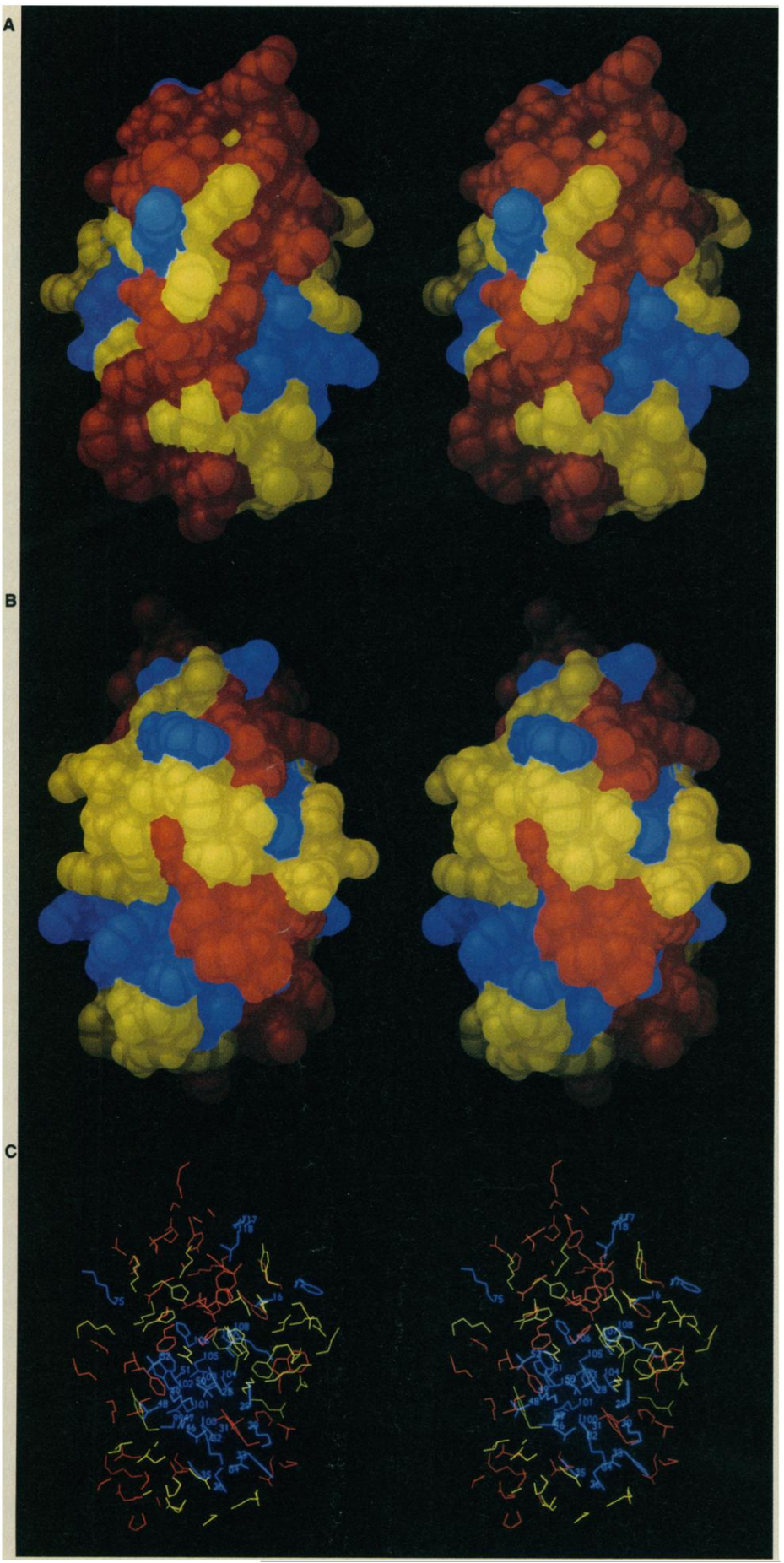

Fig. 2. Stereo pairs showing computer graphics models of MHr color-coded by immunological reactivity (red, most reactive; yellow, average; and blue, least reactive). Raster color graphics images of opposing sides (A and B) of the solid external molecular surface of MHr show that the most highly reactive surface regions form superassemblies (three-dimensional clusters) that wrap around one side (A) of the protein and almost connect on the other side (B). The four α-helices are oriented vertically, with the two-iron active site near the top. In view (**A**), the B (left) and C (right) helices appear in front connected by the BC loop (top) and in view (**B**), the NH_2-terminal loop appears in front of the A (left) and D (right) helices. The molecular surface accessible to a water-sized (1.4 Å radius) probe sphere was calculated analytically with AMS and RAMS (*24*). This surface is composed of convex regions formed by the solvent-accessible van der Waals surface of individual atoms and concave regions smoothing over small gaps and crevices inaccessible to the probe sphere. Molecular surfaces are especially appropriate for examining intermolecular interactions because they focus on surface regions accessible to interacting molecules. A skeletal model (**C**) of all MHr side chains (no main-chain bonds), shows that the least reactive residues (blue lines identified by residue number at the Cα position) form one major cluster exposed on two opposite sides of the protein and largely surrounded elsewhere by side chains with average (yellow lines) and most (red lines) immunological reactivity. This core region of low reactivity extends through the end of the helix bundle distant from the two oxygen-binding iron atoms. View (C), with the A and B helices in front, is intermediate between orientations (A) and (B) and related to them by 45° rotations about the vertical axis. This vector graphics image was rendered with the use of programs GRAMPS and GRANNY (*26*). Molecular surface, packing density, sphere accessibility radius, and exposed area calculations used individual van der Waals atomic radii that included implicit hydrogen atoms: O(sp, carbonyl), 1.40 Å; O(sp^2, carboxyl), 1.40 Å; OH, 1.60 Å; N(sp^2), 1.54 Å; NH, 1.70 Å; NH_2, 1.80 Å; NH_3, 2.00 Å; CH, 2.00 Å; CH_2, 2.00 Å; CH_3, 2.00 Å; C(sp^2), 1.74 Å; C(sp^3); 1.86 Å; S, 1.80 Å; SH, 1.85 Å; and Fe, 1.72 Å.

The exposed surface area per residue is the lowest for the least immunoreactive residues (Table 1), supporting a relation between surface exposure and antigenicity (*14*). However, this trend is not continued: the most antigenic residues do not have more exposed surface area per residue than those with average antigenicity.

When total solvent-exposed surface area is divided into side-chain and main-chain contributions (Table 1), the least reactive residues are again distinguished (Monte Carlo method, $P = 0.04$). The ratio of main-chain to side-chain surface area is similar for positions with most and average reactivity and lower for the least reactive positions. In the folded protein, the side-chain surface exposure can also be measured as a percentage of the total surface exposure of the same side-chain conformation in the absence of the other protein residues. This percentage side-chain exposure (Table 1) is marginally lower for the least reactive positions ($P = 0.10$), largely as a result of a decreased hydrophobic (carbon atom) contribution ($P = 0.02$). Similarly, the immunological reactivity profiles match carbon exposure better than noncarbon exposure; the least reactive positions seem to associate with areas of concave surface shape in which the most exposed atoms are frequently the ends of polar side chains.

Although secondary structure is not a dominant factor in the degree of antigenicity, the least reactive residues are predominantly helical. Twenty-seven least reactive residues are distributed among the four helices, whereas the remaining four residues occur in the NH_2- or COOH-terminal loops. This bias may be due in part to the peptide assay system, since the 3.6-residue helical repeat (visible in most of the structural plots) may make a hexapeptide less able to mimic a surface region or conformational determinant of a protein. However, except for the lack of most reactive residues in the A helix, each of the four helices, as well as the longest loop, contains residues in all three categories of immunological reactivity. For example, the NH_2-terminal loop includes residues with most (4 to 10), average (1 to 3 and 11 to 15), and least (16 to 17) reactivity. The shorter loops taken together also include all three categories of reactivity. Areas of average reactivity occur in every helix and in every connecting or terminal loop region of the MHr secondary structure.

To assess the topography and three-dimensional arrangement of most reactive versus least reactive positions, the three categories of immunological reactivity were mapped by color-coding onto the solvent-exposed molecular surface of the protein by means of computer graphics (Fig. 2). Although these categories were identified from peptide homologs of the MHr sequence, the immunological reactivity varies relatively smoothly over the three-dimensional surface of the protein as well as the linear sequence. Along the polypeptide chain, the most and least reactive residues are always separated by residues with average reactivity. On the molecular surface, the most reactive residues cluster into patches, as do the least reactive residues, and the two sets of patches are usually separated by surface area from residues of average reactivity. The most frequently recognized positions form superassemblies wrapping around the MHr molecule (Fig. 2, A and B) to encompass the more flexible and less tightly packed areas of the protein. Residues from least reactive positions cluster to a cylindrical core within the most tightly packed portion of the protein extending through the molecule to include residues at two surface regions (Fig. 2C). The clustering of residues with like antigenic properties and the relatively smooth variation of immunological reactivity at the molecular surface found by this work suggest that conformational or assembled determinants may often consist of residues contributed by parts of two or more linear sequences, as has also been shown by others (*3, 15*).

Structural biases for antigenicity. Our results show that, although the entire MHr molecule can be antigenic, stereochemical and biophysical properties bias the frequency of antiprotein responses, as assayed with peptides. The protein chemistry of the most antigenic superassemblies is dominated by high mobility and low packing density (allowing local side-chain rearrangements and induced fit) (*16*), convex shape (facilitating the release of water molecules), and significant negative electrostatic potential at the molecular surface. Any given site may not have all of these attributes, but each feature appears to bias a given site toward antigenicity. The least antigenic regions are usually the more ordered regions of the protein with tightly packed internal side chains and concave surface shape sequestering bound water molecules. In general, structural parameters more clearly differentiate the least reactive positions, which have fewer residues in loops of nonrepetitive secondary structure, less exposed surface area per residue, a smaller percentage of surface area contributed by main-chain atoms, and a decreased hydrophobic contribution to percent side-chain exposure.

Shape and mobility have been previously implicated in the antigenic response to proteins (*6, 14, 15, 17*). In the static structure of MHr, every hexapeptide has exposed surface area and every residue belongs to at least one hexapeptide accessible to a spherical probe the size of an antibody binding domain. Novotný *et al.* (*14*) have proposed that antibody accessibility (measured with a spherical probe of fixed 10 Å radius) is the primary intrinsic determinant of antigenicity and thus predict little or no antigenicity for MHr residues 53 to 60. In our experiments, the hexapeptide starting at residue 54 is one of the most frequently antigenic (recognized by six of the seven antisera), suggesting that for a small single-domain protein, shape accessibility is an insufficient criterion for predicting antigenicity, although it may be quite important for large multimeric structures such as virus capsids. Barlow *et al.* (*15*) have suggested that protein antigenic sites delineated by peptides fall in loops or protruding surface regions characterized as "continuous," that is, >50 percent surrounded (within a 10 Å radius) by atoms of neighboring ($i \pm 3$) residues. For MHr, this method predicts "best" antigenic sites in two of the four antigenic minima and only four of the six antigenic maxima mapped with hexapeptides, suggesting that surface continuity is not an overriding factor.

The most critical aspect of shape for protein antigenicity appears to be the association of the least reactive residues with neighboring, narrow, concave surface regions containing bound water molecules. Thus, correlations of antigenicity with shape may reflect the relative difficulty of displacing bound water molecules. Antibody-antigen interaction depends on complementary interactions between the surfaces of the two proteins and also on the driving force resulting from the exclusion of water. The superassemblies of antigenic sites identified here have predominantly convex surface areas, where individual water molecules have fewer contacts with the protein, so that their release is kinetically favored and less likely to be rate-limiting.

In MHr, the most antigenic sites tend to have significant negative electrostatic potentials. This may be biased by the net negative charge on MHr, as many antibodies are charge-complementary to their antigens (*4, 18*). However, many haptens known to bind antibodies (nitrates, carboxylates, sulfonates, arsonates, and phosphate compounds) are also negatively charged. Moreover, mutants that escape monoclonal antibody binding often appear to have charge changes, frequently resulting from mutations to positive residues (*19*). Thus, the immune response is apparently sensitive to electrostatic forces, whether or not the bias is toward regions of negative potential. Electrostatic attraction may increase the rate of antibody-antigen complex formation by promoting facilitated diffusion and may also stabilize an initial complex, allowing subsequent slower steps to occur before the antigen and antibody diffuse apart. Electrostatic forces are very strong in the absence of water and may stabilize the final antigen-antibody interface significantly, as in other

known protein-protein interfaces (*18, 20*).

In a system as diverse as the immune response, which has the potential to generate some 10^7 immunoglobulins with affinities for antigens spanning at least six orders of magnitude (*21*), many patterns of antibody-antigen interactions can be expected. Thus, one would expect to find antibodies that do not match any given observed trend. The surprise in the analysis of the antigenic sites characterized here is not in their variations, but in their commonalities, which show that stereochemical properties of the antigen bias immunological reactivity. The identification of common attributes is a preliminary, but necessary, step in both the prediction and understanding of antigenic recognition.

REFERENCES AND NOTES

1. W. A. Hendrickson, G. L. Klippenstein, K. B. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2160 (1975); R. E. Stenkamp, L. C. Sieker, L. H. Jensen, *Acta Crystallogr. Sect. B* **39**, 697 (1983); S. Sheriff, W. A. Hendrickson, R. E. Stenkamp, L. C. Sieker, L. H. Jensen, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1104 (1985); S. Sheriff, W. A. Hendrickson, J. L. Smith, *J. Mol. Biol.*, in press.
2. D. W. Metzger, L.-K. Ch'ng, A. Miller, E. E. Sercarz, *Eur. J. Immunol.* **14**, 87 (1984).
3. A. G. Amit, R. A. Mariuzza, S. E. V. Phillips, R. J. Poljak, *Science* **233**, 747 (1986); R. Jemmerson and Y. Paterson, *ibid.* **232**, 1001 (1986).
4. D. C. Benjamin *et al.*, *Annu. Rev. Immunol.* **2**, 67 (1984); J. A. Berzofsky, *Science* **229**, 932 (1985).
5. Although immunization protocols (*6, 7*) differed between rabbits Nos. 1, 2, and 3 (Scripps) and rabbits Nos. 4, 5, and 6 (CSL), the responses were not distinguishable by this attribute. Rabbits at Scripps (*6*) received three injections of 200 μg purified MHr: at day 0, diluted 1:1 in complete Freund's adjuvant, injected subcutaneously along the back; at day 14, diluted 1:1 in incomplete Freund's adjuvant, subcutaneously; and at day 21, in alum at 4 mg/ml, intraperitoneally. The rabbits were bled at day 30. Rabbits at CSL (*7*) received two intramuscular injections of 5 mg of MHr, each: at day 0, diluted 1:2 in complete Freund's adjuvant; at day 14, diluted 1:2 in incomplete Freund's adjuvant. The rabbits were bled at day 28. No antibody to MHr was detected in any "preimmune" antiserum.
6. J. A. Tainer *et al.*, *Nature (London)* **312**, 127 (1984); J. A. Tainer, E. D. Getzoff, Y. Paterson, A. J. Olson, R. A. Lerner, *Annu. Rev. Immunol.* **3**, 501 (1985).
7. H. M. Geysen, R. H. Meloen, S. J. Barteling, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3998 (1984); H. M. Geysen, S. J. Barteling, R. H. Meloen, *ibid.* **82**, 178 (1985); S. J. Rodda, H. M. Geysen, T. J. Mason, P. G. Schoofs, *Mol. Immunol.* **23**, 603 (1986).
8. The sequence of MHr used for peptide synthesis was that of Klippenstein [G. L. Klippenstein, J. L. Cote, S. E. Ludlan, *Biochemistry* **15**, 1128 (1976)] corrected by a sequence inversion of residues 34 and 35 (J. L. Smith and W. A. Hendrickson, unpublished results). At CSL, comprehensive sets of overlapping pentameric and hexameric peptides were synthesized on polyethylene rods (*7*), with the use of both the *t*-butyloxycarbonyl (BOC) and 9-fluoroenylmethyloxycarbonyl (FMOC) strategies. At Scripps, thirteen 10- and 14-residue peptides, covering most of the MHr sequence (*6*), were synthesized as described [R. A. Houghten, W. C. Chang, C. H. Li, *Int. J. Pept. Protein Res.* **16**, 311 (1980)].
9. At CSL, antisera were diluted appropriately and assayed by ELISA against rod-coupled peptides. At each dilution of a given antiserum, the test background was obtained from the nonreacting peptides (*7*). Analysis of the corrected extinctions showed that they were linearly dependent on the concentration of the antiserum used in the test (r^2 values typically in the range 0.97 to 1.00). Titers were estimated from the data to be the reciprocal of the dilution of the antiserum that would give an extinction of 1.00 above the test background. Titers less than 200 (twice the slope of the background) were ignored and those greater than 2000 were not further resolved. At Scripps the reactivity of anti-MHr was tested with peptides conjugated to one of four carrier molecules: keyhole limpet hemocyanin (KLH), bovine serum albumin (BSA), and two polymer sizes of poly-L-lysine. Conjugated peptides (10 to 20 per carrier molecule) were adjusted to 50 pmol per well of microtiter plates, and reactivity was measured by ELISA (*6*). For immunoprecipitation assays, anti-MHr was incubated with ^{125}I-labeled peptides in phosphate-buffered saline (PBS) solution, and antigen-antibody complexes were precipitated with *Staphylococcus aureus* (*6*).
10. For protein competition experiments in solution, which were done at Scripps, anti-MHr was first incubated with 0.1, 1, 2, 5, and 10 μg of unlabeled MHr for 60 minutes at 4 °C, then reacted with ^{125}I-labeled 10- and 14-residue peptides in PBS, and assayed by immunoprecipitation. MHr absorption of anti-MHr binding to 10- and 14-residue peptides was also measured by ELISA. Portions (50 μl) of serially diluted anti-MHr were incubated overnight with and without 50 pmol of MHr per well of microtiter plates. On the following day, 25 μl of the antiserum solutions were transferred to microtiter plates, containing 50 pmol of conjugated peptides per well, and the plates were subjected to ELISA assay (*9*). Peptides were scored as positive in all cases where prior incubation of the antisera with MHr contributed to loss of reactivity with the peptides. At CSL, antisera for competition experiments were diluted, on the basis of titers in previous tests, to a level expected to give an optical density of 1 above background in ELISA's. MHr was added to give a final concentration of 0.1 mg/ml, and the solution was incubated at 4°C overnight. The competed antisera in the presence of MHr were then tested in duplicates by ELISA against rod-coupled peptides (*9*) in parallel with controls prepared without added MHr.
11. Z. A. Moudallal, J. P. Briand, M. H. V. Van Regenmortel, *EMBO J.* **4**, 1231 (1985).
12. A χ^2 test was applied to compare the experimental immunological reactivity patterns of the seven rabbits to duplicate distributions determined by averaging 100 randomly generated sets of responses, with the number of peptides for each rabbit fixed at the observed values. Using the five frequency categories (none to four rabbits) with sufficient responses (more than four peptides) in the random distribution we obtained $\chi^2 = 18.95$, giving, for four degrees of freedom, a probability <0.001 that the experimentally observed distribution occurred randomly. The higher-than-expected incidence of both frequently reactive (four, five, or six rabbits) and unreactive (no rabbits) hexapeptides indicates the statistical validity of consensus aspects of the measured antigenic responses in rabbits. Furthermore, hexapeptides responded to with a given frequency (either high or low) were more likely to be numerically adjacent in the experimental, rather than the randomly generated, data. To assess the statistical significance of the trends shown in Table 1, we used Monte Carlo methods to group the 118 residues of MHr randomly into most (48 residues), average (39 residues), and least (31 residues) reactive positions, 1010 times. From these expected random distributions, the mean and standard deviation for each structure-associated property was defined and used to assess the statistical significance of the experimentally observed values (see text).
13. E. D. Getzoff *et al.*, *Nature (London)* **306**, 287 (1983); S. J. Weiner *et al.*, *J. Am. Chem. Soc.* **106**, 765 (1984).
14. J. Novotný *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 226 (1986); D. W. Fanning, J. A. Smith, G. D. Rose, *Biopolymers* **25**, 63 (1986).
15. J. M. Thornton, M. S. Edwards, W. R. Taylor, D. J. Barlow, *EMBO J.* **5**, 409 (1986); D. J. Barlow, M. S. Edwards, J. M. Thornton, *Nature (London)* **322**, 747 (1986).
16. E. D. Getzoff *et al.*, *Science*, **235**, 1191 (1987).
17. G. R. Moore and R. J. P. Williams, *Eur. J. Biochem.* **103**, 543 (1980); E. Westhof *et al.*, *Nature (London)* **311**, 123 (1984); R. J. P. Williams and G. R. Moore, *Trends Biochem. Sci.* **10**, 96 (1985).
18. M. J. Crumpton, in *The Antigens*, M. Sela, Ed. (Academic Press, New York, 1974), vol. 2, p. 133; F. Karush, *Adv. Immunol.* **2**, 1 (1962).
19. H. M. Geysen, unpublished data.
20. C. Chothia and J. Janin, *Nature (London)* **256**, 705 (1975); E. D. Getzoff, J. A. Tainer, A. J. Olson, *Biophys. J.* **49**, 191 (1986).
21. M. W. Steward and J. Steensgaard, in *Antibody Affinity* (CRC Press, Boca Raton, FL, 1983), p. 69; E. D. Max, in *Fundamental Immunology*, W. E. Paul, Ed. (Raven, New York, 1984), p. 191.
22. T. P. Hopp and K. R. Woods, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3824 (1981); T. P. Hopp, *J. Immunol. Methods* **88**, 1 (1986).
23. E. D. Getzoff and J. A. Tainer, unpublished method.
24. M. L. Connolly, *Science* **221**, 709 (1983); *J. Appl. Crystallogr.* **16**, 548 (1983).
25. B. K. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971); F. J. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977); F. M. Richards, *Methods Enzymol.* **115**, 440 (1985).
26. T. J. O'Donnell and A. J. Olson, *Comput. Graphics* **15**, 133 (1981); M. L. Connolly and A. J. Olson, *Comput. Chem.* **9**, 1 (1985).
27. A. Shrake and J. A. Rupley, *J. Mol. Biol.* **79**, 351 (1973).

28. We thank S. Sheriff, W. A. Hendrickson and J. L. Smith for use of their MHr coordinates and temperature factors corrected for crystal contacts. We also thank R. Bott, D. R. Davies, W. A. Hendrickson, W. P. Jencks, N. Klinman, A. J. Olson, E. Padlan, Y. Paterson, M. E. Pique, J. Sayre, S. Sheriff, and A. Tramontano, for helpful discussions regarding the manuscript. Supported in part by grants from the National Institutes of Health to E.D.G. and R.A.L.

11 July 1986; accepted 31 December 1986