wild-type GAL4 or GAL4-B17—the latter activator is GAL4 (1-147) fused to an acidic activating region encoded by an *Escherichia coli* genomic fragment[2]. Figure 2 also shows that GAL4-VP16 worked, albeit about 10-fold less efficiently, when the UAS_G was positioned at −1180 or +1850. Using a different cell line (CHO-DUKX (ref. 6)), GAL4-VP16 worked at about the same level of efficiency from all three positions (Fig. 1*c*). Activity was not detected in either cell line for either GAL4 or GAL4-B17 when the UAS_G was positioned at −1180 or +1850 and in no case did GAL4 (1-147) activate, confirming previous results[6]. The start sites of the induced RNAs for all constructs were identical to those of the glucocorticoid-receptor-stimulated transcripts initiated from the wild-type MMTV LTR (not shown). Immunoprecipitation and western blotting of the various activator proteins revealed that all were produced at equal levels (within a factor of two or three) in both cell types (data not shown). In all of these experiments the reporter bears a binding site for the mammalian activator nuclear factor (NF1) near the TATA sequence, and we do not know what contributions, if any, this protein made to the activation we observed. It is possible, for example, that the differences observed between the two cell lines reflect different concentrations of this auxiliary factor.

GAL4-VP16 activated only slightly more efficiently from a UAS_G (which includes four GAL4 binding sites) than from a single strong binding site (17 mer, Fig. 1*d*), indicating that activation may be near-maximal from the single site. In addition, GAL4-VP16, bound at position −110, activates transcription about 10-fold more efficiently than the glucocorticoid receptor (GR) bound to the multiple sites found in the wild-type MMTV promoter (Fig. 1d). Finally, Fig 1*d* shows that GAL4-VP16 bound at −110 elicits a level of CAT activity that is about that produced by a CAT gene driven by an adjacent simian virus 40 (SV40) or Rous sarcoma virus (RSV) promoter. For reasons we do not understand, using an SV40 promoter as a measure, we find that wild-type GAL4 is substantially less active than reported by Webster *et al.*[7].

Why is GAL4-VP16 such a powerful activator? It is highly unlikely that the distinctive properties of the molecule can be ascribed to, for example, increased DNA binding. A variety of experiments suggest that GAL4 (1-147) forms stable dimers (Carey, M., Kakidani, H. and M.P., in preparation) and that this fragment, as well as other GAL4 derivatives, efficiently fill the GAL4 binding sites *in vivo*[17]. Instead, we imagine that the acidic region of VP16 interacts with unusual avidity with some component of the transcriptional apparatus, perhaps the TATA-binding factor TFIID (ref. 18). In other experiments, we found that GAL4-VP16, like native VP16, inhibited transcription initiating at promoters lacking GAL4 binding sites. As described elsewhere[19], we believe this inhibitory effect (which we have called squelching[19,20]) is an unavoidable consequence of expression of a strong activating region. If so, the continuous expression at high levels of a powerful activator would be harmful to cells, but such an activator could profitably be used by a virus that expressed it transiently[20].

We are grateful to Steve McKnight for suggesting this experiment. We thank T. Maniatis for helpful comments. This work was supported by grants from the American Cancer Society and National Institutes of Health to M.P. I.S. is a K.M. Hunter P.D.F. of the National Cancer Institute of Canada.

Received 25 May; accepted 2 September 1988

1. Ma, J. & Ptashne, M. *Cell* **48**, 847-853 (1987).
2. Ma, J. & Ptashne, M. *Cell* **51**, 113-119 (1987).
3. Brent, R. & Ptashne, M. *Cell* **43**, 729-736 (1985).
4. Hope, I. & Struhl, K. *Cell* **46**, 885-894 (1986).
5. Gill, G. & Ptashne, M. *Cell* **51**, 121-126 (1987).
6. Kakidani, H. & Ptashne, M. *Cell* **52**, 161-167 (1988).
7. Webster, N., Jin, J. R., Green, S., Hollis, M. & Chambon, P. *Cell* **52**, 169-178 (1988).
11. Triezenberg, S. J., LaMarco, K. L. & McKnight, S. L. *Genes Dev.* **2**, 730-742 (1988).
12. Triezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. *Genes Dev.* **2**, 718-729 (1988).
13. O'Hare, P. & Goding, C. R. *Cell* **52**, 435-445 (1988).
14. Preston, C. M., Frame, M. C. & Campbell, M. E. M. *Cell* **52**, 425-434 (1988).
15. McKnight, J. L. C., Kristie, T. M. & Roizman, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7061-7065 (1987).
16. West, R., Yocum, R. & Ptashne, M. *Molec. cell. Biol.* **4**, 2467-2478 (1984).
17. Giniger, E. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 383-386 (1988).
18. Lin, Y.-S. *et al. Cell* **54**, 659-664 (1988).
19. Gill, G. & Ptashne, M. *Nature* **334**, 721-724 (1988).
20. Ptashne, M. *Nature* (1988).
21. Cato, A., Miksicek, R., Schuz, G., Arnemann, J. & Beato, M. *EMBO J.* **5**, 2237-2240 (1986).
22. Ellis, L. *et al. Cell* **45**, 721-732 (1986).
23. McGeady, M. L. *et al. DNA* **5**, 289-298 (1986).

Structure of antibody hypervariable loops reproduced by a conformational search algorithm

Robert E. Bruccoleri*, Edgar Haber* & Jiří Novotný*

Cellular and Molecular Research Laboratory,
Massachusetts General Hospital & Harvard Medical School, Boston, Massachusetts 02114, USA

The antigen-combining site of antibody molecules consists of six separate loops supported by a conserved β-sheet framework; antibody specificity arises from length and sequence variation of these 'hypervariable' loops[1] and can be manipulated by transferring sets of loops between different frameworks[2]. Irregular loops are the most difficult parts of protein structure to understand and to model correctly[3-6]. Here, we describe two computer experiments where all the hypervariable loops were deleted from X-ray structures of mouse immunoglobulins and reconstructed using the conformational search program CONGEN[7]. A protocol was developed for reconstruction of the hypervariable loops in McPC 603 antibody. Calculated loop conformations were generated and a model of the combining site was built from selected low-energy conformations. We then modelled hypervariable loops in another antibody molecule, HyHEL-5. Both models agreed well with the known crystal structures. Our results hold out promise for the success of future modelling studies of complete antigen-combining sites from amino acid sequences.

Current techniques of loop-structure modelling, and antibody-binding site modelling in particular[8-13], involve a variety of methods. Protein conformation is determined by values of backbone and side-chain torsional angles ϕ (C-N-Cα-C), ψ (N-Cα-C-N) and χ (N-Cα-Cβ-Cγ, Cα-Cβ-Cγ-Cδ and so on). Three-dimensional modelling involves finding specific values for all these torsional degrees of freedom. A particularly promising approach is therefore to use automatic computer algorithms that uniformly sample the complete conformational space of a polypeptide chain segment[7,14-16]. Theoretically, all the backbone and side-chain conformations compatible with the rest of the protein structure can be generated. The lowest free-energy conformation should correspond to the naturally occuring one. In practice, technical problems associated with an exhaustive sampling of conformational space restrict searches to short polypeptide segments only. A successful protocol for hypervariable loop reconstruction could later be extended to automatic modelling of combining sites of unknown three-dimensional structures solely from their amino acid sequences.

We used X-ray atomic coordinates of the mouse immunoglobulin McPC 603 (ref. 17) as our starting structure, and defined endpoints of the hypervariable loops as shown in Fig. 1. These six loops, 46 amino acid residues altogether, were deleted from the structure. The shortest loop, H1, contained five

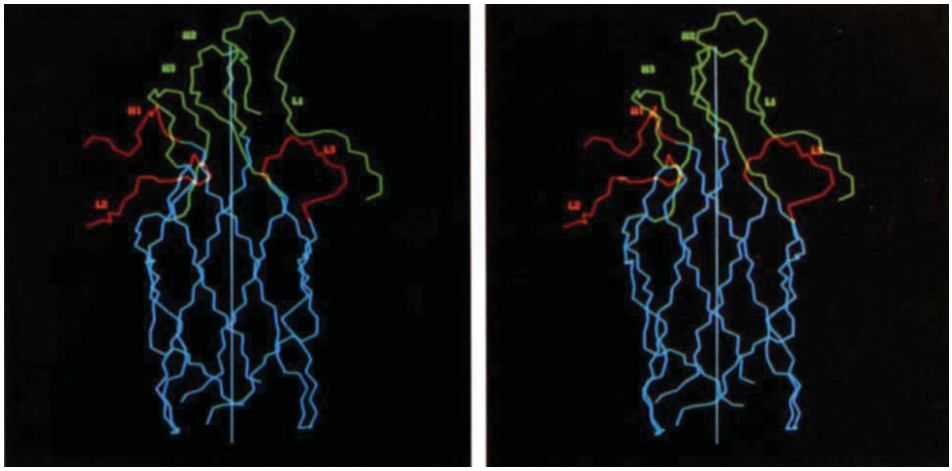

Fig. 1 Stereoscopic view of the hypervariable loops of McPC 603 with the underlying β-sheets. By tight non-covalent association, two curved β-sheets from the VL and VH domains create a twisted, elliptical β-barrel[3,21,22]. The polypeptide backbone of this barrel can be approximated by mathematical hyperboloidal surfaces[22] and the long axis of the least-squares fitted surface becomes a convenient reference axis of the β-barrel. The fitted axis is drawn as a white vertical line, the β-sheet backbones are drawn in blue lines (the VL domain sheet in front, the VH domain sheet behind and the six hypervariable loops (heavy-chain loops H1, residues 28–32; H2, residues 50–58; H3, residues 102–109; light-chain loops L1, residues 26–37; L2, residues 56–61; L3, residues 97–102; consecutive numbering system) are in red and green. The three shorter loops, H1, L2 and L3, drawn in red, have lower average axis coordinates, so they are closer to the centre of the barrel than the other three loops, H2, H3 and L1, drawn in green. In our construction protocol, the low loops were constructed before the high loops.

residues; the longest, L1, contained 12 (see Fig. 1 legend for loop nomenclature).

The CONGEN search algorithm[7] uses a selectable angular grid to sample the torsional degrees of freedom in a given polypeptide segment, using the amino- and carboxy-terminals as fixed endpoints. In the searches, both *cis* and *trans* proline peptide bonds are considered. Each sampling run generates a set of loop conformations which satisfy the fixed endpoint condition (the modified Gō and Scheraga chain-closure algorithm)[18,19] and have acceptable potential energies[20]. These energies are used to distinguish between 'good' and 'bad' conformations and depend critically on short-range interactions between loops and, consequently, on the order of loop construction. Our order of loop reconstruction L2-H1-L3-H2-H3-L1 was chosen on the basis of the relative positions of the loops in the local frame of reference provided by the β-sheets of the VL-VH domain interface[21,22] (Fig. 1). The tight non-covalent interaction of these sheets creates an elliptical 'β-barrel' whose long axis nearly coincides with the two fold axis of VL–VH domain pseudosymmetry. In this reference frame, the three shorter loops (L2, L3 and H1) are placed lower (closer to the midpoint of the β-barrel interface) than the others. They do not interact with each other and provide a natural basis for the construction of the remaining 'high' loops, H2, H3 and L1. These 'low' loops were built first, H2 next, and loops H3 and L1, which interact closely, were built last.

From the set of conformations generated for each loop, 'the best one' was selected to be incorporated into the final model

Table 1 Description of McPC 603 combining site reconstruction

Loop	Run and conformation number	Total conformations found	Energy (kcal)	Root mean square (Å) Total	Root mean square (Å) Backbone	Surface (Å^2)
L2	56LMCP7-75*	159	−14.6	1.9	1.6	360
	56LMCP7-89		−16.1	2.4	2.0	397
H1	28HMCP1-44*	46	−17.1	1.7	0.7	344
	28HMCP1-43		−10.7	1.6	0.7	329
L3	97LMCP7-42*	306	−20.8	1.4	0.8	320
	97LMCP7-95		−21.8	3.0	1.5	375
H2	50HMCP5-81	576	−18.8	2.1	2.0	101
	50HMCP5-75		−18.5	2.2	2.2	126
	50HMCP5-71		−18.2	2.1	2.1	100
	50HMCP5-76		−17.9	2.2	2.1	123
	50HMCP5-170*		−17.8	1.6	1.0	80
	50HMCP5-175		−17.8	1.7	1.0	96
	50HMCP27-681*	912	−39.6	2.4	1.9	550
	50HMCP27-874		−33.8	4.0	1.5	584
H3	102HMCP17-176*	511	−64.5	2.9	1.1	369
	102HMCP17-408		−62.2	2.7	0.9	394
L1	26LMCP23-1344*	2,287	−49.0	4.0	3.5	799
	26LMCP23-1348		−46.9	4.4	3.8	810

Six polypeptide chain segments (see Fig. 1) were deleted from the McPC 603 structure and were reconstructed in separate computer runs, in the order listed in the table. In the L2 loop, residues 61–59 were searched in the C→N direction and the Gō-Scheraga chain-closing algorithm[7,19] was applied to residues 56–58. In the H1 loop residues 28 and 32 were searched and the chain-closing was applied to residues 29–31. In the L3 loop, residues 97–99 were searched successively and chain-closing applied to residues 100–102. In the H3 loop the residues were searched successively from the amino-terminus and the chain-closing routine was applied to residues 107–109. The H2 and L1 loops were too long to be constructed as one segment and 'real space renormalization'[30] was applied. In H2, the N-terminal tripeptide Ala-Ser-Arg was searched first (see the top six conformations listed in the table). From the six lowest-energy conformations, all within a 1 kcal interval, the one with the smallest surface was taken as the starting conformation for the search over the rest of the loop (see the lower two conformations listed in the table). In the L1 loop, three separate runs constructed dipeptides Ser-Gln, Ser-Leu and Leu-Asn. In the second and third runs, the four lowest-energy conformations from the previous run were all included as starting conformations. The C-terminal dipeptide Lys-Asn was then constructed separately. Finally, the lowest-energy conformations of the peptides Ser-Glu-Ser-Leu-Leu-Asn (residues 26–31) and Lys-Asn (residues 36–37) were included as starting conformations into the search over the rest of the loop. The angular grid on which the torsional spaces were searched was 30° and the non-bonded interactions were computed within the cut-off distance of 5 Å, with dielectric constant.

Table 2 Description of HyHEL-5 combining site reconstruction

Loop	Run and conformation number	Total conformations found	Energy (kcal)	Root mean square (Å) Total	Root mean square (Å) Backbone	Surface (Å^2)
L2	49LHYHEL55-76*	100	−57.6	1.7	0.8	268
	49LHYHEL55-89		−56.3	2.3	1.3	294
	49LHYHEL55-20		−55.2	1.9	1.1	286
H1	28HHYHEL55-21*	37	−26.1	1.8	1.1	489
	28HHYHEL55-22		−22.9	1.9	1.2	498
L3	90LHYHEL56-21*	33	−20.4	4.1	1.1	357
	90LHYHEL56-18		−5.5	4.4	1.4	368
H2	50HHYHEL56-4*	6	−0.3	3.1	2.1	205
	50HHYHEL56-3		2.5	3.2	2.1	214
H3	100HHYHEL55-11*	14	−47.3	2.7	1.0	359
	100HHYHEL55-2		−46.7	1.5	0.9	381
L1	26LHYHEL55-11	13	−35.1	2.0	0.4	367
	26LHYHEL55-7*		−32.8	1.8	0.6	362

Amino acid sequences of variable domains of McPC 603 and HyHEL-5 light and heavy chains were aligned to maximize sequence homology (43% identical residues in the VH domain and 54% identical residues in the VL domain). Endpoints of HyHEL-5 hypervariable loops were then determined as residue positions coinciding with the endpoints of McPC 603 hypervariable loops determined previously (see Fig. 3 legend). The six HyHEL-5 hypervariable loops defined in this way (32 residues altogether) were deleted from the HyHEL-5 crystallographic structure and reconstructed using the protocol developed for construction of the McPC 603 combining site. The order of loop construction, Gō-Scheraga chain-closure, and so on was as described in Table 1. More specifically, the L2 loop residues 54–52 were searched in the C→N direction and the chain-closing algorithm[7,19] was applied to residues 49–51. In the H1 loop the residues 28 and 32 were searched and chain-closing applied to residues 29–31. In the L3 loop, the residues 90 and 91 were searched successively and chain-closing applied to residues 92–94. In the H2 loop the N-terminal tetrapeptide 50–53 was searched first and the chain-closing was applied to residues 54–56. In the H3 loop residue 100 was searched and chain-closing applied to residues 101–103. In L1, the residues 26 and 30 were searched and chain-closing applied to residues 27–29. Because of the short lengths of the HyHEL-5 loops H2 and L1, whole loops could be constructed in single search runs. The angular grid on which the torsional spaces were searched was 30° and the non-bonded interactions were computed using cut-off distance of 8 Å, with the distance-dependent dielectric constant used for all six loops. The lowest-energy conformations of each run are given; * indicates those included in the model displayed in Fig. 3.

of McPC 603 as follows. All the conformations were ranked by their empirical potential energies[20]; the two lowest-energy conformations were examined in detail. If calculated energies of these two structures differed by about 2 kcal (three times the Boltzmann factor kT, indicating that both conformations are well populated at room temperature), the one with the smaller solvent-exposed surface[23] was selected. Otherwise, the lowest-energy structure was selected. The H2 and L1 loops were too long to be constructed in a single computer run. Instead, selected results of searches over shorter sections of the loop were included into new searches over the rest of the loop as starting conformations (see Table 1). The success of our reconstruction was gauged against the original X-ray structure, by evaluating root-mean-square (r.m.s.) deviations between the computed and the crystal conformations.

Details of individual loop constructions in McPC 603 are described in Tables 1 and 3. Overall, the best r.m.s. conformations were not often the lowest energy ones and therefore could not be selected in a true modelling experiment where the final structure would be unknown. Nevertheless, there were low-energy conformations generated which still matched the native structure fairly well. The r.m.s. agreement of our model, constructed from the few selected lowest-energy conformations, is 2.4 Å total and 1.7 Å backbone r.m.s. shifts from the X-ray structure (Fig. 2). The model shows close agreement with the crystal; for example, the rare *cis* peptide bond conformation of Pro L101 was accurately reproduced. Two isolated regions show large disagreements with the crystal structure: the tyrosine side chain ring H103, which had been misplaced by rotation of about 160° around its χ_1 torsional angle (Fig. 2*d*) and the backbone of the middle of the L1 loop, residues 32–35 (Fig. 2*b*).

Each of the six McPC 603 loops was very closely matched by at least one of the computed conformations. If these 'best' r.m.s. constructions could be assembled into a model, its shift from the X-ray structure would be 1.7 Å, 1.3 Å on the backbone only. This level of accuracy approaches the limits of atomic resolution of the McPC 603 structure[17] (0.5–1.0 Å).

The reconstruction protocol developed for McPC 603 hypervariable loops was next applied to another antibody structure, as described for the McPC 603 loops; the only change to the above construction protocol was a consistent use of the distance-dependent dielectric constant, $\varepsilon = R$ (R is distance between the two interacting charges), in the evaluation of electrostatic interactions. As for McPC 603, each of the hypervariable loops was matched closely by at least one calculated conformation. The model of the HyHEL-5 combining site, made from selected lowest-energy conformations showed a 2.6 Å r.m.s. shift from the X-ray structure, 1.4 Å r.m.s. from the backbone. Thus, our construction scheme developed on the McPC 603 molecule could be used, with only small technical modifications, on a different molecule to give an equally good result.

Figure 3 shows details of our HyHEL-5 model. Loop backbones show good agreement with the X-ray structure, except for the H2 loop residues H51–H54. Most of the side chains match the X-ray structure equally well, the exceptions being Tyr H101 and Arg L92. The poor agreement in the H2 modelling is due to inaccurate positioning of the side chain Phe H29 during the H1 loop construction. The shift in the Phe ring position,

Table 3 CONGEN reconstruction of combining sites in antibodies McPC 603 and HyHEL-5

Loop	Length	Root mean square to crystal (Å) total	Root mean square to crystal (Å) backbone	CPU time*
H1 in McPC 603	5	1.7	0.7	4 hours
H1 in HyHEL-5	5	1.8	1.1	3 hours
H2 in McPC 603	9	2.1	1.6	5 days
H2 in HyHEL-5	7	3.1	2.1	20 minutes
H3 in McPC 603	8	2.9	1.1	7 days
H3 in HyHEL-5	4	2.7	1.0	2 hours
L1 in McPC 603	12	3.0	2.6	7 days
L1 in HyHEL-5	5	1.8	0.6	40 minutes
L2 in McPC 603	6	1.9	1.6	8 hours
L2 in HyHEL-5	6	1.7	0.8	37 hours
L3 in McPC 603	6	1.4	0.8	5 hours
L3 in HyHEL-5	5	4.1	1.1	12 hours
All loops, McPC 603	46	2.4	1.7	20 days
All loops, HyHEL-5	32	2.6	1.4	2.25 days

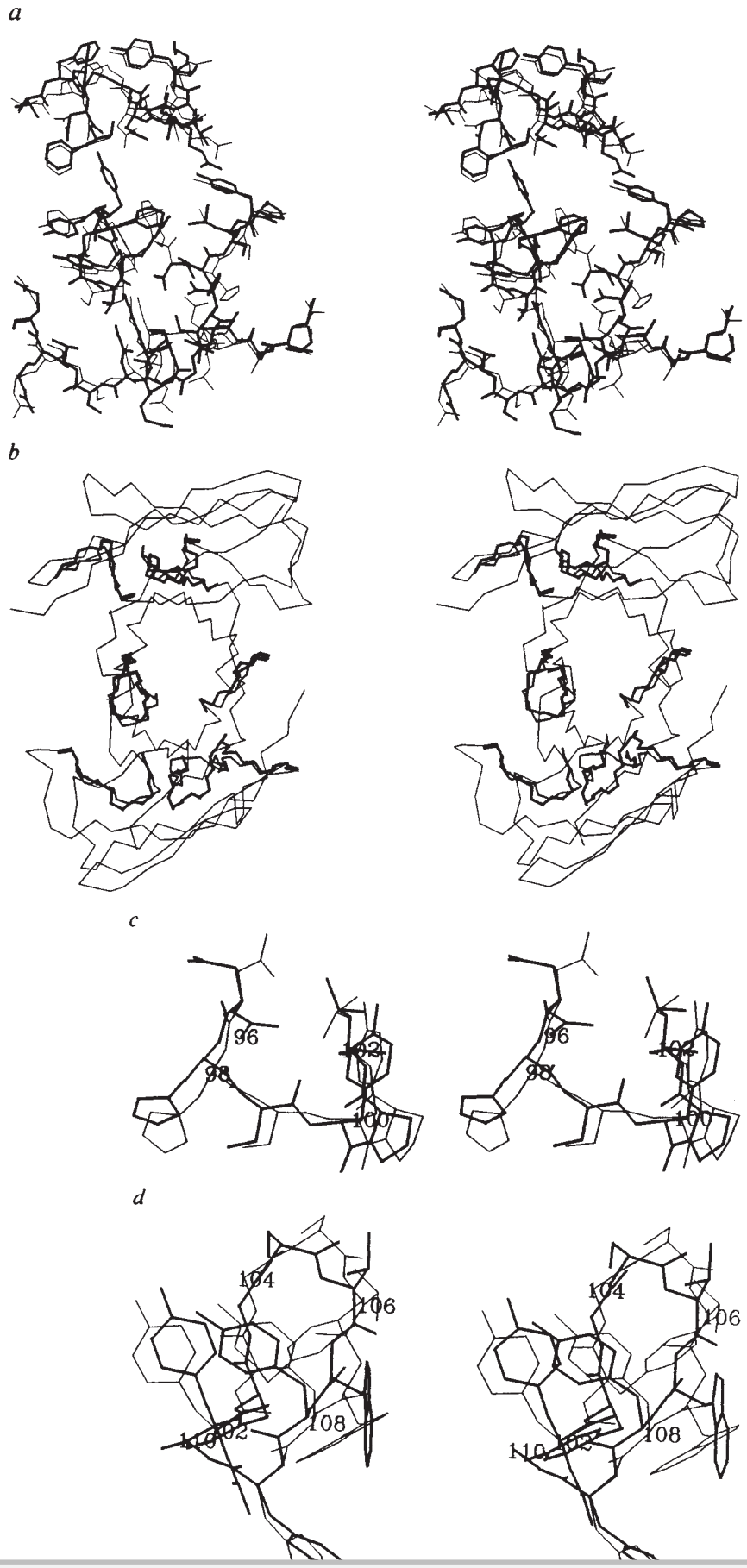

Fig. 2 Stereoscopic view of the McPC 603 antigen combining site, as reconstructed by CONGEN (see Table 1). Light lines trace the crystallographic structure, heavy lines are CONGEN-generated conformations. *a*, Comparison of all six hypervariable loops in the X-ray structure and those constructed by CONGEN. The loops are clockwise from the top: H2, H-chain residues 50–58 (Ala-Ser-Arg-Asn-Lys-Gly-Asn-Lys-Tyr); L3, L-chain residues 97–102 (Asp-His-Ser-Tyr-Pro-Leu); L1, L-chain residues 26–37 (Ser-Gln-Ser-Leu-Leu-Asn-Ser-Gly-Asn-Gln-Lys-Asn); L2, L-chain residues 56–61 (Gly-Ala-Ser-Thr-Arg-Glu); H3, H-chain residues 102–109 (Tyr-Tyr-Gly-Ser-Thr-Trp-Tyr-Phe); and H1, H-chain residues 28–32 (Thr-Phe-Ser-Asp-Phe). *b*, Backbones of the framework residues outside the hypervariable loops are traced through α-carbons only, hypervariable loop backbones are drawn as N, Cα and C atoms. Orientation of the combining site was as above. The middle part of the L1 loop, where there is the greatest discrepancy between the crystallographic and calculated structures, is poorly visible and disordered in the crystal[17]. *c*, A detailed picture of the L3 loop (total r.m.s. shift between the crystallographic and calculated conformations 1.4 Å, backbone shift 0.8 Å). Note the *cis* proline residue L101. *d*, A detailed picture of the H3 loop (total r.m.s. shift between the crystallographic and calculated conformations 2.9 Å, backbone shift 1.1 Å). Note the displacements of the side-chain ring Tyr H103 and, to a lesser extent, Trp H107. These inaccuracies are largely responsible for the unsatisfactory r.m.s. value.

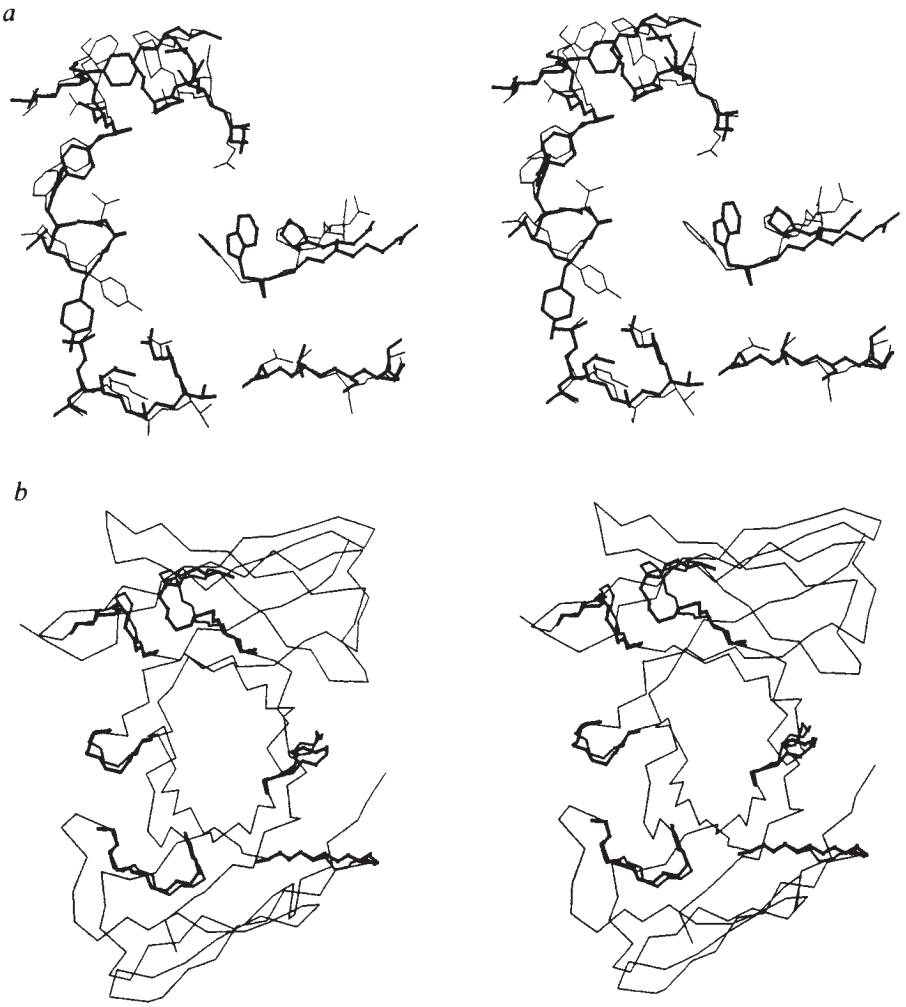

Fig. 3 Stereoscopic view of the HyHEL-5 antigen combining site, as reconstructed by CONGEN (see Table 2). Light lines trace the crystallographic structure, heavy lines are CONGEN-generated conformations. *a*, Comparison of all six hypervariable loops in the X-ray structure and as constructed by CONGEN. The loops are, clockwise from the top: H2, H-chain residues 50–56 (Glu-Ile-Leu-Pro-Gly-Ser-Gly); L3, L-chain residues 90–94 (Trp-Gly-Arg-Asn-Pro); L1, L-chain residues 26–30 (Ser-Ser-Ser-Val-Asn); L2, L-chain residues 49–54 (Asp-Thr-Ser-Lys-Leu-Ala); H3, H-chain residues 100–103 (Asp-Tyr-Asp-Phe); and H1, H-chain residues 28–32 (Thr-Phe-Ser-Asp-Tyr). *b*, Backbones of the framework residues outside the hypervariable loops are traced through α-carbons only, hypervariable loop backbones are drawn as N, Cα and C atoms. Orientation of the combining site is as above.

although small, interfered with the correct placement of Pro H54 because of atomic volume overlaps and resulted in misplacement of the middle part of this loop. Discrepancies like these provide us with important stimuli for future work. We expect that improvements in energy ranking (for example, a better representation of solvent-modified electrostatic interactions) will allow selection of better-quality conformations and construction of more accurate models. Empirical free-energy functions[24–27] should be particularly useful. Likewise, canonical structures developed by Chothia and Lesk[13], could be used to improve on the loop selection process.

The conceptual basis for the construction of polypeptide loops on a fixed backbone framework is only beginning to be understood. *A priori*, there was no guarantee that local potential energies of the loops and their local environments would lead to accurate models of the structure. The success of the construction protocol may imply that local interactions in protein structure[28], particularly loops[29], are more important than generally believed.

This work was supported by grants from National Institute of Health and the Office for Naval Research. The program CONGEN is available by request from R.E.B.

Received 7 March; accepted 22 August 1988.

1. Wu, T. T. & Kabat, E. A. *J. exp. Med* **132,** 211–250 (1970).
2. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. *Nature* **321,** 522–525 (1986).
3. Richardson, J. *Adv. Protein Chem.* **34,** 167–339 (1981).
4. Sibanda, B. L. & Thornton, J. *Nature* **316,** 170–174 (1985).
5. Rose, G. D., Gierasch, L. M. & Smith, J. A. *Adv. Protein Chem.* **37,** 1–50 (1985).
6. Leszczynski, J. F. & Rose, G. D. *Science* **234,** 849–855 (1986).
7. Bruccoleri, R. E. & Karplus, M. *Biopolymers* **26,** 137–168 (1987).
8. Kabat, E. A. & Wu, T. T. *Proc. natn. Acad. Sci. U.S.A.* **69,** 960–964 (1972).
9. Chothia, C. *et al. Science* **233,** 755–758 (1986).
10. de la Paz, P., Sutton, B. J., Darsley, M. J. & Rees, A. R. *EMBO J.* **5,** 415–425 (1986).
11. Fine, R. M., Wang, H., Shenkin, P. S., Yarmush, D. L. & Levinthal, C. *Proteins* **1,** 342–362 (1986).
12. Smith-Gill, S. J. *et al. J. molec. Biol.* **194,** 713–724 (1987).
13. Chothia, C. & Lesk, A. M. *J. molec. Biol.* **196,** 901–917 (1987).
14. Shih, H. L., Brady, J. & Karplus, M. *Proc. natn. Acad. Sci. U.S.A.* **82,** 1697–1700 (1985).
15. Snow, M. E. & Amzel, M. L. *Proteins* **1,** 276–279 (1986).
16. Moult, J. & James, M. N. G. *Proteins* **1,** 146–163 (1986).
17. Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. *J. molec. Biol.* **190,** 593–604 (1986).
18. Gō, N. & Scheraga, H. A. *Macromolecules* **3,** 178–187 (1970).
19. Bruccoleri, R. E. & Karplus, M. *Macromolecules* **18,** 2767–2773 (1985).
20. Brooks, B. *et al. J. comput. Chem.* **4,** 187–217 (1983).
21. Chothia, C., Novotny, J., Bruccoleri, R. E. & Karplus, M. *J. molec. Biol.* **186,** 651–663 (1985).
22. Novotny, J. & Haber, E. *Proc. natn. Acad. Sci. U.S.A.* **82,** 4592–4596 (1985).
23. Lee, B. K. & Richards, F. M. *J. molec. Biol.* **55,** 379–400 (1971).
24. Rashin, A. A. *Biopolymers* **23,** 1605–1620 (1984).
25. Novotny, J., Bruccoleri, E. R. & Karplus, M. *J. molec. Biol.* **177,** 787–818 (1984).
26. Eisenberg, D. & McLachlan, A. *Nature* **319,** 199–203 (1986).
27. Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H. A. *Proc. natn. Acad. Sci. U.S.A.* **84,** 3086–3090 (1987).
28. Chou, P. Y. & Fasman, G. *Biochemistry* **16,** 222–244 (1974).
29. Dyson, H. J. *et al. Nature* **318,** 480–483 (1985).
30. Pincus, M. R., Klausner, R. D. & Scheraga, H. A. *Proc. natn. Acad. Sci. U.S.A.* **79,** 5107–5110 (1982).