PROTEINS: Structure, Function, and Genetics 14:483–498 (1992)

Modeling the Anti-CEA Antibody Combining Site by Homology and Conformational Search

Maria T. Mas,[1] Kenneth C. Smith,[2] David L. Yarmush,[2] Kazuo Aisaka,[1] and Richard M. Fine[2]
[1]Physical Biochemistry Section, Division of Biology, Beckman Research Institute of the City of Hope, Duarte, California 91010, and [2]Department of Biology, Columbia University, New York, New York 10027

ABSTRACT **A model for an antibody specific for the carcinoembryonic antigen (CEA) has been constructed using a method which combines the concept of canonical structures with conformational search. A conformational search technique is introduced which couples random generation of backbone loop conformations to a simulated annealing method for assigning side chain conformations. This technique was used both to verify conformations selected from the set of known canonical structures and to explore conformations available to the H3 loop in CEA ab initio. Canonical structures are not available for H3 due to its variability in length, sequence, and observed conformation in known antibody structures. Analysis of the results of conformational search resulted in three equally probable conformations for H3 loop in CEA. Force field energies, solvation free energies, exposure of charged residues and burial of hydrophobic residues, and packing of hydrophobic residues at the base of the loop were used as selection criteria. The existence of three equally plausible structures may reflect the high degree of flexibility expected for an exposed loop of this length. The nature of the combining site and features which could be important to interaction with antigen are discussed.**

Key words: CEA, antibody, CDR, homology modeling, canonical structures, rotamers, tweak, conformational search

INTRODUCTION

Antibody molecules comprise a fundamentally important class of biological molecules whose structure, function, and genetics have been widely studied.[1,2] Several hundred sequences for light and heavy chains of antibody molecules have been reported[3] while a smaller but significant number (~20) of three-dimensional structures have been determined crystallographically.[4–19] Comparison of available sequences and crystal structures confirms the early conjecture of Kabat and Wu that both the structural and functional diversity of antibody molecules arise from sequence diversity in only ~10% of the molecule. This diversity is concentrated in six surface hypervariable loops, three from the light chain and three from the heavy chain, which assemble during folding to form the antibody combining site. Despite considerable diversity among the hypervariable loops in length, sequence, and conformation, and despite the existence of distinct classes of sequences for the remainder of the molecule, the structure of this framework is remarkably well conserved across class boundaries and in diverse organisms.

These properties of antibodies make them a natural target for two kinds of studies. The first of these is theoretical modeling aimed toward predicting the nature of the combining site for sequences whose structure remains undetermined. The second is experimental mutagenesis aimed toward altering the nature of the combining site and its concomitant binding and specificity. This paper reports the construction of a model of a monoclonal antibody specific for the carcinoembryonic antigen (CEA), which is a known colon cancer cell marker.[20] The purpose of the model building is to (1) characterize the nature of the antibody combining sites; (2) identify key residues which play a role in antigen binding and are thus targets for mutagenesis; and (3) help in the design and interpretation of genetic engineering experiments whose aim is to modify the antibody specificity and affinity. A later paper[21] will present the results and discussion of initial site directed mutagenesis experiments in the context of the current model. The approach used in the present paper to

Abbreviations used: CEA, monoclonal antibody CEA.66-E3, specific for carcinoembryonic antigen; CDR, complementarity-determining region; EM, Eisenberg–McLachlan solvation free energy.

Received May 17, 1991; revision accepted May 5, 1992.
Address reprint requests to R.M. Fine (present address), Biosym Technologies Inc., 4 Century Dr., Parsippany, NJ 07054.
Present address of Kenneth C. Smith: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032.
Present address of Kazuo Aisaka: Kyowa Hakko Kogyo Company, Ltd., Tokyo Research Laboratories, 3-6-6 Asahimachi, Machidashi, Tokyo, Japan.

construct the model of the CEA antibody combines knowledge based modeling and energetic conformational search methods.

BACKGROUND

Several approaches to modeling hypervariable loops in proteins have been reported in the literature. Early attempts were based on crystallographic hand-building methods, using interactive modeling tools to adjust the internal coordinates for loop amino acids to construct an initial loop model.[22–25] This conformation was refined (in some cases) using force field energy minimization. The difficulty with such an approach is one which is now widely recognized: the energy surface topography for a protein exhibits many local minima. These tend to trap energy minimized loops near their hand-built starting conformation. The minimized conformation is thus strongly dependent on the intuition of the modeler in constructing a "nearly correct" starting conformation by hand. Intuition on conformations of proteins or of loops has historically proven elusive and difficult to refine.

More recent approaches have considered multiple conformations for loops in proteins. Methods fall into two general classes.

1. In *knowledge-based modeling* the collection of solved crystal structures in the Protein Database[26] is used to suggest possible conformations for the hypervariable loop of interest. The choice is made based on similarity in length, sequence, and if the loop conformation chosen is not that of a hypervariable loop in a known crystal structure, similarity in backbone geometry at the ends of the loops. Protein Database searches for structures with similar backbone geometry at the end of the loops are easily accomplished using a distance matrix matching method introduced by Jones.[27]

The knowledge-based approach has been refined specifically for antibodies by Lesk and Chothia[28,29] who found that hypervariable loops in six known crystal structures adopted only a few distinct conformations named canonical structures. Key residues and contacts were suggested as determinants which select among the ensemble of canonical structure conformations. Their analysis was sufficient to successfully predict the conformations of selected hypervariable loops in four known antibodies prior to crystallographic determination.

The key limitation to any knowledge based modeling approach is the limited size of the Protein Databank. Simply stated, there is no guarantee that the database is large enough to contain an example for a given structure. Evidence has been presented to suggest that the Protein Databank is sufficient to contain multiple examples of all possible backbone conformations of six adjacent amino acids in proteins.[30] This sufficiency, however, breaks down for longer stretches (R. Fine and S. Bryant, unpublished). For classifications such as that of Lesk and Chothia, the number of available structures of antibody loops is particularly small. Their canonical structure hypothesis can be experimentally tested by site-directed mutagenesis. In a recent paper, Strong et al.[19] have found that the conformation of a hypervariable loop can be changed by changing the identity of a residue not considered a canonical determinant by Lesk and Chothia which participates in interhypervariable-loop contacts. Whereas these observations clearly illustrate limits to the canonical structure hypothesis in its present form, there is unquestionable value in the observed correlation between conserved structure and conserved contacts which should be considered in modeling antibody combining sites.

2. In *conformational search* the internal coordinates describing the conformations of the hypervariable loops are varied to explore all possible conformations. Selection is made based on energetic criteria which ideally include a description of solvation free energy. Methods have been developed based on systematic search in internal coordinates,[31,32] random generation of backbone conformations,[33,34] Monte Carlo,[35] and high temperature molecular dynamics.[36] While these methods are not limited by the presence of a relevant template in the Protein Database, they are limited by the speed of present day computers since their computational cost is high.

An ideal approach to modeling antibody structures by homology would be one which incorporates the best features from both knowledge-based modeling and conformational search. An interesting combined approach has been described by Martin et al.[37] Their method depends on the length of the loop under consideration. For short loops ($N \leq 5$ where N is the number of amino acids) the conformation is determined using an algorithm of Bruccoleri and Karplus and developed in a program called CONGEN.[31] This algorithm constructs possible conformations using systematic search in N-6 backbone torsions, where N is the total number of backbone torsional angles which are varied. The backbone geometry at the ends of the loop required to correctly splice into the remaining framework provide sufficient constraints to analytically solve for the remaining six torsions.[38] For intermediate lengths ($6 \leq N \leq 7$), the database of solved crystal structures is searched for loop conformations consistent with the backbone geometry at the ends of the loop, using the A. Jones database search method described earlier.[27] For longer lengths a combination of methods is used. First the method of Jones is used to identify candidates from the Protein Database which exhibit correct geometry at the ends of the loop. This is followed by deletion of the midsection of each loop, and finally reconstruction of this midsection using the torsional search method of Bruccoleri and

Karplus.[31] This approach provides a clear recipe which is systematic in its preference for conformational search. However, the choice of conformational search methods for short loops seems somewhat arbitrary. The geometries at the ends of the loops provide severe constraints on intervening torsions; as a result, knowledge-based and conformational search-based approaches will work equally well for short loops (exceptions to this rule may occur for unusual combinations of glycines and prolines, where conformational search may be required). For intermediate lengths ($6 \leq N \leq 7$) conformational search becomes expensive, but has demonstrated the ability to saturate the conformational space available to the loop.[34] The data in reference 30 suggest that the available crystal structures provide a complete set of backbone geometries in this length regime (excluding unusual combinations of glycines and prolines). For longer lengths, neither approach can be expected to work well. The combination of methods used by Martin et al.[37] assumes that the bases of long loops will be well-represented in the collection of solved structures. This assumption, imposed from necessity since the systematic search for long loops using the CONGEN search method is prohibitive, is both interesting and provocative but open to question.

In modeling the CEA antibody, we have used a combined approach which is quite different from that of Martin et al.[37] Our approach relies heavily on the canonical structure hypothesis of Lesk and Chothia.[28,29] We have augmented this knowledge-based approach with conformational search specifically for two purposes: (1) to check the conclusions from application of the canonical structure hypothesis for selected loops; and (2) to investigate the conformations of the single long loop H3 in CEA for which there is no prediction from the canonical structure hypothesis. The conformational search technique used is the TWEAK algorithm described elsewhere[33,34] for generation of a random ensemble of backbone conformations, coupled to side chain modeling methods described in this paper. The advantage of the conformational search algorithms chosen is that they generate an ensemble of independent conformations for long loops quickly. Whereas this ensemble may not be sufficient to span the conformational space available to the entire loop, it will span the conformational space available to residues near the base of the loop. Thus questions of the sufficiency of the database to describe the base conformations of these loops can be circumvented. Our approach to H3 was motivated by an examination of the conformations of H3 loops in experimentally determined antibody structures in the Protein Databank. The conformation of H3 is dictated in large part by its conformation near its base; this conformation is determined by the packing of large hydrophobic residues which are part of and which flank H3 in sequence. Given the sequence variability of H3 we did not wish to rely on the database to provide a correct base conformation.

Conformations of five of the six hypervariable loops were identified based upon canonical structures. Conformations for selected loops were verified using conformational search. The unusual length (12 amino acids) and composition of H3 in CEA with two consecutive prolines near its base and several hydrophobic residues flanking an Asp at the center of the loop are unique for this antibody. A preliminary study using canonical structures alone found difficulty in identifying conformations for this loop which simultaneously exposed the Asp and minimized the exposure of hydrophobic surface area (K. Aisaka and M. Mas, unpublished). These difficulties dictated the need to employ extensive conformational search for this loop. H3 has not been analyzed as extensively as other loops in the literature, due in part to structural variability in known crystal structures and in part to prohibitive lengths for conformational search.

MATERIALS AND METHODS

Calculations were carried out on a STAR ST100 array processor linked to a MicroVax II in the Biological Computer Facility of Columbia University, on a Convex C220 in the Department of Biochemistry and Molecular Biophysics at Columbia University, and a Silicon Graphics Personal Iris workstation in the Biosym Corporation. Several software packages were used for the present work. PAKGGRAF,[39] a torsional modeling program due to Levinthal and co-workers, was modified to incorporate random loop generation and side chain modeling using rotamer libraries as described below. CHARMM[40] coupled to GEMM[41] was used for cartesian minimizations and energy evaluation of H3 on the STAR ST100. Minimizations of shorter loops, L1 and H2, were performed using X-PLOR[42] on the Convex C220. The force field used in PAKGGRAF used nonbond parameters taken from Hagler et al.[43] and torsional parameters from ECCEP.[44] The force field used in CHARMM/GEMM was PARAM19;[40] that used in X-PLOR was essentially identical for the problems dealt with in this work. The reason for using two disparate force fields is that a force field optimized for cartesian molecular mechanics will differ from one optimized for torsional molecular mechanics, particularly in the description of interactions between atoms separated by three bonds (1–4 interactions).

All database loop searches were performed using the graphics display program INSIGHT (Biosym). Sequence and structural alignments within the antibody family were performed using the HOMOLOGY package (Biosym). Solvent accessibilities were computed using a program provided by Dr. Fred Richards at Yale. Eisenberg-McLachlan[45] solvation free

TABLE I. Sequences and Numbering for Hypervariable Loops of CEA, Aligned With Selected Backbone Templates From the Protein Database Used in Homology Modeling*

L1	26	27	28	29	30	31	32						
CEA	Ser	Gln	Asp	Val	Gly	Ala	Ala						
REI	Ser	Gln	Asp	Ile	Ile	Lys	Val						
L2	50	51	52	53	54	55	56						
CEA	Trp	Ala	Ser	Thr	Arg	His	Thr						
MCPC603	Gly	Ala	Ser	Thr	Arg	Glu	Ser						
L3	91	92	93	94	95	96							
CEA	Tyr	Ser	Gly	Tyr	Pro	Leu							
REI	Tyr	Glu	Ser	Leu	Pro	Tyr							
H1	26	27	28	29	30	31	32						
CEA	Gly	Phe	Thr	Phe	Ser	Arg	Tyr						
MCPC603	Gly	Phe	Thr	Phe	Ser	Asp	Phe						
H2	52a	53	54	55									
CEA	Ser	Gly	Gly	Ser									
KOL	Asp	Asp	Gly	Ser									
H3	95	96	97	98	99	100	a	b	c	i	j	k	101
CEA	Pro	Pro	Leu	Ile	Ser	Leu	Val	Ala	Asp	Tyr	Ala	Met.	Asp

*Numbering is that of Kabat et al.[3]

energies were evaluated from these solvent accessibilities using a program developed for that purpose. Cavities were identified using a program provided by Alex Rashin.

All protein coordinates were obtained from the Brookhaven Protein Database. The sequence of the CEA antibody was as published[3,46] with corrections reported in a later paper.[21]

Modeling Strategy

Our overall modeling strategy was to assign loop conformations in order of increasing difficulty. In determining this order, consideration of loop length, sequence homology, degree of expected interaction with other CDRs and framework, and the degree to which the sequence could be assigned to one of the canonical structures were considered. The first loop modeled was L2. The conformations of all L2 loops of antibodies in the Protein Database are similar; only one canonical structure is given by Lesk and Chothia. Conformational search has shown that this loop is determined by end vector geometry and backbone contacts which are largely independent of sequence.[34] The next four loops modeled in order were H1, H2, L3, and L1. Each of these was assigned backbone conformations using the canonical structure approach of Lesk and Chothia.[28,29] Side chain conformations were assigned as described below. H3 was modeled using extensive conformational search. The sequences of hypervariable loops of CEA and of the selected templates are shown in Table I. Amino acids are numbered according to Kabat et al.[3]

The framework for the CEA structure was chosen as that of MCPC603 based on sequence similarity. In making this choice particular attention was paid to the framework residues in contact with hypervariable loops. Since energetic criteria were to be used, the structure of MCPC603 was subjected to convergent minimization using the program CHARMM/GEMM with a distance dependent dielectric. The movement of the hypervariable loops and the framework during these minimizations has been discussed elsewhere.[34]

Procedures were introduced into the interactive modeling program PAKGGRAF specifically for modeling loops. These could be invoked by issuing specific commands. Two types of commands and underlying procedures were introduced. The first of these generated an ensemble of random backbone conformations for a loop of interest. The second generated side chain conformations consistent with a given backbone conformation.

The first command, GENLOOP, initiated the TWEAK procedure described earlier[33,34] to generate an ensemble of random conformations for the backbone of a loop. The TWEAK algorithm begins by assigning a set of random values to the ϕ and ψ angles in the loop. Then, distances which describe the geometry needed for the two ends of the loop to correctly splice into the remainder of the molecule are used to construct a set of Lagrange multipliers equations. These are linearized in the backbone torsions describing the loop (θs) and solved interatively in the course of minimizing the function $\Sigma\Delta\theta^2$. $\Delta\theta$ measures deviation of a ϕ or ψ from its original random value; hence, the initial random torsion angle assignments are minimally disturbed. As loop conformations are generated, they are screened for cor-

rect chirality at the base. They are optionally screened for overlaps with selected portions of the molecule using energy or van der Waals contact criteria. The conformations which pass these screens are written to disk as a "library" of possible loop conformations. The TWEAK algorithm is fast since each iteration requires only the inversion of a 4×4 matrix (a convergent result is typically obtained after 5–20 iterations). For the loop modeling described in this paper, conformations were accepted if the van der Waals energy of interaction of backbone and C_α atoms for the loop residues with framework atoms and with atoms of loops already modelled was less than 1×10^5 kcal/mol. A group-based cutoff of 6 Å was used in these energy evaluations to restrict energy evaluations to neighboring residues.

Side chains were added to generated loop conformations using a set of procedures introduced for that purpose. These procedures were based on the concept of a rotamer library. The rotamer library used was that of Ponder and Richards.[47] This library contains a list of the most likely conformations for side chain χ angles as determined by surveying highly resolved crystal structures. All χ angles are not well determined by such a survey. The Ponder and Richards library typically specifies only χ_1 χ_2. The remaining χ angles were treated either as all *trans* or by generating explicit conformations which treated each chi angle as a 2- or 3-fold rotor. Though the latter is a more complete description of possible conformations, the former was found to be nearly as effective in replicating known structures when used in the procedures described below.

The CUSTOMIZE command initiated a procedure described as follows. For each amino acid to be modeled in the loop each rotamer was selected from the library and subjected to torsional energy minimization in place in the molecule to relieve overlaps. The backbone atoms of the loop and all atoms of the remainder of the protein were considered in computing the interaction energy minimized. A group-based cutoff of 6 Å was used in the minimizations.

Since the force field used in PAKGGRAF contained torsional barriers, the minimizations of amino acid rotamers generally converged to yield well-determined conformations. The minimized rotamers for a particular amino acid were compared to each other and duplications eliminated. Conformations were considered duplicates if the root mean square (rms) for χ angles matched within 10° (after considerations of stereochemical symmetry). The remaining set of minimized rotamers defines "customized" rotamer library for that particular amino acid in that particular location in the protein. Once a customized set of rotamers had been constructed, one of the three procedures could be invoked to search for low energy or high probability combinations of rotamers.

The GRINDSIDE command initiated an exhaustive, systematic search for the lowest energy combination of customized rotamers. This was done by considering all possible combinations of customized rotamers for loop amino acids (plus any contacting amino acids identified as important to determination of the correct loop rotamers). The number of possibilities, however, becomes prohibitive for problems exceeding a few amino acids. For longer loops two other procedures were installed.

The ANNEALSIDE command initiated a simulated annealing Monte Carlo side chain placement procedure. This procedure began by placing a randomly chosen customized rotamer for each amino acid in the loop (as above). An amino acid in the loop was selected at random and a rotamer was selected also at random from the customized rotamer library. The energy for this rotamer was compared to that for the rotamer which it replaced. If the new energy was lower the new rotamer was retained. If the new energy was higher a Bolzmann factor was computed using a specified temperature. The new rotamer was accepted or rejected according to the Metropolis Monte Carlo scheme.[48] The temperature, initially set to a high value ($kT = 1{,}000$ kcal/mol), was lowered by a selectable fixed fraction every complete pass through all amino acids in the loop. This fraction was set to a few percent for the work described in this paper. The procedure was stopped when no rotamer replacements occurred on several successive passes. This simulated annealing algorithm was repeated 100–250 times for each loop conformation. Subsequent passes of annealing were quicker since information on acceptance vs. temperature could be retained from the first pass. Thus they were started with a temperature corresponding to a specified acceptance in the initial annealing (here, 90%). The 10 lowest energy combinations of rotamers were written to disk for interactive viewing using the display program INSIGHT. The lowest energy conformation was automatically selected as the preferred selection; however, this selection could be replaced interactively with any of the low energy conformations retained on disk after visual review. The simulated annealing software was modeled on an algorithm written by Peter Shenkin during his stay in the Columbia Biology Computer Facility.

The NOVOSIDE command invoked the quickest and most flexible rotamer assignment procedure. After initialization, the first amino acid in the loop was tested to see if it was "bad." If that amino acid was determined to be "bad," it was replaced with the "best" replacement from the library. This procedure was repeated on the second amino acid in the loop and so on until all amino acids had been considered. If any changes occurred, a new pass through all of the amino acids was initiated until an entire pass exhibited no further changes.

Three aspects of the NOVOSIDE procedure were selectable: (1) the method for initializing rotamer

assignments; (2) the method for defining "bad" rotamers as candidates for replacement; and (3) the method for selecting the "best" replacement from the customized library. Loop amino acids could be initialized with the lowest energy customized rotamer (this energy excluded interactions with all other side chain atoms in the loop), the highest probability customized rotamer (defined by frequency of occurrence in the Protein Database), or simply the current state of the molecule. The definition of "bad" was supplied by providing a maximum energy or van der Waals overlap (now including interactions with other loop side chains): a large negative energy selected all amino acids as replaceable. The "best" replacement was normally defined as the lowest energy rotamer in the customized rotamer library (including interaction with other loop side chains). An alternate procedure selected the highest probability rotamer which passes a specified energy or overlap cutoff as "best."

The three side chain search procedures satisfy different needs in modeling side chains using rotamer libraries. GRINDSIDE is guaranteed to find the lowest energy combination of rotamers (the "global" minimum for the problem) but is restricted to a few amino acids. ANNEALSIDE uses the well-known simulated annealing technique from random starts to find representative low energy combinations for large problems in finite computer time. It will approach the global minimum as continued computer time is invested. NOVOSIDE is designed to quickly find a reasonable local minimum in energy or probability. The options installed permit the user to adjust the amount of bias toward an initialized set of side chain conformations which may be a hand-built model. For modeling performed here, all of these procedures were tried and compared. For consistency in constructing a final model, NOVOSIDE was employed to assign rotamers to randomly generated loops prior to cartesian minimization. NOVOSIDE was initialized with lowest energy rotamers, considered all amino acids during replacement, and selected the lowest (total) energy rotamer from the library. ANNEALSIDE, which is computationally more expensive, was used to explore alternative packings of side chains after cartesian minimization. To prevent errors in evaluating packings from imprecise backbone conformational specification, a "shifted" energy evaluation could be selectively invoked which replaced $F(1/r)$ by $F(1/\sqrt{r^2 + \epsilon^2})$ in the force field (with ϵ adjusted interactively). This has the effect of moving all atoms away from each other, permitting packings which might otherwise be missed due to incorrect backbone geometry. Final selected packings reported here are for $\epsilon = 0$.

It is worth commenting that these side chain search programs were not used for replacement of single, isolated side chains on the framework of CEA, or of side chains identified as participating in conserved contacts between a hypervariable loop and its context. There is considerable evidence that the side chain conformations of amino acid replacements follow the side chain conformations in original structures to the extent that similar atoms occupy similar positions in three-dimensional space.[49] A MUTATE procedure was installed in PAKGGRAF, which follows the conformational preference of similar heavy atoms. The results were examined for unfavorable overlaps or high energies and alternative rotamers selected only in these cases.

After generation of a library of structures followed by placement of side chain rotamers using NOVOSIDE, each library entry was subjected to 500 cycles of conjugate gradient minimization using CHARMM/GEMM on a STAR ST100. Final assignment of side chain conformations were done using the ANNEALSIDE algorithm as described.

APPLICATION TO CEA AND RESULTS

The framework for CEA was first prepared by mutating all residue differences between CEA and MCPC603. The mutations were performed as follows. If a side chain for a residue was isolated, i.e., it could not make contact with other residues requiring replacement, it was mutated using the MUTATE procedure described above. Side chains which were not isolated were replaced using the ANNEALSIDE procedure. An exception was mutation to a proline: these were modeled using GENLOOP to find side chain and backbone conformations for the proline plus two residues immediately preceding and following the proline (both *cis* and *trans* prolines were tried; the lowest energy conformation was selected).

L2

The backbone conformation of L2 is highly conserved,[28] and well determined by end constraints and interactions of the backbone atoms of the loop with surrounding framework.[34] The CEA conformation was taken as that of MCPC603. The side chains were mutated from those of MCPC603, and side chain rotamers selected using ANNEALSIDE. Since extensive conformational search had been performed on MCPC603 using GENLOOP in the past, no further conformational search was performed.

H1

The conformation of H1 has been discussed by Chothia et al. as being highly constrained by a hydrophobic anchor in the center of the loop.[29] In CEA, this anchor is Phe; this residue is identical in MCPC603. The residue identity between CEA and MCPC603 is high (5/7 identities); for these reasons the backbone conformation from MCPC603 was selected. Simple side chain replacement was used in all but Arg-31 (Asp-31 in MCPC603), which exhibited van der Waals contacts on such replacement.

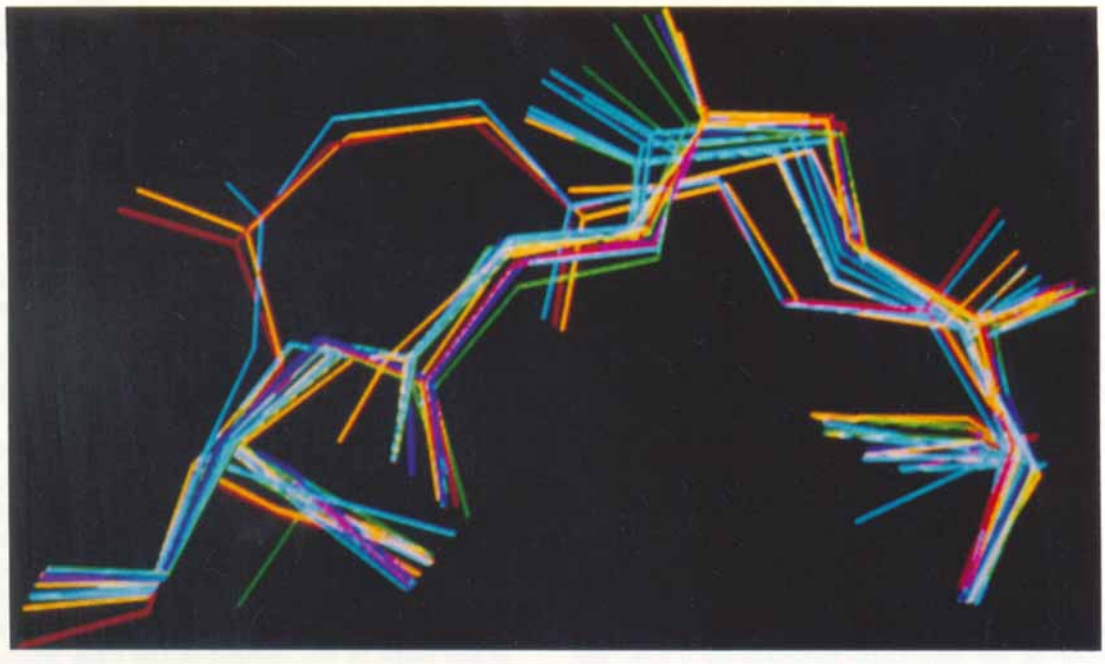

Fig. 1. Results of minimizations from 250 randomly generated conformations of H2. Fifty-five structures were found whose energy was within 10 kcal/mol of the lowest energy found. Representative conformations are shown in the figure. Two classes of conformations are evident. The more populated class (toward the bottom of the figure) closely follows the canonical structure for the loop. The less populated class, consisting of three structures toward the top in the figure, represent an alternate conformation for H2. In these structures the central glycines occupy regions of the Ramachandran map inaccessible to non-Gly residues. The energies of the three top conformations, evaluated using the Param19 force field, were 2, 3, and 9 kcal/mol from the lowest energy found. The backbone torsions for the two central glycines of these conformations were ($\phi1$, $\psi1$ | $\phi2$, $\psi2$) = (−154, −156 | −74, 58), (−139, −148 | −74, 63), and (−160, −150 | −66, 98), respectively.

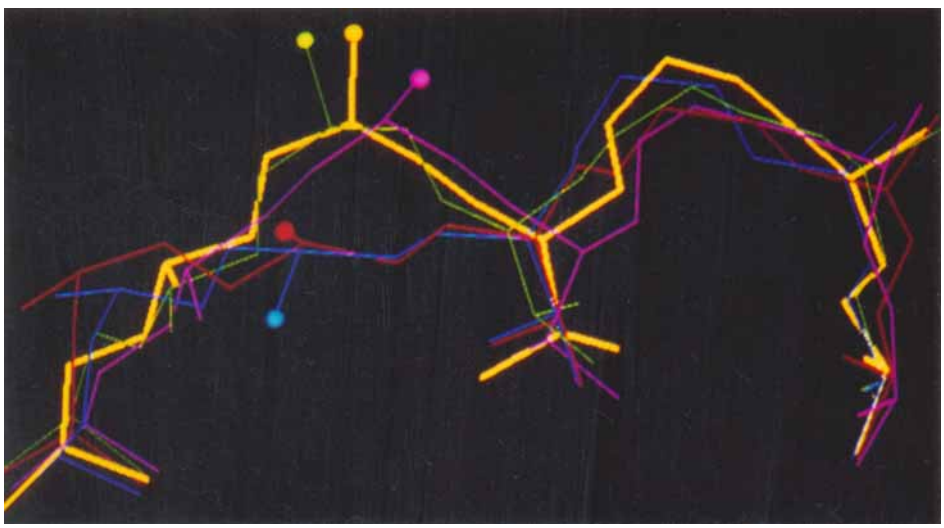

Fig. 2. Results from 250 minimizations from randomly generated conformations of L1. The TWEAK and NOVOSIDE algorithms were used to generate 250 conformations for the entire 7-residue loop. These were then subjected to convergent minimization in CHARMM. The resulting conformations were scanned to select those whose Val-L29 C_β and C_γ coordinates superimposed to within 0.5 Å rms on those of the canonical structure. The selected conformations were screened visually to ensure that conserved contacts with framework residues were maintained. The four resulting loop structures are shown superimposed on the canonical structure (heavy yellow trace). The conformations are similar in the region L30–L32, but differ in the region L26–L29 (to the right in the figure). The C_β of Asp-28 are shown as balls.

This residue was set to the lowest energy rotamer from the Ponder and Richards library.

H2

The sequence for H2 in CEA exhibits the conserved contacts identified by Chothia et al. as determining a specific H2 canonical structure (#3 in ref. 29). Due to the existence of Gly in position 54 in both CEA and KOL, the backbone of the latter was selected as a template for H2. Two substitutions were required from the KOL: Asp→Ser (52a) and Asp→Gly (53). Straightforward side chain replacement using the MUTATE procedure described above was carried out with torsional minimization applied to the resulting Ser χ angles. An identical Ser side chain assignment was found using GRINDSIDE to select the lowest energy minimized rotamer for this residue alone.

The resulting model containing H1 and H2 exhibited a carbonyl–carbonyl overlap. Upon investigation, it was discovered that the backbone conformation of H1 had shifted during the initial minimization of the MCPC603 molecule by 0.6 Å rms (backbone heavy atoms). H2 did not exhibit this overlap against the original crystal conformation of H1; we replaced the H1 conformation with that observed in the original crystal structure which alleviated this problem.

H2 has two glycines in the middle of the loop. There are no crystal structures available for H2 loops with two adjacent glycines. The completeness of the Protein Database in providing all possible conformations of loops of this length with two adjacent glycines in the appropriate positions is open to question. Therefore as a check on the possible effect of these glycines on the final H2 conformation, we have turned to conformational search. GENLOOP and NOVOSIDE were used to generate 250 random structures for H2. These were subjected to convergent minimization as described in Methods. Fifty-five structures whose final force field energy was within 10 kcal/mol of the lowest energy found were collected and examined visually. These structures fell into two classes, shown in Figure 1. The most frequently found class corresponds to the canonical structure. A second less populated class, appearing toward the top in the figure, has members whose glycine residues fall in regions of the Ramachandran map normally inaccessible to non-Gly residues. Relative energies, along with ϕ and ψ values, are given in the figure legend. Our original suspicion that two adjacent glycines could dictate the existence of an alternate conformation for the loop is confirmed by these results. For the final CEA model we have selected the canonical structure, since the frequency with which minimized loop conformations of comparable energy are found should correlate with the size of the "basins of attractions" corresponding to the minima. The size of a given basin of attraction at room temperature correlates with the probability for finding the loop within that region of conformational space. The selection must be made, however, with the caveat that there exists an alternate conformation whose existence is unique to the CEA sequence. It is possible that this alternate may be partially populated in native CEA.

L3

In the analysis of Lesk and Chothia, the L3 of both MCPC603 and REI have the same conformation and, therefore, belong to the same canonical structure. Two positions, 95 (Pro) and 90 (Gln in REI and Asn in MCPC603) were considered critical in this assignment. The CEA sequence shows a Pro at position 95 and an Gln in position 90. L3 conformations from both MCP and REI were spliced into the CEA model. The placement of Gln at position 90 was found to be energetically incompatible with the backbone conformation of MCP603 (Asn) but energetically compatible with that of REI (Gln). This was due to a subtle shift in backbone conformations of the loops in the context of the CEA framework. Therefore, the conformation of REI was used as the backbone conformation for the loop. After mutation, side chains did not exhibit bad contacts.

L1

L1 has a conformation which is constrained by a hydrophobic anchor at position 29. It has the same length in CEA and REI; all of the key residues identified by Chothia et al.[29] as determining the structure of the loop showed high similarity. Therefore, the structure of L1 from REI was chosen. Side chain replacement was straightforward and performed using the MUTATE procedure described above.

To further investigate possible backbone conformations for L1, we have used the TWEAK and NOVOSIDE algorithms to generate 250 random conformations for the 7-residue loop. These were then subjected to energy minimization as described in Methods. Minimized conformations were selected whose hydrophobic anchor residue Val-L29 superimposed on that of the canonical structure to within 0.5 Å rms (C_β and C_γ atoms only). These conformations were then examined visually to ensure that conserved contacts with framework residues were maintained. Four loop structures were found which satisfied these requirements. The four structures are shown in Figure 2, superimposed on the canonical structure. The conformations are similar to each other and to the canonical structure from L30–L32, but differ from L26–L29. The variation is considerable: the most disparate from the canonical structure exhibits an Asp-L28 C_α which deviates from that of the canonical structure by 4.1 Å. It is difficult to select between the conformations based on energetic or solvent exposure criteria: Asp-L28 is exposed in all conformations while force field energy

differences are dominated by electrostatic interactions of the Asp. These are exaggerated due to evaluation of these interactions in vacuum. The canonical structure and closely related conformations pack against conserved residues L68 and L69 in the framework. This packing is common in known crystal structures. Based on this packing argument we have retained the canonical structure in the CEA antibody model though our results strongly indicate that the additional loop conformations shown in Figure 1 should be considered as alternate conformations.

H3

The H3 loop is the longest loop considered, and has little homology to H3 in other crystallographically determined structures. We first examined the H3 loop by using the A. Jones algorithm.[27] This procedure extracted nine loops of the same length from the data base which did not exhibit gross overlaps with the framework of the molecule. All nine exhibited no sequence homology to the H3 loop. Side chains were placed using the NOVOSIDE strategy described above. The resulting loops were used as the first nine entries in a library of starting conformations for minimization.

An initial set of 100 backbone conformations was generated for H3 using GENLOOP. For both generation and minimization, the loop residues were defined to be H95–H100k. During minimization, all loop atoms were free to move. The atoms of two conserved residues at each end of the loops were tethered to their starting positions using a harmonic constraint of 10 kcal/Å. All other atoms were held fixed. Solvent was not explicitly included in these minimizations; however, a distance dependent force field was employed to partially account for electrostatic solvent screening. The resulting H3 loop conformations were examined visually, and compared with known crystal structures in the Protein Database.

The backbone conformation of the bases of H3 loops in available crystal structures fall into two classes. In a first "kinked" class there is a sharp turn at Asp 101 which is absent in a second ("extended") class. All of the 100 loops generated belong to the "kinked" class as a consequence of treating this conserved Asp as part of the fixed framework during loop generation. Since there was no a priori way to assert that this conformation was the only conformation available to H3 in CEA, 800 additional conformations were generated and minimized with the Asp free to move.

The minimized H3 structures were initially evaluated in the context of the remainder of the CEA model using two energetic criteria. The first was the force field energy of the structure. The second was the Eisenberg-McLachlan (EM) scale of solvation free energy derived from partition of amino acids between water and octanol.[45] These two "energies" are not additive: interactions such as hydrogen bonds considered explicitly in a force field are included in an average way in the EM solvation free energy by using octanol to model a protein interior. A plot of force field energies vs. the solvation energies for the 150 lowest energy structures is shown in Figure 3. We should note that none of these lowest 150 structures resulted from minimization of the nine conformations of the loop extracted from the database. Little correlation between solvation free energies and force field energies is evident in these plots. The two lowest force field energy structures have high solvation free energies. This is due in part to electrostatics: the conformations of these loops have strong hydrogen bonds between the oxygens of Asp 100c and the backbone NH groups of the loop which result in burial of both oxygens of the Asp-100c (Fig. 4). Such conformations will have relatively high solvation free energies. The lowest solvation free energy structures exhibit an Asp which is exposed to solvent, resulting in lower solvation free energies and higher force field energies since they do not include the effects of strong interactions of Asp-100c with the protein.

To investigate the problem of simultaneously exposing Asp-100c to solvent and minimizing the surface area of flanking hydrophobic residues, we have plotted the exposed hydrophobic surface area of the

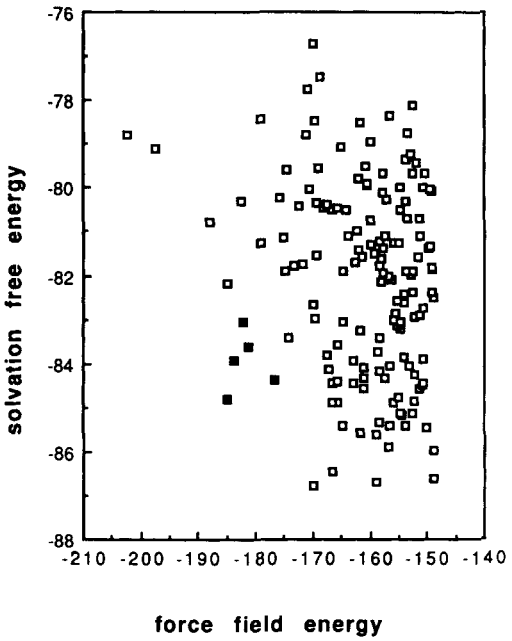

Fig. 3. A scatter plot of the Eisenberg–McLachlan solvation free energy vs. the force field energy of the 150 lowest force field energy conformations for H3 generated and minimized as described in the text. Minimized conformations are used for these calculations. Energies and free energies are in kcal/mol. The five filled-in squares represent the five structures selected for detailed consideration as described in the text.

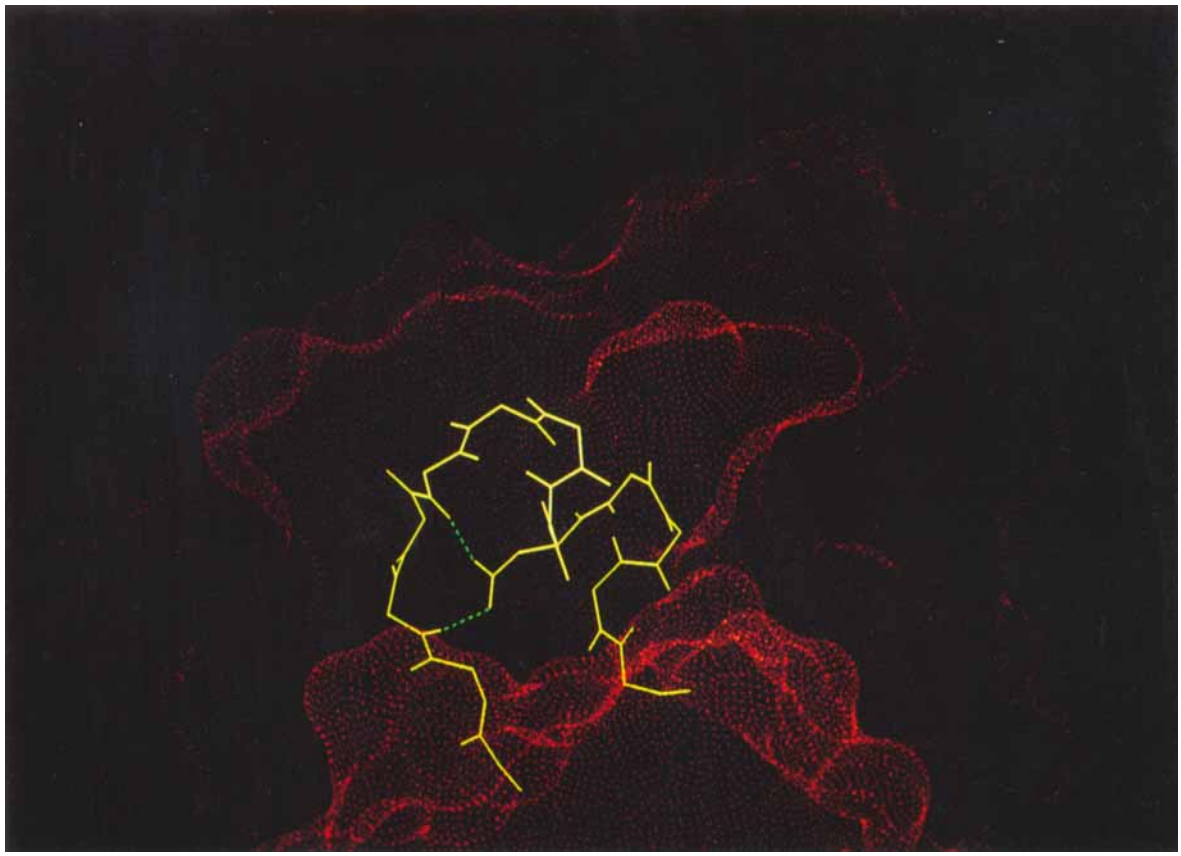

Fig. 4. A sample conformation which ranks artificially low in force field energy and high in EM solvation free energy. Strong hydrogen bonding interactions "solvate" the charges on the Asp resulting in inaccessibility to true solvent.

model vs. the exposed surface area of Asp 100c for the 150 structures described above in Figure 5. An anticipated problem mentioned in the introduction was the difficulty in simultaneously minimizing the hydrophobic surface area of hydrophobic residues flanking Asp-100c while exposing the Asp. From this figure it is clear that the two criteria are not mutually exclusive. That the EM scale is effective in ranking the loops in order of hydrophobic surface area is evident from Figure 6, which shows the correlation between hydrophobic surface area and the evaluated EM solvation free energy.

As representative structures which satisfy the requirements that (1) the force field energy is low; (2) the EM solvation energy is low; and (3) the oxygens of Asp-100c are partly exposed, we examined the 5 minimized structures whose CHARMM energies were lower than −175 and whose EM solvation energies were lower than −82. All five structures exposed the Asp though oxygen atoms are buried to varying degrees. They are shown in the context of the remainder of the CEA model in Figure 7. It is evident from the figure that these structures span a considerable range in conformation though the force

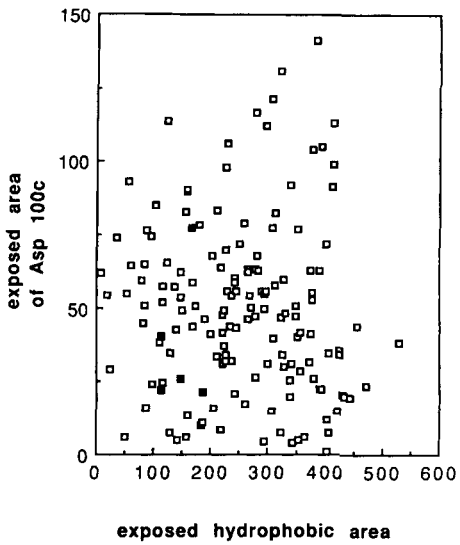

Fig. 5. A scatter plot of the surface accessible area of Asp-100c vs. the relative exposed hydrophobic area of the light and heavy variable domains for conformations of H3. Relative exposures were generated by subtracting the lowest area found. Selection of data and definition of filled squares are as described for Figure 3. Areas are in Å^2.

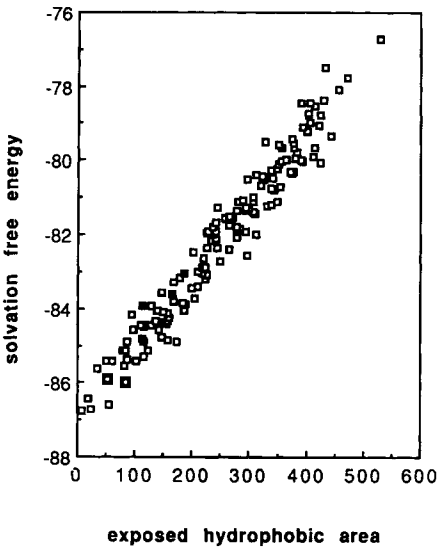

Fig. 6. A scatter plot of the EM solvation free energy vs. the relative exposed hydrophobic area for the light and heavy variable domains for conformations of H3. Selection of data and definition of filled squares are as described for Figure 3. Free energies are in kcal/mol; areas are in Å^2. Relative exposed hydrophobic area was generated as described in Figure 5.

field energies are within 7 kcal/mol and the EM solvation free energies are within 2 kcal/mol. Conformations A and B to the left in the figure do not exhibit a kink at their base as described earlier. A kink at the base of the loop similar to that in MCPC603 is present in the two conformations in the center of the figure (C and D) and the one to the right (E).

Examination of the base of these loops reveals a pattern for the unkinked structures. The side chain of Met-100k extends into solvent, leaving a solvent-inaccessible cavity which was occupied by Phe-100k in the original MCPC603 structure. At the same time the conserved salt bridge between Asp-H101 and Arg-H94 is broken, leaving Asp-100c buried with a small exposure of its terminal oxygen to the hydrophobic cavity. For conformations A and B, the cavity has total surface areas of 18.14 and 20.37 Å^2. The ratio of polar to nonpolar surface area, using a definition given in ref. 50, is 0.39 and 0.42 for the two conformations, consistent with our description of the cavity as hydrophobic. The cavity volumes are 72.04 and 75.57 Å^3, respectively; though of sufficient size to hold a water molecule, no energetically favorable placements were found using the search method described in ref. 51. To investigate this pattern of side chain packing, ANNEALSIDE was applied to the backbone conformations for the loops shown in Figure 7 as described in Methods. Values for the overlap relief parameter ϵ described in Methods of 0.6 and 0.0 Å were tried. The pattern of an exposed Phe-100k and a broken Asp-H101–Arg-H94 salt bridge persisted in all side chain packings found. In the kinked structures, the pattern is quite different: the Met-100k fills the cavity leaving the Asp-H101–Arg-H94 salt bridge intact. For these combined reasons we consider the unkinked structures unlikely. We will therefore limit discussion of H3 in the intact model to the three kinked structures C, D, and E in Figure 7. It is interesting to note that we have found cavities at the base of unkinked H3 loops in several crystal structures available in the Brookhaven Protein Data Bank, namely in HyHEL-5 (2HFL), NEW (3FAB), and 4-4-20 (4FAB). The range of cavity surface area and volume is slightly less than that described above (40–60 Å^2 and 9.5–15 Å^3), whereas the ratio of polar to nonpolar surface area is higher (0.55–1.84). In spite of the higher ratio of polar to nonpolar surface, the same algorithm used above could not identify an energetically favorable placement for a water molecule (none was listed by crystallographers). Thus, the existence of a cavity alone might not discriminate the kinked from the unkinked class of H3 loops. These crystal structures do not, however, result in the energetically unfavorable burial of a charged residue.

All three structures are displayed in the remainder of the CEA model in Figure 8. From the picture, the similarity in the conformation of the Met-100k side chain is clear (yellow residue at the bottom of the cleft), as is the similarity of the conformations of the two prolines which comprise the N-terminal end of the loop. A second conformation for these prolines was found but only in nonkinked H3 conformations. From this point, the conformations diverge resulting in quite different locations for the tops of the loops. This could reflect the conformational flexibility expected for loops of this size: long loops are frequently characterized by high temperature factors or by an absence of well-defined electron density in crystallographic data.

DISCUSSION

It is generally accepted that the primary determinant of antibody specificity is complementarity of surface. The surface of the combining site of the model for the CEA antibody is shown in Figure 9. The combining site of the CEA model is characterized by a central depression lined by the polar aromatic residues Tyr-91 and Tyr-94 from L3, the hydrophobic residues Ala-100j and Leu-100 from H3, and prolines H95 and H96 from the base of H3. A prominent feature of the combining site is the fully exposed side chain of Trp-L50. A relatively large portion of the surface is comprised of atoms from nonpolar residues, shown in green. From the few antibody–antigen complexes which are known, a striking feature is the charge complementarity between the surfaces of the antibody and its antigen. The charged residues in the hypervariable loops of CEA, Asp-28 from L1, and Arg-54 from L2, Arg 31 from H1, and Asp 100c from H3, are located at the

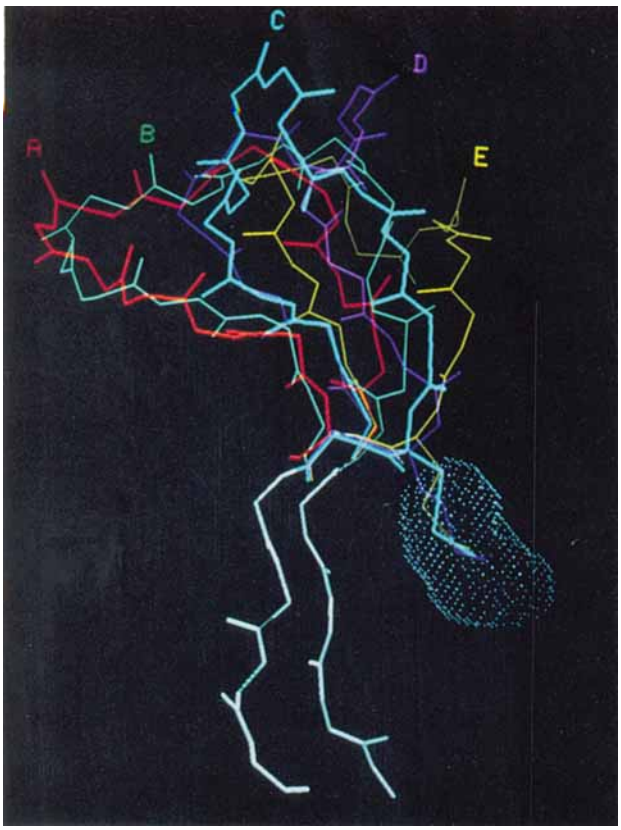

Fig. 7. The five conformations of H3 selected on the basis of low force field energy and Eisenberg–McLachlan solvation free energies. The conformations C, D, and E have a kink at their base similar to that found in MCPC603. The kink is absent in conformations A and B. The dotted surface represents a cavity found when conformations A or B are placed in the context of the remainder of the CEA model. This cavity is filled by Met 100k in conformations C, D, and E as shown.

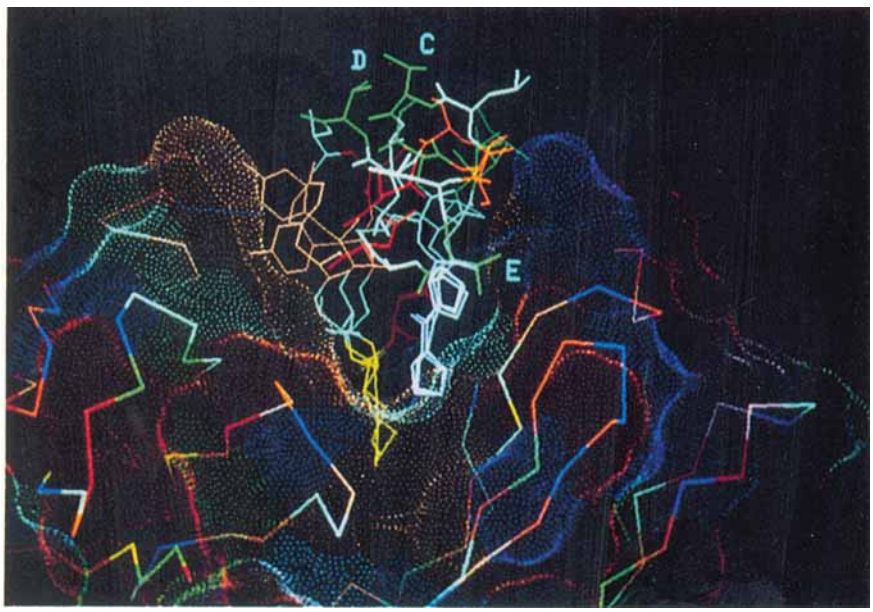

Fig. 8. The three kinked conformations of H3 discussed in the text are shown superimposed on the remainder of the antibody molecule. Only the C_α trace and the solvent accessible surface of the remainder of the molecule are shown. The surface shown is the solvent accessible surface generated using a Connolly algorithm[60] after removing the H3 loop from the molecule. The surface is color coded by the hydrophobic nature of the amino acids.

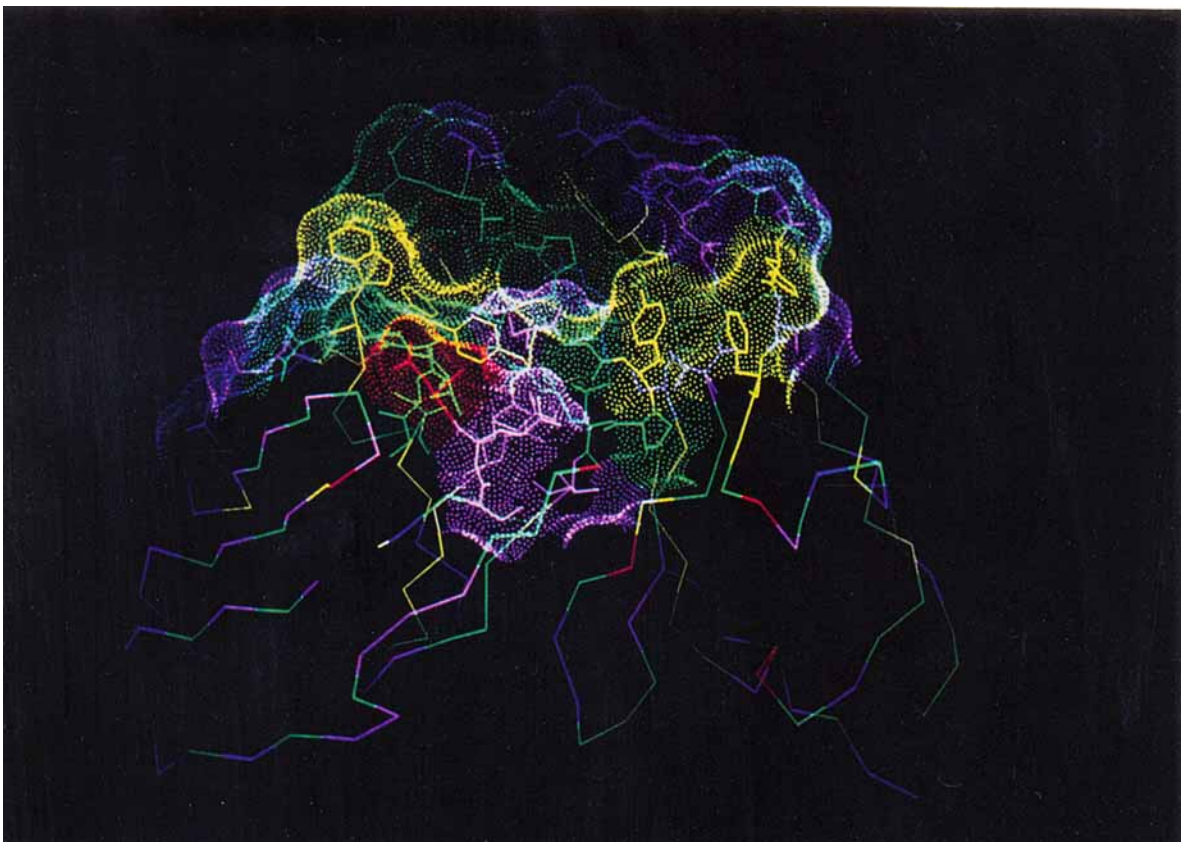

Fig. 9. The surface of the CEA antibody model with conformation D of H3 selected. The solvent accessible surface of the combining site has been colored to highlight certain features. Arg and Lys are blue; Asp and Glu are red. Nonpolar residues are green (Gly, Met, Pro, Leu, Ile, Val, Ala). Aliphatic polar residues are purple (Ser, Thr, Gln, Asn). All aromatic residues are yellow (His, Trp, Phe, and Tyr). The residues included at atomic detail and used in the surface calculation are 26–34, 50–56, 89–97 from the light chain and 26–35, 50–56, and 94–102 from the heavy chain. This set includes additional residues which are variable in sequence in the complementarity determining regions defined by Kabat et al.[3]

periphery of the combining site. Hydrogen bonding complementarity is also expected to play a key role at the interface. The purple surface in Figure 9 is surface contributed by polar side chains which can potentially form hydrogen bonds to antigen.

Another feature of this combining site which is common to many known sequences of antibodies is its high content of aromatic residues, particularly tyrosines. The reason for such concentration has been suggested by Padlan[52] to be the relatively large hydrophobic surface and small number of side chain angles for aromatic residues which will be frozen on binding. These side chains can, therefore, contribute significantly to the binding energy through the hydrophobic effect while paying a relatively small price in entropy. Tyrosines can further contribute a hydrogen bond, thus providing specificity to the hydrophobic surface of the antibody. Initial site-directed mutagenesis studies of the hypervariable region of this antibody are described in detail in a later paper.[21]

A key problem in modeling in general is the selection of folded vs. misfolded structures.[50,53] A problem, reflected not only by this study but by others in the literature, is the lack of a consistent way to incorporate all features known to contribute to the total conformational free energy of a protein in a consistent fashion. In self-criticism, it is clear that our use of a force field to characterize a conformation, even augmented by solvation free energy scales based on transfer from water to octanol, is incorrect: attractive van der Waals (London) interactions have been explicitly counted between protein atoms but not between protein atoms and water atoms.[37,53] The inclusion of explicit waters could solve this problem, but at a cost: the use of explicit waters cannot properly model the hydrophobic effect in comparing two conformations. This is due to the fact that the hydrophobic effect is entropic: it is a property which depends on the dynamic behavior of waters surrounding the selected conformation of the loop. To compare two disparate loop structures, properly accounting for the hydrophobic effect, would require a free energy simulation which slowly

changed the conformation of the loop between the two structures, which is not feasible at the present time.

Scales for the hydrophobic effect (and solvation free energy in general) continue to evolve. Recently, several authors have suggested that the magnitude of the hydrophobic effect is stronger than indicated in earlier studies.[54–56] In our work only the proportionality between the hydrophobic surface and hydrophobicity has been used to rank structures; a shift in the underlying slope would not affect this ranking. The assertion that burial of charged or polar groups should be proportional to surface area is more problematic: electrostatic properties are not in general directly proportional to exposed surface area.[57] It might be interesting in the future to explore the shape-dependent hydrophobic method described previously[55,56,58] as a means to evaluate trial structures. The hydrophobic solvation for the neutral protein would be computed using this approach. Contributions to solvation from polar or charged nature of protein atoms could be added using continuum electrostatics to "charge" the molecule.[59] Attractive London forces between protein atoms could be either completely removed[51] or, if included, the effect of missing solvent interactions could be included using continuum methods.[56] The molecular mechanics force field could be used to evaluate internal energies only (bond stretch, angle bend, torsions, cross-terms). Unlike an earlier paper which used molecular dynamics followed by minimization to convert randomly generated high energy structures to low energy structures, in the present paper we have employed minimization only. Molecular dynamics followed by minimization, or simulated annealing followed by minimization, is considerably more efficient in converting initial randomly generated conformations to low energy conformations than is energy minimization alone.[34] This conversion efficiency is obtained at the cost of depopulating large regions of conformational space which may be accessible at room temperature and which may be characterized by solvation free energies sufficiently large to ameliorate observed differences in force field energies. Minimization was selected over simulated annealing followed by minimization for this study as a means to relieve bad van der Waals overlaps in the initial randomly generated conformations without permitting large conformational changes. This provided an ensemble of conformations which could be ranked using a variety of criteria, including the EM scale for solvation free energy.

Cavities present yet another area which is both central to the present work and poorly characterized in proteins, both theoretically and experimentally. Cavities clearly exist in protein interiors, and from the examples given, can exist at the base of H3. From the limited distribution of sizes of known cavities in proteins,[56] it is clear that they incur a free energy cost which increases with cavity size.[56,58] Since the matrix of the protein seems not to relax around this cavity in crystallographic structures, the appropriate metaphor for the analysis of the creation of the cavities discussed at the base of H3 is that of creating a cavity in a hydrocarbon solid rather than in a hydrocarbon liquid.[58] A recent analysis of this process[58] yields similar estimates for the free energy of cavity formation to those derived earlier from the distribution of cavity volumes in known proteins.[56] We will use the estimate of 60 cal/mol/Å^3 to estimate the free energy cost of creating cavities at the base of H3. This gives a cost estimate of 2.4 to 4.5 kcal/mol for cavity volumes of 40–75 Å^3. Where cavities do exist, they are often correlated with function or increased flexibility.[56] This leads to a functional conjecture on H3: altering residues at the base of H3 may provide a mechanism to "switch" between the extended more mobile conformation and the kinked well-packed conformation, regulating the flexibility of the loop at its base by opening or closing the cavity discussed here.

CONCLUSIONS

We have constructed a model of a clinically important antibody (CEA) specific for the carcinoembryonic antigen. The purpose in constructing this model is to provide a framework for the design and interpretation of site-directed mutations whose aim is to alter the specificity and binding of the antibody. In a very real sense, the conclusions of this paper are to be found in the interpretation of the binding site and the results of these site-directed mutations which are presented in a later paper. However, it is possible to state several conclusions here which may be of general interest to predictive studies of antibody and protein structure in general.

1. Conformational search provides a useful tool to explore the uniqueness of results from knowledge based modeling and to suggest alternate conformations. It can be especially useful to examine cases where unusual sequences, such as combinations of glycines or prolines, exist in a loop.

2. Though conformations of H3 in solved crystal structures in Protein Database exhibit considerable variability in length, sequence, and conformation, their base conformations fall into two classes which we have called kinked and unkinked.

3. Conformational search procedures applied to H3 in CEA clearly suggest that these are the only conformations available to the loop. Packing of residues at the base of the loop, illustrated by the creation of a solvent inaccessible cavity, can be used to differentiate between these structures although a similar cavity has been found in crystal structures deposited in the Protein Database.

4. The inclusion of effects of solvent can be criti-

cal to evaluation of loop structures especially when charged amino acids can be buried.

ACKNOWLEDGMENTS

This work was supported in part by grants P01 CA 43904 from the National Cancer Institute and Cancer Center Support Grant CA33572 (M.T.M.), NIH Grant RR-00442 and NSF equipment Grant DMB-84-02496 (R.M.F., D.L.Y. and K.C.S.), and ONR Grant N0014-90-J-1713 (K.C.S.). Partial support from Kyowa Hakko Kogyo Co. for one of us (K.A.) is gratefully acknowledged. M. Mas would like to thank Dr. John E. Shively (Division of Immunology, Beckman Research Institute) and Dr. Arthur D. Riggs (Division of Biology) for suggesting this work. We would like to warmly acknowledge the unflagging interest and encouragement of Dr. Cyrus Levinthal. The side chain programs described in this work owe significantly to earlier unpublished algorithm exploration by Dr. Peter Shenkin. We also thank Dr. Alex Rashin (Biosym) for implementing and enhancing his cavity program for our use.

REFERENCES

1. Alzari, P.M., Lascombe, M.-B., Poljak, R.J. Three-dimensional structure of antibodies. Annu. Rev. Immunol. 6: 555–580, 1988.
2. Davies, D.R., Padlan, E.A. Antibody-antigen complexes. Annu. Rev. Biochem. 59:439–473, 1990.
3. Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M., Gottesman, K.S. Sequence of Proteins of Immunological Interest. Washington, DC: U.S. Department of Health and Human Services, National Institute of Health, 1987.
4. Amzel, L.M., Poljak, R.J., Saul, F., Varga, J.M., Richards, F.F. The three-dimensional structure of a combining region ligand complex of immunoglobulin NEW at 3.5 Å resolution. Proc. Natl. Acad. Sci. U.S.A. 71:1427–1430, 1974.
5. Edmundson, A.B., Ely, K.R., Girling, R.L., Abola, E.E., Schiffer, M., Westholm, F.A., Fausch, M.D., Deutsch, H.F. Binding of 2,4-dinitrophenyl compounds and other small molecules to a crystalline λ-type Bence-Jones dimer. Biochemistry 13:3816–3827, 1974.
6. Segal, D.M., Padlan, E.A., Cohen, G.H., Rudikoff, S., Potter, M., Davies, D.R. The three dimensional structure of a phosphocholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. Proc. Natl. Acad. Sci. U.S.A. 71:4298–4302, 1974.
7. Epp, O., Lattman, E.E., Schiffer, M., Huber, R., Palm, W. The molecular structure of a dimer composed of the variable portions of the Bence Jones protein Rei refined at 2.0 Å resolution. Biochemistry 14:4943–4952, 1975.
8. Saul, F.A., Amzel, L.M., Poljak, R.J. Preliminary refinement and structural analysis of the Fab fragment from the human immunoglobulin New at 2.0 Å. J. Biol. Chem. 253: 585–597, 1978.
9. Marquart, M., Deisenhoffer, J., Huber, R., Palm, W. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. J. Mol. Biol. 141: 369–391, 1980.
10. Furey, W., Jr., Wang, B.C., Yoo, C.S., Sax, M. Structure of a novel Bence-Jones protein (Rhe) fragment at 1.6 Å resolution. J. Mol. Biol. 167:661–692, 1983.
11. Amit, A.G., Mariuzza, R.A., Phillips, S.E.V., Poljak, R.J. Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. Science 233:747–753, 1986.
12. Satow, Y., Cohen, G.H., Padlan, E.A., Davies, D.R. Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Angstroms. J. Mol. Biol. 190: 593–604, 1986.
13. Suh, S.W., Bhat, T.N., Navia, M.A., Cohen, G.H., Rao, D.N., Rudikoff, S., Davies, D.R. The galactan-binding immunoglobulin Fab J539: An x-ray diffraction study at 2.6 Å resolution. Proteins 1:74–80, 1986.
14. Tulloch, P.A., Colman, P.M., Davies, P.C., Laver, W.G., Webster, R.G., Air, G.M. Electron and x-ray diffraction studies of influenza neuraminidase complexed with monoclonal antibodies. J. Mol. Biol. 190–225, 1986.
15. Colman, P.M., Laver, W.G., Varghese, J.N., Baker, A.T., Tulloch, P.A., Air, G.M., Webster, R.G. Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. Nature (London) 326:358–363, 1987.
16. Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C., Davies, D.R. Three-dimensional structure of an antibody-antigen complex. Proc. Natl. Acad. Sci. U.S.A. 84:8075–8079, 1987.
17. Herron, J.N., He, X.-M., Mason, M.L., Voss, Jr., E.W., Edmundson, A.B. Three-dimensional structure of a fluorescein-fab complex crystallized in 2-methyl-2,4-pentanediol. Proteins 5:271–280, 1989.
18. Padlan, E.A., Silverton, E.W., Sheriff, S., Cohen, G.H., Smith-Gill, S.J., Davies, D.R. Structure of an antibody-antigen complex: Crystal structure of the HyHEL-10 Fab-lysozyme complex. Proc. Natl. Acad. Sci. U.S.A. 86:5938–5942, 1989.
19. Strong, R.K., Campbell, R., Rose, D.R., Petsko, G.A., Sharon, J., Margolies, M.N. Three-dimensional structure of murine anti-*p*-azophenylarsonate Fab 36-71. 1. X-ray crystallography, site-directed mutagenesis, and modeling of the complex with hapten. Biochemistry 30:3739–3748, 1991.
20. Wagener, C., Yang, Y.H.J., Crawford, F.G., Shively, J.E. Monoclonal antibodies for carcinoembryonic antigen and related antigens as a model system: A systematic approach for the determination of epitope specificities of monoclonal antibodies. J. Immunol. 130:2308–2315, 1983.
21. Hu, S.-Z., Aisaka, K., Riggs, A.D., Mas, M.T. Site-directed mutagenesis of the anti-CEA antibody combining site. Manuscript in preparation.
22. Padlan, E.A., Davies, D.R., Pecht, I., Givol, D., Wright, C. Model-building studies of antigen-binding sites: The hapten-binding site of MOPC-315. Cold Spring Harbor Symp. Quant. Biol. 41:627–637, 1976.
23. Mainhart, C.R., Potter, M., Feldmann, R.J. A refined model for the variable domains (Fv) of the J539 β(1,6)-D-galactan-binding immunoglobulin. Mol. Immunol. 21: 469–478, 1984.
24. Feldmann, R.J., Potter, M., Glaudemans, P.J. A hypothetical space-filling model of the V-regions of the galactan-binding myeloma immunoglobulin J539. Mol. Immunol. 18:683–698, 1981.
25. Pawlita, N., Mushinski, E., Feldmann, R.J., Potter, M. A monoclonal antibody that defines an idiotype with two subsites in galactan-binding myeloma proteins. J. Exp. Med. 154:1946–1956, 1981.
26. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein databank: A computer based archival file for macromolecular structure. J. Mol. Biol. 112:535–542, 1977.
27. Jones, T.A., Thirup, S. Using known substructures in protein model building and crystallography. EMBO J. 5:819–822, 1986.
28. Chothia, C., Lesk, A.M. Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. 196: 901–917, 1987.
29. Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M., Poljak, R.J. Conformations of immunoglobulin hypervariable regions. Nature (London) 342:877–883, 1989.
30. Unger, R., Harel, D., Wherland, S., Sussman, J.L. A 3D building blocks approach to analyzing and predicting structure of proteins. Proteins 5:355–373, 1989.
31. Bruccoleri, R.E., Karplus, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers 26:137–168, 1987.
32. Moult, J., James, M.N.G. An algorithm for determining

the conformation of polypeptide segments in proteins by systematic search. Proteins 1:146–163, 1986.
33. Shenkin, P.S., Yarmush, D.L., Fine, R.M., Wang, H., Levinthal, C. Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ring-like structures. Biopolymers 26:2053–2085, 1987.
34. Fine, R.M., Wang, H., Shenkin, P.S., Yarmush, D.L., Levinthal, C. Predicting antibody hypervariable loop conformations II: Minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations. Proteins 1:342–362, 1986.
35. Dudek, M.J., Scheraga, H.A. Protein structure prediction using a combination of sequence homology and global energy minimization. I. Global energy minimization of surface loops. J. Comp. Chem. 11:121–151, 1990.
36. Bruccoleri, R., Karplus, M. Conformational sampling using high temperature molecular dynamics. Biopolymers 29:1847–1862, 1990.
37. Martin, A.C.R., Cheetham, J.C., Rees A.R. Modeling antibody hypervariable loops: A combined algorithm. Proc. Natl. Acad. Sci. U.S.A. 86:9268–9272, 1989.
38. Go, N., Scheraga, H.A. Ring closure and local conformational deformation of chain molecules. Macromolecules 3: 178–187, 1970.
39. Katz, L., Levinthal, C. Interactive computer graphics and representation of complex biological structures. Annu. Rev. Biophys. Bioeng. 1:465–504, 1972.
40. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comp. Chem. 4:187–217, 1983. The force field used in CHARMM is discussed in detail in the Harvard Ph.D. thesis of Walter E. Reiher, 1985.
41. GEMM (Generate, Emulate, Manipulate Macromolecules) is the name of an implementation of molecular dynamics and minimization code written for a Star ST-100 array processor by Bernard Brooks at the National Institute of Health (unpublished).
42. Brünger, A.T., Karplus, M., Petsko, G.A. Crystallographic refinement by simulated annealing: Application to crambin. Acta Crystallogr. A45:50–61, 1989.
43. Hagler, A.T., Huler, E., Lifson, S. Energy functions for peptides and proteins. J. Am. Chem. Soc. 96:5319–5327, 1974.
44. Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A. J. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsion potentials for the naturally occurring amino acids. Phys. Chem. 79:2361–2381, 1975.
45. Eisenberg, D., McLachlan, A.D. Solvation energy in protein folding and binding. Nature (London) 319:199–203, 1986.
46. Cabilly, S., Riggs, A.D., Pande, H., Shively, J.E., Holmes, W.E., Rey, M., Perry, L.J., Wetzel, R., Heyneker, H.L. Immunoglobulin transcripts and molecular history of a hybridoma that produces antibody to carcinoembryonic antigen. Proc. Natl. Acad. Sci. U.S.A. 81:3273–3277, 1984.
47. Ponder, J.W., Richards, F.M. Tertiary templates for proteins. J. Mol. Biol. 193:775–791, 1987.
48. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. J. Equation of state calculations by fast computing machines. Chem. Phys. 21:1087–1092, 1953.
49. Sutcliffe, M.J., Hayes, F.R.F., Blundel, T. Knowledge based modeling of homologous protein, part II: Rules for the conformations of substituted side chains. Protein Eng. 1:385–392, 1987.
50. Novotny, J., Bruccoleri, R., Karplus, M. An analysis of incorrectly folded protein models. J. Mol. Biol. 177:787–818, 1984.
51. Rashin, A.A., Iofin, M., Honig, B. Internal cavities and buried waters in globular proteins. Biochemistry 25:3619–3625, 1986.
52. Padlan, E.A. On the nature of antibody combining sites: Unusual structural features that may confer on these sites an enhanced capacity for binding ligands. Proteins 7:112–124, 1990.
53. Novotny, J., Rashin, A.A., Bruccoleri, R.E. Criteria that discriminate between native proteins and incorrectly folded models. Proteins 4:19–30, 1988.
54. Dill, K.A. The meaning of hydrophobicity. Science 250: 297, 1990.
55. Sharp, K.A., Nicholls, A., Fine, R.M., Honig, B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. Science 252:106–109, 1991.
56. Flores, F., Tomasi, J. Evaluation of the dispersion contribution to the solvation energy. A simple computational model in a continuum approximation. J. Comp. Chem. 10: 616–627, 1989.
57. Gilson, M., Rashin, A., Honig, B. On the calculation of electrostatic interactions in proteins. J. Mol. Biol. 183: 503–516, 1985.
58. Nicholls, A., Sharp, K., Honig, B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 11:281–296, 1991.
59. Sharp, K., Honig, B. Calculating the total electrostatic energies with the nonlinear Poisson-Boltzman equation. J. Phys. Chem. 94:7684–7692, 1990.
60. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. Science 306:287–290, 1983.