PROTEINS: Structure, Function, and Genetics 1:267-279 (1986)

Calculating Three-Dimensional Changes in Protein Structure Due to Amino-Acid Substitutions: The Variable Region of Immunoglobulins

Mark E. Snow and L. Mario Amzel
The Laboratory for Molecular Structure and Function, Department of Biophysics, The Johns Hopkins University, School of Medicine, Baltimore, Maryland 21205

ABSTRACT **A procedure (coupled perturbation procedure, CPP) is introduced as a specific method for calculating the detailed three-dimensional structure of a protein molecule which has a nummber of amino-acid substitutions relative to some previously determined "parent" protein structure. The accuracy of the procedure is tested by calculating the conformation of a region of the human immunoglobulin fragment Fab Kol based on the analogous region of the human immunoglobulin fragment Fab New. Both structures have previously been determined crystallographically. The calculated model is accurate to the extent that both of the sequence differences in the region are modeled correctly and that conformational changes in a number of nearby residues are correctly identified. CPP is shown to give better results than other commonly used modeling procedures when applied to the same problem.**

Key words: molecular model building, energy minimization, homology modeling, site-specific mutagenesis

INTRODUCTION

Background

Many situations arise in which one is interested in the three-dimensional structures of a group of protein molecules that differ from each other in only a small number of amino-acid residues. Site-directed mutagenesis experiments offer many examples of such situations. Other examples arise in the study of protein families such as the globins or immunoglobulins or in the study of a single protein from a variety of species. If one of the structures in such a group of proteins has been determined experimentally, it should be possible to calculate the structure of another protein molecule in the group by replacing the necessary side-chains in the known structure and evaluating the resultant change in the conformational energy surface.*[1] Unfortunately, to date no reliable procedure has been developed to generate detailed and accurate† models of unknown structures by this method. When such a procedure becomes available, it will greatly reduce the number of structures which have to be determined experimentally, and it will allow one to evaluate the three-dimensional structure of a protein before the molecule is produced by site-directed mutagenesis techniques.

Previous reports of comparative model building have dealt almost exclusively with cases involving distantly related proteins.[1-7] These reports have done an excellent job of providing approximate models which could be used to design new experiments, including some experiments to probe the validity of the proposed structures themselves. These studies do not address, however, the problem of obtaining a detailed and accurate (as defined above) model for a protein of unknown structure. In the only attempt to analyze the accuracy of such models, Read et al.[8] compared the structure of *Streptomyces griseus* trypsin (SGT) with the models produced by Jurasek et al[9] and by Greer.[3] Both models were based on the expected structural homology of SGT to bovine trypsin and other serine proteases. Read et al[8] conclude that, as expected, the models are much more accurate in homologous than in nonhomologous regions. In addition, predictions involving nonconserved residues in homologous regions are more susceptible to errors in side-chain conformation. Since many functionally important amino acids (those involved in catalysis, substrate specificity, etc.) belong to this class, these errors represent a serious drawback for this kind of modeling. Greater understanding of the strengths and limitations of comparative model building of distantly related proteins will be gained when the three-dimen-

Received July 29, 1986; accepted September 25, 1986.

Address reprint requests to Dr. L. Mario Amzel, The Laboratory of Molecular Structure and Function, Department of Biophysics, The Johns Hopkins University, School of Medicine, Baltimore, MD 21205.

* The conformational energy surface of a molecule is the multidimensional "surface" defined by evaluating the energy of the molecule as a function of the coordinates of its atoms. Combinations of atomic coordinates representing dihedral angles, valence bond angles, bond lengths, etc., are more commonly used as independent variables.

†If a model is to be useful in addressing questions of binding or enzymatic mechanism, it must not only have the overall fold of the main chain correct; it must also be accurate with regard to the positioning of the individual amino-acid side-chains in their correct conformational energy-wells.

sional structures of more of the proteins modeled become available for comparison.

Fewer modeling studies have been reported for proteins differing in only a few amino acids. The conformation of a stable variant protein which results from the substitution of only a few surface amino-acid residues is very likely to closely resemble that of the "parent" molecule. In such cases, all of the structural changes are likely to occur in the regions surrounding the replaced amino acids and to affect mainly the side-chains of the new residues and of those residues which contact the old and/or new side-chains. Comparative modeling in such cases may consist of attaching the new side-chains(s) to the backbone in some conformation that is then subjected to energy minimization using semi-empirical energy potential functions. Because the conformational energy surface contains many local energy minima, the conformation of a residue in the final model is dependent on the initial choice of conformation. In previous studies, two methods have been used to choose the conformation(s) of the side-chain(s) which are subjected to energy refinement.

In the first approach (maximum overlap procedure, MOP) (for example, see references 1,2,10), the new side-chain(s) is built in a carefully chosen conformation; then the entire structure is energy refined. Methods for choosing the initial conformation for the new side-chain(s) include following the side-chain dihedrals from the residue in the parent structure as far as possible and selecting dihedrals chosen from calculated low-energy conformations of the new residue in peptides with similar phi and psi values. This procedure is very easy to implement and will work in many cases, but, in many other cases, it will result in a structure which is trapped in a local energy minimum.

In the second approach (minimum perturbation procedure, MPP),[11] after building the new side-chain(s), a conformational search is performed about the dihedral angles of the side-chain(s) and the low-energy conformations are identified. These structures are then energy minimized using constraints which assure that only atoms close to the replaced residue are allowed to move. These constraints are introduced to keep portions of the molecule assumed to be unchanged by the introduction of the new residue close to their positions in the parent structure.[11] Although this procedure avoids many of the problems of MOP, it is still possible for it to fail to find the lowest energy conformation for the replaced residue and for neighboring residues as well. This is particularly true in cases where the interactions of the new side-chain with neighboring residues are substantially different from the analogous interactions in the parent structure. Comparing the known three-dimensional structures of several immunoglobulins in regions where they differ in only a few amino acids, we found that in some cases the side-chains of conserved amino acids that are adjacent to replaced residues change in conformation to a different energy minimum which is not accessible from the original conformation by energy minimization procedures.

In this study we have used an approach (coupled perturbation procedure, CPP) in which conformational searches are performed not only for the side-chain(s) of the replaced residue(s) but also for other side-chains which are identified as belonging to the "dependency set" of the replacement(s). The low-energy structures found in these searches are then subjected to energy minimization. As in the MPP, the refinement is limited to a region surrounding the new residue(s). Provisions are taken as part of the procedure to limit the amount of computation necessary to find all of the low-energy conformations which are used as starting points for energy refinement.

Before using the method to calculate the atomic coordinates of a protein of unknown structure, a procedure should be tested by modeling a protein of known structure. This test has not been performed in most previous reports. In this study we have tested the procedure by using the reported coordinates of two human immunoglobulin fragments, Fab New[12] and Fab Kol.[13] Since the structures of these proteins were determined from different crystal forms, in different laboratories using different procedures, we feel that this modeling problem constitutes a fairly rigorous test of the procedure.

Specifically, we have calculated the structure of a region of the immunoglobulin fragment Fab Kol based on the corresponding region of the fragment Fab New. The region (in the V_L, near the V_L-C_L interface) contains two amino-acid differences but is embedded in a larger region containing many substitutions, insertions, and deletions.

Overview of Coupled Perturbation Procedure

The CPP is a specific method for calculating the detailed three-dimensional structure of a protein molecule which has a number of amino-acid substitutions relative to some previously determined "parent" protein structure. Before initiating the modeling procedure, a conformational energy search is performed, in the parent structure, about all of the dihedral angles of each side-chain to be replaced. Such a search should verify that the conformation of the residue which is observed in the crystal structure lies in the lowest energy well according to the conformaional energy equation,[11] demonstrating that the energy equations could calculate the parent structure correctly based on itself.

For each substitution to be introduced, the set of residues whose conformation is likely to be affected by the substitution (the dependency set) is identified. For each set of dependent residues, it is then necessary to find all of the low-energy conformations which will be used as starting points for energy refinement. To avoid the computational expense of performing a

simultaneous conformational search about all of the side-chains in the dependency set, restricted search-volume techniques are used to find the refinement starting points. Finally, each of these low-energy structures is energy refined. The relative conformational energies of the refined structures are compared and the lowest energy structure is taken as the model.

MATERIALS AND METHODS

All calculations were performed by using the Protein Conformational Analysis Package (PCAP) developed in this laboratory.[14] In all work described here, "unified atoms" are used; i.e., hydrogens are not explicitly included, an "atom" is a unit such as a methyl group, hydroxyl group, etc.

Energy Calculations

A semi-empirical potential energy function similar to that used in other programs[15-18] was used. The functional form of the potential is

$$E = E_b + E_a + E_p + E_t + E_v + E_h + E_e + E_w$$

where

$$E_b = \sum_i \frac{1}{2} k_{b_i} (b_i - b_{i_0})^2,$$

b_i = *bond length*, b_{i_0} = *optimum value*,

$$E_a = \sum_i \frac{1}{2} k_{a_i} (\theta_i - \theta_{i_0})^2 ,$$

θ_i = *bond angle*, θ_{i0} = *optimum value*,

$$E_p = \sum_i k_{p_i} [cos\ (m\phi_i) + 1],$$

ϕ_i = *single bond dihedral angle*,

$$E_t = \sum_i \frac{1}{2} k_{\tau_i} (\tau_i - \tau_{i_0})^2,$$

τ_i = *double bond dihedral angle*,

$$E_v = \sum_i \sum_j A r_{i,j}^{-12} + B r_{i,j}^{-6},$$

$r_{i,j}$ = *nonbonded atom distance*,

$$E_h = \sum_i \sum_j C r_{i,j}^{-12} + D r_{i,j}^{-10},$$

$r_{i,j}$ = *hydrogen bond distance*,

$$E_e = \sum_i \sum_j \frac{e_i e_j}{\epsilon r_{i,j}},$$

$$E_w = \sum_i \frac{1}{2} k_{w_i} \| \vec{x}_{c_i} - \vec{x}_{o_i} \|^2,$$

$\vec{x}_{c_i}$, $\vec{x}_{o_i}$ = *current, tether position for atom i.*

The E_w term (penalty term, reference 1) was used during refinement only—not for conformational searches or for comparing energies of different model structures. This term tethers the model to the parent structure and may also be used to avoid "edge effects" at the boundary of the refinement region. A cutoff distance of 9 Å was used for the electrostatic, Van der Waals, and hydrogen bond terms to decrease the computational size of the problem. Lennard-Jones parameters were calculated from the Slater-Kirkwood formula by using the polarizabilities, effective outer shell electron numbers, and Van der Waals radii of Gelin and Karplus.[15] Partial charges were taken from Momany et al.[16] where they were obtained by using the CNDO/2 method. A distance dependent dielectric constant of $\epsilon = 1.2\ r$ is used, where r is the interatomic distance. Hydrogen bond parameters were taken from Gelin and Karplus.[15]

Structure Alignments

Structures to be compared were aligned by using a two-step procedure. First, a general 3×3 transformation matrix was found that minimized the sum of the squares of the distances between equivalent atoms in the two structures. Then the procedure of Diamond[19] was used to separate this matrix into a true rotation and an expansion/contraction. The rotation matrix was then used for structural alignment.

Dependency Sets

The dependency set associated with an amino-acid substitution consists of all those residues which can come within Van der Waals contact distance of any conformation of either the original side-chain or its replacement. All such residues can easily be identified by performing conformational searches about all of the side-chain dihedral angles of both the original and the replacement amino acid.

When modeling a replacement from a smaller to a larger side-chain, the important dependent residues are those which are not in Van der Waals contact with the original side-chain but may contact (or even interfere with) the replacement side-chain. Conversely, when modeling a replacement from a larger to a smaller residue, the important dependent residues will be those which were within Van der Waals contact distance of the original residue, but for which contact could conceivably be broken when the residue is replaced.

Conformational Searches

Conformational searches were performed about side-chain dihedrals in steps of 15 degrees. A point in conformation space was considered a refinement starting point (RSP) if the conformational energy at that point was lower than the energy of any point for 2 steps (30 degrees) in every dimension of the space. Occasionally this criterion selects a structure which has a conformational energy as much as hundreds of

kcal/mol greater than other RSPs identified by the procedure. Such structures correspond to local minima in very unfavorable regions of conformation space and are rejected, i.e., are not used as starting points for refinement.

Restriction of the Search-Volume

Rather than performing a simultaneous conformational search about the side-chain dihedral angles of two interacting residues, the side-chain of one of the residues is temporarily removed past the beta-carbon (i.e., the residue is temporarily replaced with an alanine) while a conformational search is performed on the other residue. A conformational search can then be performed on the first residue while the second residue is held fixed in the low-energy conformations identified by the first search. Although this method does not search the entire conformation space defined by the side-chains of the residues, it does find at least one point in each energy-well of the conformational energy surface. Since such searches are always followed by energy refinement, finding one point in each energy-well is all that is necessary. The technique can easily be extended to problems involving more than two residues.

Energy Refinement

A conjugate gradients function minimization procedure[20] was used for all energy refinements. Residues within 8 Å of the replacement being modeled were included in the region in which atoms are allowed to move subject to the refinement procedure. Side-chains in this region were allowed to move freely while backbone atoms were subject to an 80 kcal/mol $Å^2$ penalty for movement.

RESULTS AND DISCUSSION

The immunoglobulin region being modeled in this experiment is located in the portion of the V_L closest to the C_L (Fig. 1). The immunoglobulin fragments Fab New and Fab Kol contain two amino-acid differences in this region. Alanine 83 in Fab New is a serine in Fab Kol while leucine 105 in Fab New is a valine in Fab Kol.

It was first necessary to verify that, in the Fab New crystal structure, these residues are in a conformational energy minimum according to the energy equation. A conformational energy search about the side-chain dihedral angles of leucine 105 (Fig. 2) confirms that this residue lies in a conformational energy minimum. It was not necessary to perform such a search for alanine 83 because this residue has no side-chain dihedral degrees of freedom. These results indicate that the conformational energy equation agrees well with the crystallographic data in this region of the Fab New structure.

To determine the residues which make up the dependency set for the alanine 83 → serine substitution, a conformational search was performed on a serine at position 83. Tyrosine 85 and leucine 38 were identified as potential contact residues. These residues along with alanine/serine 83 were defined as the dependency set for this substitution.

For the leucine 105 → valine substitution, the residues valine 107, tyrosine 85, and isoleucine 74 were identified as residues which could potentially lose Van der Waals contacts due to the substitution. These residues along with leucine/valine 105 were defined as the dependency set for this substitution.

The only residue common to the two dependency sets is tyrosine 85. A conformational search performed on tyrosine 85 against all combinations of alanine/serine at position 83 and leucine/valine at position 105 indicated that there is only one conformational energy-well of low energy for tyrosine 85. The facts that tyrosine 85 is the only residue common to both dependency sets and that this residue shows only one reasonable conformational energy-well were taken to indicate that the two substitutions could reasonably be modeled sequentially rather than concurrently.

The alanine 83 → serine substitution was modeled first: Using the restricted search-volume technique, conformational searches were performed, with only backbone plus a beta-methyl group at position 83, on leucine 38 and tyrosine 85 to locate wells of low conformational energy for these residues. Since leucine 38 and tyrosine 85 are themselves independent, these conformational searches were also done sequentially rather than concurrently. Only one energetically acceptable conformation was found for the tyrosine and four were found for leucine. The four structures corresponding to these results were built and, in each case, a conformational search was performed on a serine at position 83. Three of the four starting structures had three acceptable conformations for the serine; the fourth had only two. The 11 resultant structures were built and each was subjected to 80 cycles of conjugate gradients energy refinement. The side-chain dihedral angles (before and after refinement) and the relative conformational energies of the 11 model structures are shown in Table I. It is apparent from these data that, in this case, the search procedure has sampled conformation space more finely than one point per energy-well. Although the structures have not been refined to completion,** it is apparent that the structures A00 and A01 are refining to the same point as are the structures A10 and A11, the structures A20 and A21, and the structures A30, A31, and A32. The lowest energy structure (A01.R) after this initial refinement was used as a starting point for the introduction of the second substitution.

Conformational searches were performed on isoleucine 74, tyrosine 85, and valine 107 while the side-chain at position 105 was reduced to a beta-methyl group in accordance with the restricted search volume technique. Twelve energy-wells of reasonable conformational energy were found for these residues.

**Since further refinement is done after the introduction of the leucine 105 → valine substitution, refinement was not carried to completion at this point.

TABLE I. Side-Chain Dihedral Angles and Relative Conformational Energies of Structures Following the Alanine 83 → Serine Substitution**

	Conformational energy	ser 83 χ_1*	leu 38 χ_1*	leu 38 χ_2*	tyr 85 χ_1*	tyr 85 χ_2*
A00.R	0.64	15.0/40.4	291.9/278.5	81.4/94.7	292.5/290.1	105.6/83.6
A01.R	0.00	60.0/43.9	291.9/278.2	81.4/94.5	292.5/290.2	105.6/84.7
A10.R	7.40	60.0/46.2	51.9/44.0	51.4/80.2	292.5/289.8	105.6/84.6
A11.R	7.60	15.0/43.9	51.4/43.9	51.4/80.2	292.5/289.7	105.6/83.4
A12.R	5.41	285.0/280.9	51.9/44.1	51.4/82.1	292.5/289.4	105.6/81.9
A20.R	6.03	60.0/46.6	51.9/52.8	291.4/281.7	292.5/290.0	105.6/84.5
A21.R	3.83	15.0/43.7	51.9/52.7	291.4/281.7	292.5/289.9	105.6/83.3
A22.R	4.63	285.0/281.8	51.9/53.0	291.4/282.3	292.5/289.5	105.6/82.1
A30.R	6.20	90.0/45.0	218.7/358.4	156.4/156.3	292.5/289.8	105.6/83.6
A31.R	5.01	60.0/44.6	218.7/356.3	156.4/157.2	292.5/289.7	105.6/83.5
A32.R	5.77	15.0/42.7	218.7/357.6	156.4/156.6	292.5/289.6	105.6/83.2

*Before/after refinement.
**Conformational energies are relative to the lowest energy structure (A01.R) and are expressed in kcal/mol. Side-chain dihedral angle values are in degrees.

TABLE II. Side-Chain Dihedral Angles and Relative Conformational Energies of Structures Following the Leucine 105 → Valine Substitution**

	Conformational energy	val 105 χ_1*	ile 74 χ_1*	ile 74 χ_2*	val 107 χ_1*
V000.R	4.91	75.0/71.7	167.1/165.9	43.2/45.6	180.9/177.8
V001.R	3.87	285.0/285.8	167.1/166.0	43.2/45.6	180.9/177.9
V010.R	1.41	75.0/71.8	167.1/154.1	238.2/228.6	180.9/177.7
V011.R	0.00	285.0/285.7	167.1/154.0	238.2/228.7	180.9/177.9
V020.R	3.75	75.0/71.7	77.1/96.9	163.2/178.8	180.9/177.8
V021.R	1.24	285.0/285.6	77.1/96.5	163.2/177.3	180.9/177.9
V030.R	1.64	75.0/71.5	302.1/302.8	103.2/98.4	180.9/177.7
V031.R	1.60	285.0/285.8	302.1/302.7	103.2/89.5	180.9/177.8
V100.R	6.97	75.0/71.5	167.1/165.9	43.2/45.6	45.9/66.1
V101.R	6.23	285.0/286.1	167.1/166.0	43.2/45.6	45.9/66.3
V110.R	5.20	75.0/71.7	167.1/154.0	238.2/228.3	45.9/66.2
V111.R	2.29	285.0/286.0	167.1/154.0	238.2/228.4	45.9/66.3
V120.R	7.18	75.0/71.7	77.1/97.0	163.2/179.3	45.9/66.1
V121.R	2.77	285.0/285.9	77.1/97.0	163.2/179.4	45.9/66.2
V130.R	5.25	75.0/71.4	302.1/302.7	103.2/89.4	45.9/66.0
V131.R	3.97	285.0/286.1	302.1/302.8	103.2/89.3	45.9/66.2
V200.R	5.44	75.0/71.7	167.1/165.9	43.2/45.6	285.9/310.2
V201.R	4.43	285.0/285.8	167.1/165.9	43.2/45.6	285.9/310.4
V210.R	5.74	75.0/71.8	167.1/154.1	238.2/229.0	285.9/310.1
V211.R	4.24	285.0/285.6	167.1/154.1	238.2/229.4	285.9/310.3
V220.R	4.59	75.0/71.7	77.1/96.9	163.2/179.2	285.9/310.2
V221.R	1.82	285.0/285.6	77.1/96.3	163.2/176.7	285.9/310.3
V230.R	4.24	75.0/71.5	302.1/302.7	103.2/89.4	285.9/310.2
V231.R	4.43	285.0/285.7	302.1/302.7	103.2/89.3	285.9/310.3

*Before/after refinement.
**Conformational energies are relative to the lowest energy structure (V011.R) and are expressed in kcal/mol. Side-chain dihedral angle values are in degrees.

Structures corresponding to these 12 energy-wells were built and, in each case, a conformational search was performed on a valine at position 105. In each case, two reasonable conformational wells were found for the valine. This gave 24 starting structures for energy refinement. Each was subjected to 80 cycles of conjugate gradients refinement. The results of these refinements (side-chain dihedral angles and conformational energies of the 24 model structures) are shown in Table II. These structures have been refined to near completion.‡ There are slight differences in side-chain dihedral angles (never greater than 2.5

‡During the final energy minimization of this modeling problem, we have found that 80 cycles of conjugate gradients refinement will result in a structure which has an all-atom rms movement of less than 0.001 Å during the last refinement cycle.

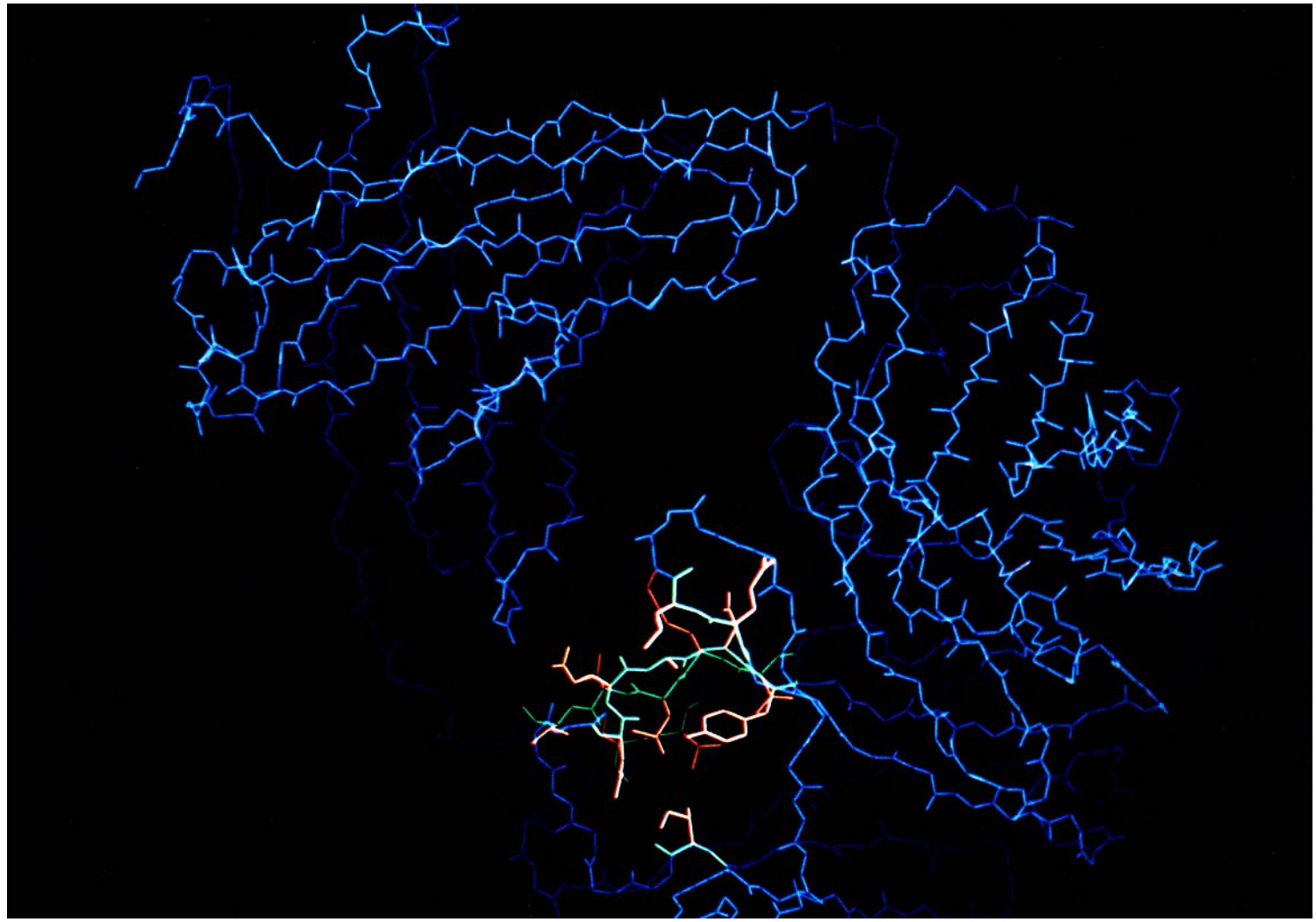

Figure 1a.

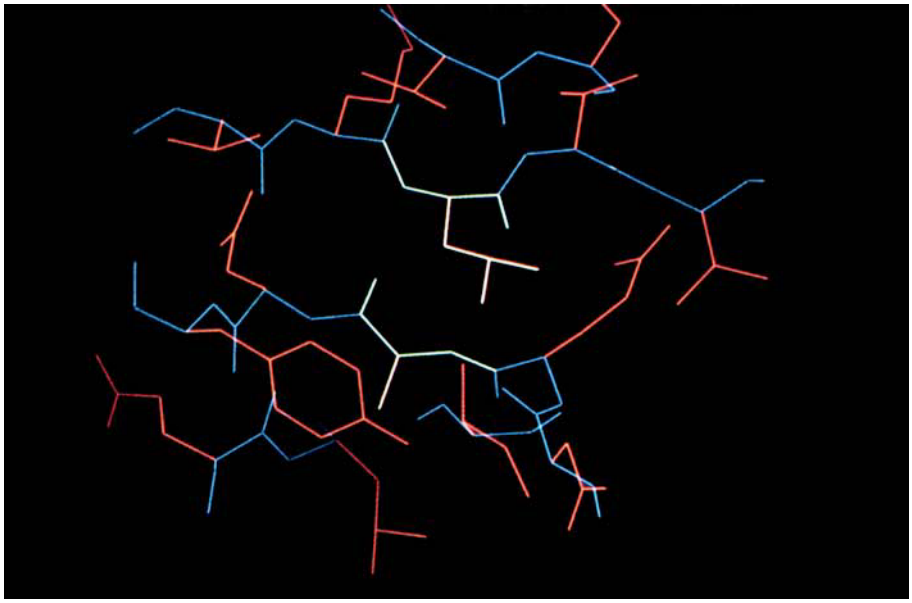

Figure 1b.

Fig. 1. The region of the immunoglobulin fragment Fab New which was used as the basis for the model lies in the V_L, but away from the antigen combining site. a: The location of this region (green backbone/orange side-chains) in relation to the entire Fab fragment (blue, backbone only). b: Detail of the region. The highlighted residues (green backbone/white side-chains) are alanine 83 and leucine 105. The sequence of Fab Kol shows a serine and a valine, respectively, at these positions.

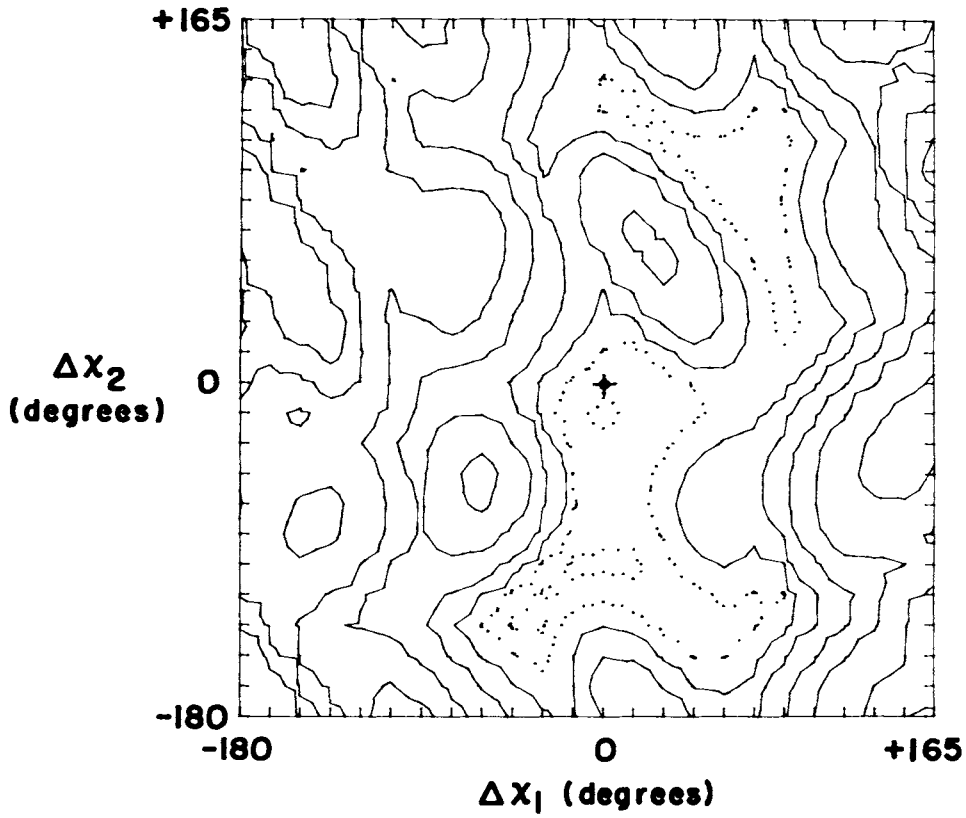

Fig. 2. Conformational energy search about the side-chain dihedral angles of leucine 105 in Fab New. The lowest energy contours are shown with dotted lines. The point at the value (0,0) of ($\Delta\chi_1$, $\Delta\chi_2$) corresponds to the conformation of the residue observed in the crystal structure. This conformation lies on the edge of the lowest energy-well of the surface.

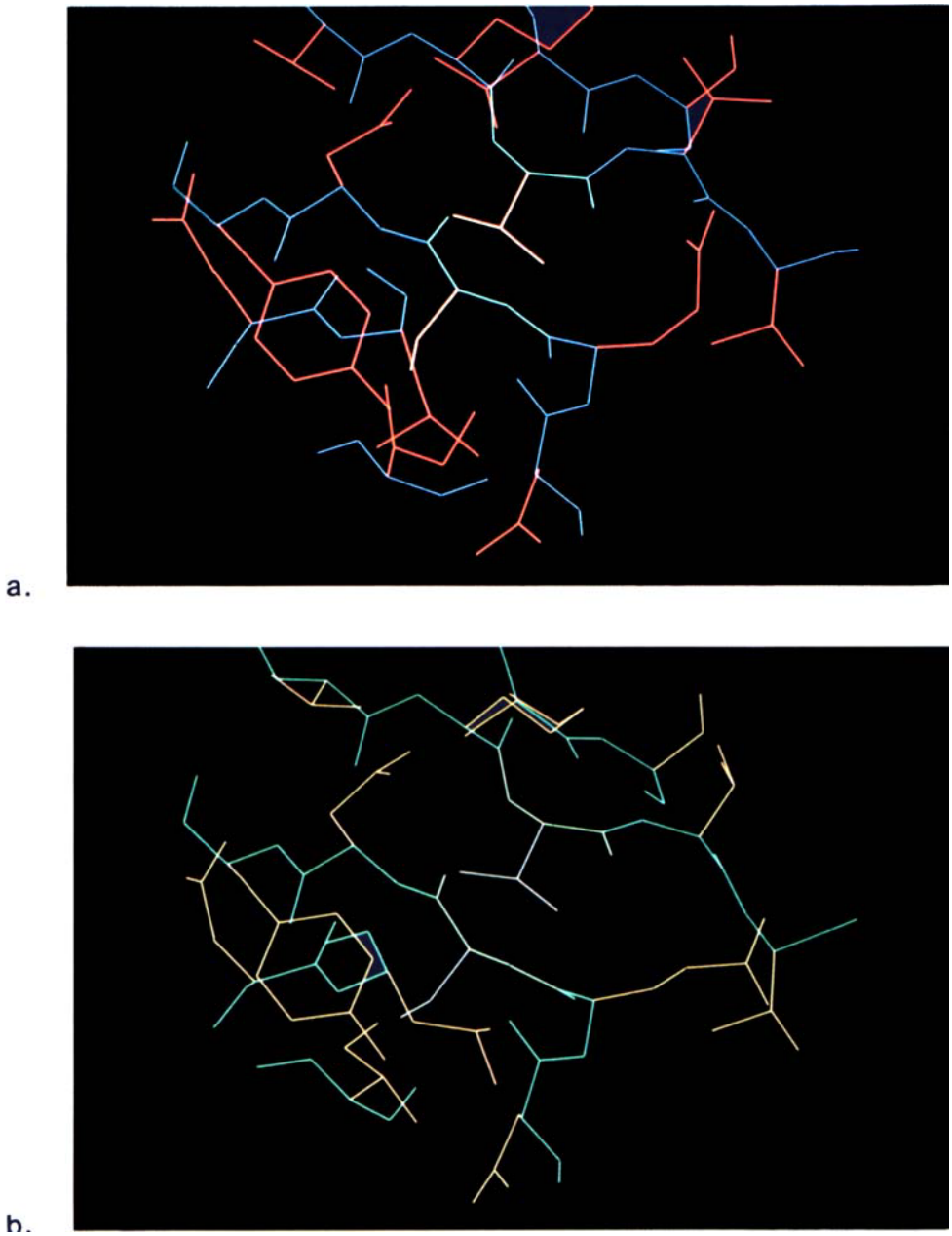

Fig. 3. Comparison of the model V011.R to the structure of Fab Kol. a, b: Detail of V011.R and Fab Kol, respectively. c: The superposition of the two backbones (V011.R is blue, Fab Kol green) and the side-chains of serine 83 and valine 105 (V011.R orange, Fab Kol white).

degrees) between residues that are in the same conformation in different models. This may be due to the effect of structural differences in other areas of the models or to the fact that, because of the penalty term in the energy equation, structures which start with even slightly different backbone conformations will not refine to exactly the same point. The "best" model structure (V011.R) has a conformational energy which is lower than any of the other structures by at least 1.2 kcal/mol. This structure was taken as the model of Fab Kol in this region.

A comparison of the model with Fab Kol (Fig. 3, Table III) shows that the calculated side-chain dihedral angles of serine 83 and valine 105 differ from those observed in Fab Kol by 15.7 and 9.1 degrees, respectively. Since these residues have three conformational minima, roughly 120 degrees apart, they have certainly been placed in the correct conforma-

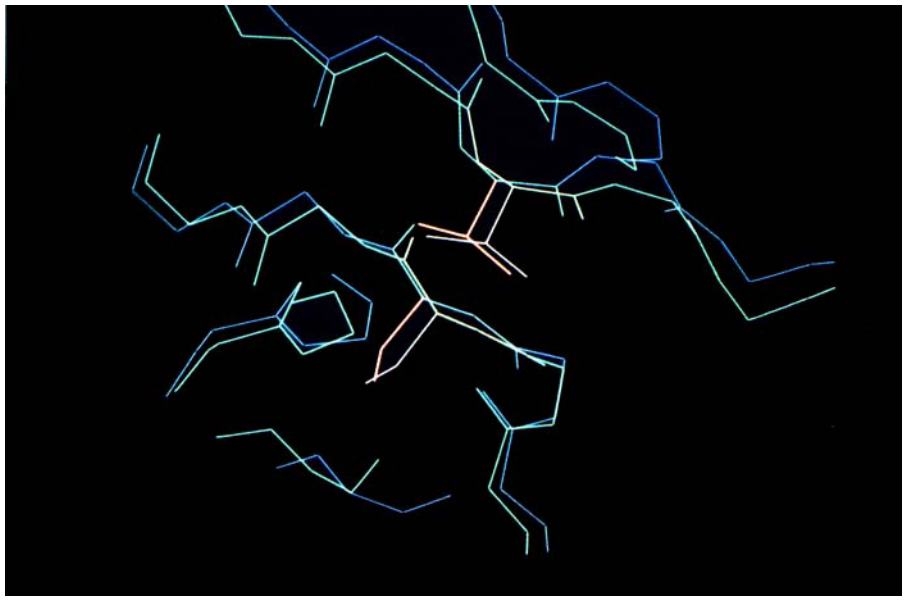

Fig. 3. continued.

TABLE III. Comparision of the Side-Chain Dihedral Angles of Serine 83 and Valine 105 in the Structures Fab Kol, Refined Fab Kol, and V011.R

Structure	ser 83 χ_1 (degrees)	val 105 χ_1 (degrees)
Fab Kol	59.6	294.8
Fab Kol (refined)	41.3	290.0
V011.R	43.9	285.7

tional well. Note that the Fab Kol structure has not been refined with our energy program in this region. If this is done, the difference between the model and the refined Fab Kol structure for the two side-chain dihedrals is only 2.6 and 4.3 degrees, respectively (Table III).

A very effective way to compare structures is to calculate rms differences in atomic positions on a residue by residue basis. All backbone as well as side-chain atoms must be included in such calculations so that the alignment of side-chains cannot be optimized at the expense of the main chain conformation. A comparison of the model structure V011.R to Fab Kol by this method (Table IV) shows that serine 83 and valine 105 have been placed correctly (0.12 Å and 0.16 Å rms error, respectively). In addition, the procedure has improved the conformation (relative to the starting conformation in Fab New) of three of the four other residues in the combined dependency sets for the two substitutions. No residue in the entire dependent region has a conformation that is further from Fab Kol in V011.R than in the starting Fab New structure.

There is one residue, isoleucine 74, which did not improve to a correct conformation as a result of the modeling procedure. A comparison of the structure Fab Kol with either Fab New or V011.R reveals why this is the case (Fig. 4). In Fab Kol, isoleucine 74 is the last residue in a strand of beta-pleated sheet. In Fab New, the sheet ends at residue 73, and isoleucine

TABLE IV. Root Mean Square Differences in Atomic Positions for Individual Amino Acids‡

Structures compared	ser 83	val 105	leu 38	ile 74	tyr 85	val 107
Fab New vs. Fab Kol	**	**	>2.*	>2.*	.18	.24
Refined New † vs. Kol	**	**	>2.*	>2.*	.17	.16
V011.R vs. Fab Kol	.12	.16	.77	>2.*	.14	.14

*For these residues, the best fit requires inversion.
**These comparisions could not be made because Fab New and Fab Kol have different residues at positions 83 and 105.
†This comparison demonstrates that the results are due to the coupled perturbation procedure (CPP) as a whole and not simply the effects of energy refinement.
‡All values in angstroms.

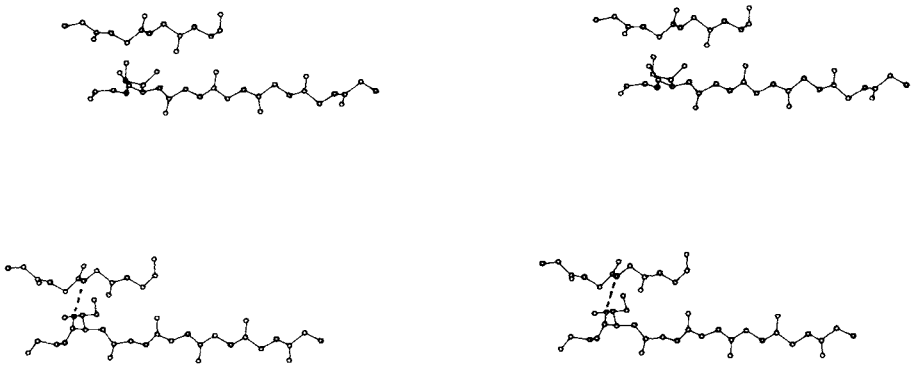

Fig. 4. Stereo drawing showing part of the beta-sheet structure of V011.R (top) and Fab Kol (bottom). The side-chain of isoleucine 74 is shown. It can be seen that this residue is part of the sheet structure in Fab Kol but is the first residue of a loop in V011.R. Note particularly that the hydrogen bond (dashed line) which is formed between the carbonyl oxygen of isoleucine 74 and the adjacent strand of sheet in Fab Kol is absent in V011.R.

74 is the first residue in a loop. Because of this difference in backbone conformation, isoleucine 74 in Fab New (or in V011.R) cannot adopt the side-chain dihedrals observed in Fab Kol without causing unfavorable Van der Waals overlap with tyrosine 85 (this was confirmed by model building). It would not be possible for the modeling procedure to calculate such changes in main chain conformation without incorporating a conformational search on phi and psi into the modeling procedure.

Since the structures of Fab New and Fab Kol were determined from different crystal forms in different laboratories, it is relevant to ask how the relative accuracy of these structure determinations affects the interpretation of our results. To briefly address this question, four regions (of comparable size to the region in this modeling problem) were chosen in which Fab New and Fab Kol show as little sequence difference as possible (zero or one sequence difference). In each case, the corresponding regions in Fab New and Fab Kol were aligned and rms C_α–C_α distances were calculated. Values of 0.32, 0.47, 0.53, and 0.37 Å were obtained. The final coordinates of a good model should agree with the experimental coordinates to within about 0.5 Å (the limit set by the differences between the two experimental structures in regions of identical sequence). A calculation comparing V011.R to Fab Kol in the modeled region gives an rms C_α–C_α distance of 0.43 Å, satisfying this criterion.

Comparison with other modeling methods: It is important to ask whether the additional computational complexity of the coupled perturbation procedure (relative to the maximum overlap or minimum perturbation procedures) is justified. To address this question it was necessary to determine whether an accurate model of Fab Kol in this region could have been calculated by one of the other procedures. Model structures representing the MOP and MPP methods were built as described below.

A maximum overlap model was built by starting with the structure of Fab New and replacing the residues alanine 83 and leucine 105 with a serine and a valine, respectively. The serine was built such that the N, C_α, C_β, C′, and backbone O were in the same positions as in the alanine. The OH was built so as to make χ_1 equal 180 degrees. The valine was also built so as to have maximum overlap with the residue (leucine 105) which it replaced. There was a choice as to which valine C_γ to build in the position previously occupied by the leucine C_γ. Both conformations were built and the conformation with lower conformational energy was taken as the starting point for energy refinement. This structure was subjected to 160 cycles of conjugate gradients energy refinement. The region of the structure allowed to move under the influence of the refinement was the union of the two regions used in the development of V011.R. The structure resulting from this refinement (FTEST) was taken as the MOP model of Fab Kol in this region.

A minimum perturbation model was built by replacing the residues alanine 83 and leucine 105 in Fab New with a serine and a valine, respectively. Conformational searches were then performed about the side-chain dihedral angles of the introduced residues. Three acceptable conformations were found for the serine and two for the valine. Six structures were built in the conformations suggested by these results and each was used as a starting point for conjugate gradients energy refinement. The conformational energies after refinement of the six structures were compared. The lowest energy structure (S21.R) was taken as the MPP model of Fab Kol in this region.

TABLE V. Root Mean Square Differences in Atomic Positions for Individual Amino Acids**

Structures compared	ser 83	val 105	leu 38	ile 74	tyr 85	val 107
FTEST. vs. Fab Kol	.79	1.25	>2.*	>2.*	.28	.14
S21.R vs. Fab Kol	.74	.14	>2.*	>2.*	.10	.15
V011.R vs. Fab Kol	.12	.16	.77	>2.*	.14	.14

*For these residues, the best fit requires inversion.
**All values in angstroms.

TABLE VI. Comparison of the Side-Chain Dihedral Angles of Serine 83 and Valine 105 in the Structures Fab Kol, FTEST, S21.R, and V011.R

Structure	Ser 83 χ_1 (degrees)	val 105 χ_1 (degrees)
Fab Kol	59.6	294.8
FTEST	268.4	180.6
S21.R	282.9	290.4
V011.R	43.9	285.7

A comparison of the three model structures (Tables V,VI) shows that the CPP model V011.R is closer to the structure of Fab Kol than either the MOP model FTEST or the MPP model S21.R. Only the CPP procedure has placed both serine 83 and valine 105 in the correct conformational energy-well. The MPP model has the valine correct while the MOP model has neither residue correct. CPP has also done a better job positioning the neighboring residue leucine 38 than either of the other methods. The residue isoleucine 74 was positioned incorrectly by all three methods.

The relative accuracy of the three model structures (Tables V,VI) is consistent with the idea that greater modeling accuracy is achieved with procedures which include more information about the conformation of both replaced and neighboring residues. This idea could, of course, be carried considerably beyond what has been discussed here. If including neighboring residues in the conformational space searched improves the procedure, wouldn't including the neighbors of the neighboring residues improve it further? This idea could hypothetically be extended to the point where the entire molecule is included in the conformational search space. Although procedures more complex than those described above have not actually been tried, there are reasons to believe that extending the complexity of the procedure beyond that used to calculate V011.R would not be a necessary part of a reliable modeling procedure. Implicit in all modeling procedures of this type is the assumption that the conformational changes resulting from small sequence changes are local in nature. If a given sequence change had structural effects which propagated through several shells of dependent residues, the programs would likely fail to calculate the structural changes regardless of the modeling procedure used. Finally, the procedure used to calculate V011.R has been shown to give good results at a reasonable level of computational complexity. Extending the procedure to include another level of dependency would increase the time and complexity of the procedure substantially.

SUMMARY AND CONCLUSIONS

The coupled perturbation procedure (CPP) has been introduced as a means of calculating the three-dimensional structural changes in protein molecules which result from amino-acid sequence substitutions. CPP uses a combination of conformational search and energy refinement techniques to find the conformation of a protein molecule which corresponds to a minimum in a semi-empirical energy equation evaluated over some region of conformational space. This conformation of the molecule is taken as a model of the structure in this region.

In order to assess the ability of CPP to calculate accurate models for proteins of unknown structure, it was first necessary to test the ability of the procedure to calculate the three-dimensional structure of a molecule with known structure. The calculation of the structure of a region of the immunoglobulin fragment Fab Kol based on the analogous region of the immunoglobulin fragment Fab New was chosen as a test calculation. This problem constitutes a fairly rigorous test of the procedure for two reasons. First, the crystallographic determinations of the structures of Fab Kol[13] and Fab New[12] were done from different crystal forms in different laboratories using different methods and refinement programs. Second, the region in which the calculation was performed is embedded in a larger region in which the two proteins contain numerous differences. The success of CPP in handling this problem gives reason to believe that the procedure could be successfully employed in "simpler" modeling problems such as calculating the structure of proteins produced by site-directed mutagenesis techniques.

The procedure has worked well although the effect of solvent has not been explicitly included in the conformational energy equation. This is likely due to the fact that only small differences between proteins

in the folded state have been considered. Under these conditions, the small solvation differences are reasonably modeled by the use of a distance-dependent dielectric constant.

In more elaborate modeling studies, there shall likely arise a need for the explicit inclusion of solvent in the energy equation. A straightforward method for the calculation of the contribution of solvent to the conformational energy makes use of the solvation energies of Eisenberg and McLachlan,[21] using the solvent accessible surface area algorithm of Richmond.[22]

The procedure cannot be expected to work when applied to problems where the amino-acid substitution(s) results in structural changes which are not local in nature. Reasons for this include the small volume of conformation space which is actually searched (only one shell of dependent residues and no backbone dihedrals), the neglect of solvent, and the neglect of entropy. The substitutions which are most likely to have far-reaching structural effects are those involving internal residues. A single amino-acid change in the interior of a protein can have large consequences. [23] It is not realistic to apply this type of modeling procedure to any problem which involves substitutions that are likely to have nonlocal effects on three-dimensional structure.

There have been a number of previous modeling studies which have addressed questions in the immunoglobulin family.[10,24–30] While many of these studies have done an excellent job of providing approximate models, in no case has the accuracy of the procedure been confirmed by correctly calculating the atomic coordinates of a known structure.

The success of CPP in this study is an encouraging step toward the goal of calculating accurate structures for naturally occurring molecules in protein families in general and in the immunoglobulin family in particular. Understanding how the amino acids in immunoglobin hypervariable regions determine the detailed conformation of the antibody combining site is one question which might be approached by these techniques.[31] Insight can be gained by modeling, for example, all the immunoglobulins that bind a given hapten,[32] all the somatic variants that appear during the maturation of an immune response[33] or hybrid molecules that combine the framework residues from one immunoglobulin with the complementarity-determining residues of another.[34] In addition, there are a growing number of proteins in the immunoglobulin family for which amino-acid sequence data are available, but for which crystal structures have not been determined. These techniques may prove useful in addressing structural questions pertaining to these molecules.

ACKNOWLEDGMENTS

The authors are grateful to Jeremy Berg, Sudhir Bedarkar, Xavier Ysern, and Claudette Snow for reading the manuscript and particularly to Ed Lattman for many useful comments and discussion regarding the work.

This work was supported by grant AI20293 from the National Institutes of Health. The Interactive Graphics Facility of the Department of Biophysics was established and is maintained by NIH and NSF grants and by a gift from the Richard-King Mellon Foundation.

REFERENCES

1. Swenson, M.K., Burgess, A.W., Scheraga, H.A. Conformational analysis of polypeptides: Application to homologous proteins. In: "Frontiers in Physicochemical Biology." Pulman, B. ed. New York: Academic Press 1978: 115–142.
2. Warme, P.K., Momany, F.A., Rumball, S.V., Tuttle, R.W., Scheraga, H.A. Computation of structures of homologous proteins. α-Lactalbumin from lysozyme. Biochemistry 13:768–782, 1974.
3. Greer, J. Comparative model-building of the mammalian serine proteases. J. Mol. Biol. 153:1027–1042, 1981.
4. Blundell, T.L., Bedarkar, S., Rinderknecht, E., Humbel, R.E. Insulin-like growth factor: A model for tertiary structure accounting for immunoreactivity and receptor binding. Proc. Natl. Acad. Sci. USA 75:180–184, 1978.
5. Blundell, T., Sibanda, B.L., Pearl, L. Three-dimensional structure, specificity and catalytic mechanism of renin. Nature 304:273–275, 1983.
6. Greer, J. Model structure for the inflammatory protein C5a. Science 228:1055–1060, 1985.
7. Palmer, K.A., Scheraga, H.A., Riordan, J.F., Vallee, B.L. A preliminary three-dimensional structure of angiogenin. Proc. Natl. Acad. Sci. USA 83:1965–1969, 1986.
8. Read, R.J., Brayer, G.D., Jurasek, L., James, M.N.G. Critical evaluation of comparative model building of *Streptomyces griseus* trypsin. Biochemistry 23:6570–6575, 1984.
9. Jurasek, L., Olafson, R.W., Johnson, P., Smillie, L.B. Miami Winter Symp. 11:93–123, 1976.
10. de la Paz, P., Sutton, B.J., Darsley, M.J., Rees, A.R. Modeling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. EMBO J. 5:415–425, 1986.
11. Shih, H. H-L., Brady, J., Karplus, M. Structure of proteins with single-site mutations: A minimum perturbation approach. Proc. Natl. Acad. Sci. USA 82:1697–1700, 1985.
12. Saul, F.A., Amzel, L.M., Poljak, R.J. Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin New at 2.0 Å resolution. J. Biol. Chem. 253:585–597, 1978.
13. Matsushima, M., Marquart, M., Jones, T.A., Colman, P.M., Bartels, K., Huber, R., Palm, W. Crystal structure of the human Fab fragment Kol and its comparison with the intact Kol molecule. J. Mol. Biol. 121:441–459, 1977.
14. Snow, Mark E. An Approach to Calculating the Structural Effects of Mutations on the Surface Residues of Proteins: Theory and Application to the System of Immunoglobulins. PhD Dissertation. Johns Hopkins University, 1986.
15. Gelin, B.R., Karplus, M. Side-chain torsional potentials: Effect of dipeptide, protein and solvent environment. Biochemistry 18:1256–1268, 1979.
16. Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A. Energy parameters in polypeptides VII. Geometric parameters, partial atomic charges, non-bonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem. 79:2361–2381, 1975.
17. Weiner, P., Kollman, P. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. J. Comput. Chem. 2:287–303, 1981.
18. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, V.C., Ghio, C., Alagona, G., Profeta, S., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
19. Diamond, R. On the comparison of conformations using linear and quadratic transformations. Acta Cryst. A32:1–10, 1976.

20. Fletcher, R., Reeves, C.M. Function minimization by conjugate gradients. Computer J. 7:149–154, 1964.
21. Eisenberg, D., McLachlan, A.D. Solvation energy in protein folding and binding. Nature 319:199–203, 1986.
22. Richmond, T. Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. J. Mol. Biol. 178:63–89, 1984.
23. Dijkstra, B.W., Weijer, W.J., Wierenga, R.K. Polypeptide chains with similar amino-acid sequences but a distinctly different conformation. FEBS Lett. 164:25–27, 1983.
24. Feldman, R.J., Potter, M., Glandemans, C.P.J. A hypothetical space-filling model of the V-regions of the galactan-binding myeloma immunoglobulin J539. Mol. Immun. 18:683–698, 1981.
25. Padlan, E.A., Davies, D.R., Pecht, I., Givol, D., Wright, C. Model building studies of antigen binding sites. The antigen binding site of MOPC315. Cold Spring Harbor Symp. Quant. Biol. 4:627–637, 1977.
26. Poljak, R.J., Amzel, L.M., Chen, B.C., Phizackerley, R.P., Saul, F. The three-dimensional structure of the Fab fragment of a human myeloma immunoglobulin at 2.0 Å resolution. Proc. Natl. Acad. Sci. U.S.A. 71:3440–3444, 1974.
27. Dwek, R.A., Wain-Hobson, S., Dower, S., Gettings, P., Sutton, B., Perkins, S.J., Givol, D. Structure of an antibody combining site by magnetic resonance. Nature 266:31–37, 1977.
28. Stanford, J.M., Wu, T.T. A predictive method for determining possible three-dimensional foldings of immunoglobulin backbones around antibody combing sites. J. Theor. Biol. 88:421–439, 1981.
29. Travers, P., Blundell, T.L., Sternberg, M.J.E., Bodmer, W.F. Structural and evolutionary analysis of HLA-D region products. Nature 310:235–238, 1984.
30. Novotny, J., Tonegawa, S., Saito, H., Kranz, D.M., Eisen, H.M. Secondary, tertiary and quaternary structure of T-cell-specific immunoglobulin-like polypeptide chains. Proc. Natl. Acad. Sci. U.S.A. 83:742–746, 1986.
31. Kabat, E.A. Antibody diversity versus antibody complementarity. Pharmacol. Rev. 34:23–38, 1982.
32. Padlan, E.A., Cohen, G.H., Davies, D.R. On the specificity of antibody/antigen interactions: Phosphoryl choline binding to McPC603 and the correlation of three-dimensional structure and sequence data. Ann. Inst. Pasteur Immunol. 136C:271–276, 1985.
33. Griffiths, G.M., Berek, C., Kaartinen, M., Milstein, C. Somatic mutation and the maturation of immune response to 2-phenyl oxazolone. Nature 312:271–275, 1984.
34. Jones, P.T., Dear, P.H., Foote, J., Neuberger, M.S., Winter, G. Replacing the complementarity determining regions in a human antibody with those from a mouse. Nature 321:522–525, 1986.
35. Chothia, C., Lesk, A.M., Levitt, M., Amit, A.G., Mariuzza, R.A., Phillips, S.E.V., Poljak, R.J. The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. Science 233:755–757, 1986.
36. 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 233:747–754, 1986.

NOTE ADDED IN PROOF

Chothia et al.[35] recently reported a comparison of a model of the six hypervariable regions of the anti-lysozyme immunoglobulin D1.3 with the structure reported by Poljak and co-workers.[36] The method used for modeling was based on carefully finding initial conformations for refinement utilizing the available immunoglobulin structures. The model presented correctly predicts many of the features of four of the six hypervariable regions while not focusing on the detail of individual side-chain conformations. The excellent approach of Chothia et al. is complementary to approaches such as CPP which could be used to extend the reliability and accuracy of the modeling procedure.