

## Predicting Antibody Hypervariable Loop Conformations II: Minimization and Molecular Dynamics Studies of MCPC603 From Many Randomly Generated Loop Conformations

R.M. Fine,<sup>1,2</sup> H. Wang,<sup>1</sup> P.S. Shenkin,<sup>1</sup> D.L. Yarmush,<sup>1</sup> and C. Levinthal<sup>1</sup>

<sup>1</sup>Department of Biological Sciences, Columbia University, New York, NY 10027; <sup>2</sup>Department of Structural Biology, Brookhaven National Laboratory, Upton, NY 11973

**ABSTRACT** We describe a method for predicting the conformations of loops in proteins and its application to four of the complementarity determining regions [CDRs] in the crystallographically determined structure of MCPC603. The method is based on the generation of a large number of randomly generated conformations for the backbone of the loop being studied, followed by either minimization or molecular dynamics followed by minimization starting from these random structures. The details of the algorithm for the generation of the loops are presented in the first paper in this series (Shenkin et al. [submitted]). The results of minimization and molecular dynamics applied to these loops is presented here. For the two shortest CDRs studied (H1 and L2, which are five and seven amino acids long), minimizations and dynamics simulations which ignore interactions of the loop amino acids beyond the carbon beta replicate the conformation of the crystal structure closely. This suggests that these loops fold independently of sequence variation. For the third CDR (L3, which is nine amino acids), those portions of the CDR near its base which are hydrogen bonded to framework are well replicated by our procedures, but the top of the loop shows significant conformational variability. This variability persists when side chain interactions for the MCPC603 sequence are included. For a fourth CDR (H3, which is 11 amino acids long), new low-energy backbone conformations are found; however, only those which are close to the crystal are compatible with the sequence when side chain interactions are taken into account. Results from minimization and dynamics on single CDRs with all other CDRs removed are presented. These allow us to explore the extent to which individual CDR conformations are determined by interactions with framework only.

**Key words:** antibodies, immunoglobulins, conformation prediction, energy minimization, random starting conformations

### INTRODUCTION

Antibody molecules comprise a family of proteins whose structural and functional diversity arise from sequence diversity in only about 10% of the mole-

cule.<sup>1</sup> Several hundred sequences of antibody light and heavy chains are known,<sup>2</sup> and at the present time, seven crystal structures have been reported at atomic resolution.<sup>3-9</sup> Careful comparison of these crystal structures has confirmed the early conjecture based on sequence analysis that those portions of the molecule conserved in sequence are conserved in structure: the complementarity determining regions (CDRs) defined by Wu and Kabat<sup>1</sup> on the basis of sequence variability have been found to be flanked by regions whose three-dimensional structure is conserved by virtue of participation in the hydrogen bonding pattern which forms the beta-barrel framework of the antibody molecule or by virtue of participation in conserved contacts between the light and heavy chain variable domains of the molecule.<sup>10-15</sup>

Antibody molecules thus form a logical target for homology model building studies, and several have been attempted.<sup>16-22</sup> In most of these studies, those portions of the molecule which fall outside of the CDRs of Kabat et al. have been assumed known. Modeling of the CDRs has then proceeded either by careful selection of backbone conformations for the CDRs from known structures or by geometric loop-building techniques. Recently, de la Paz et al. have attempted the modeling of monoclonal antibodies raised in their laboratory to hen egg white lysozyme based on the known structure of Fab NEW.<sup>20,21</sup> Chothia et al. have attempted the modeling of a subsequently solved crystal structure of an antigen-antibody complex between lysozyme and the Fab fragment of the monoclonal antibody D1.3.<sup>22</sup> Both sets of authors rely heavily on selection of possible loop conformations from those observed in solved crystals.

In this paper, we present a method for searching for the native conformations of the CDRs which is based on two assumptions. The first of these is that the native conformation of the CDR loop in question will be at least a local energy minimum for that portion

---

Received November 18, 1986; accepted January 9, 1987.

Address reprint requests to Cyrus Levinthal, Department of Biological Science, Columbia University, New York, NY 10027.

of the molecule. The second is that we can find this minimum by generating a large number of random conformations of the backbone of the CDR in question, subjecting these conformation to energy minimization or to molecular dynamics followed by energy minimization, and then selecting the lowest energy conformations from among those found for further characterization of side chain conformations. The random structures are generated in such a manner that they all join to the rest of the molecule with correct geometry. The first paper in this series describes in detail<sup>23</sup> the method for the generation of the random structures. In this paper, we present the results of minimizations from random backbone conformation of four CDRs in the known structure MCPC603.<sup>3,3a</sup> Applying our approach to a known structure allows us to test our assumptions, and to gain insight into the question of how many initial conformations are needed to find a particular CDR geometry with the use of energy minimization and molecular dynamics. In later papers we will apply the same technique to CDRs in antibody molecules whose primary sequence is known, but whose conformation is undetermined.

Our results for MCPC603 are mixed: for three of the four CDR loops examined, minimizations from between 100 and 1,000 starting conformations are sufficient to isolate either a single or several spatially similar low-energy minima which lie close to the native crystal conformation. For the fourth loop, although structures close to the native were found, additional conformations of the backbone atoms were also isolated that were lower in energy according to several force fields. The fifth and sixth loops of MCPC603, which are the longest, were found to be amenable to our technique for the generation of initial starting conformations but required prohibitive computing time for minimization given our present computational machinery (a Star ST-100 attached processor). The consequences of these observations for further modeling and the anticipated use of an additional specifically designed attached processor designed in, and presently being constructed for, our laboratory<sup>39</sup> are discussed.

## MATERIALS AND METHODS

Our approach to modeling the CDRs of antibody molecules begins with the preparation of libraries of randomly generated conformations for the heavy atoms of the hypervariable loops as presented in the first paper in this series.<sup>23</sup> These conformations are first subjected to screening which eliminates structures with van der Waals overlaps within the loop or with the rest of the molecule, and are then subjected to energy minimization and/or to molecular dynamics with the remainder of the molecule held fixed. The lowest energy conformations found in this way are then further examined by generating possible side chain orientations.

Two classes of studies have been carried out. In the first class, the target system consists of a single loop only, with the conformation of the rest of the molecule assumed known. This class is important when a new antibody sequence is found in which only one or several non-interacting CDRs are changed from those in a known structure. It is also potentially important for modeling the results of amino acid substitutions within a loop, either as a result of somatic mutation or of site-directed mutagenesis experiments. For this class, the other CDRs are present both during the van der Waals screening and the subsequent minimization or dynamics. In the second class, the target system again consists of only a single loop, but here the other CDRs are assumed to be unknown. For this class, the other CDR loops are removed both during the van der Waals screening and the subsequent minimization or dynamics. This enumeration of possible conformations of a CDR independent of the conformations of the other loops is a first step in the full exploration of all possible conformations of the binding site of the protein (see Discussion).

## The Crystal Structure and the CDRs

All of the work reported in this paper is based on the atomic coordinates for the variable domains of MCPC603 obtained from the 1MCP entry in the Brookhaven Protein Databank.<sup>24</sup> The domains consist of residues H1 to H113 of the heavy chain and L1 to L103 of the light chain, in the Kabat et al. numbering scheme used throughout this paper.<sup>2</sup> The lengths, base extents, and numbers of contacts with other CDRs and with framework are given in Table I for the light chain CDRs (L1, L2, and L3) and for the heavy chain CDRs (H1, H2, and H3). In this paper we consider H1, H3, L2, and L3, whose residue numbers and sequences are given below.

H1: [H31–H35] Asp Phe Tyr Met Glu  
H3: [H95–H102] Asn Tyr Tyr: Gly Ser Thr Trp Tyr  
Phe Asp Val  
L2: [L50–L56] Gly Ala Ser Thr Arg Glu Ser  
L3: [L89–L97] Gln Asn Asp His Ser Tyr Pro Leu  
Thr

We have chosen to work with the original definition of CDRs given by Kabat et al. based on variability in sequence.<sup>2</sup> Other definitions are possible, and several have been suggested in the literature. Novotny' et al.<sup>15</sup> define the extent of hypervariable loops which contribute to the anatomy of the antibody binding site as [H31–H32] for H1, [H96–H100] for H3, [L50–L56] for L2, and [L91–L96] for L3. These compose subregions of the CDRs defined by Kabat et al. Chothia et al.<sup>14</sup> consider the structurally variable loops of MCPC603 to be composed of [H26–H32] for H1, [H96–H101] for H3, [L50–L52] for L2, and [L91–L96] for L3. These are similarly subregions of the Kabat et al.

**TABLE I. Summary of Contacts Made by CDRs With Other CDRs and With Framework Portions of the Molecule, With Coordinates Taken From 1MCP\***

	Number of contacts: 4 Å cutoff					
	L1	L2	L3	H1	H2	H3
L1		22	49	0	0	56
L2	22		0	0	0	7
L3	49	0		20	14	43
H1	0	0	20		66	40
H2	0	0	14	66		0
H3	56	7	43	40	0	
FL	192	132	77	0	0	35
FH	0	0	17	106	203	97
PC	0	0	14	10	19	11
Length	17	7	9	5	19	11
Base(Å)	12.93	14.37	5.66	13.31	13.87	7.97

Also given: lengths and base extents of CDRs (defined as the  $C\alpha$  to  $C\alpha$  distance from the first to the last residue). The last column gives contacts made to the hapten, phosphocholine, in a separately solved structure, 2MCP. \*FL means light chain framework; FH means heavy chain framework; PC means the presumed hapten, phosphocholine, in the companion crystal structure 2MCP. Contacts are defined as pairs of heavy atoms whose centers lie within 4 Å of one another. The base extent is defined as the  $C\alpha$ - $C\alpha$  distance from the first to the last residue of the loop. The length is given as the number of amino acids.

CDRs in all cases except H1, for which the Chothia et al. definition is both longer and more N-terminal extensive. The portion of H2 defined by Chothia et al. which lies outside of the Kabat et al. CDR is oriented away from the combining site of MCPC603, and lies some 14 Å removed from the co-crystallized phosphocholine in the 2MCP crystal structure. Whereas this portion of the molecule may be structurally variable and important to the problem of immunoglobulin binding to the large surface of lysozyme, it is likely to be less important to understanding the structure of the combining site elicited by a small hapten such as phosphocholine. Our overall choice in modeling the somewhat longer CDRs defined by Kabat et al. was motivated by concern that the database of presently known structures may be too limited realistically to reduce those regions of structural variability beyond the regions shown by Kabat et al. to be hypervariable in sequence.

#### Initial Minimizations of the Crystal Structure

The crystal coordinates were subjected to 2,500 cycles of conjugate gradient cartesian energy minimization with both the DISCOVER<sup>25</sup> software from the Biosym corporation and the CHARMM/GEMM software from Harvard and the National Institutes of Health.<sup>26,27</sup> Here, as throughout the rest of the paper, studies were carried out with two programs to allow comparison of results obtained from two different force fields. The salient differences between these force fields are discussed below. In these minimizations, all atoms are allowed to move. The minimized structure obtained from DISCOVER was used for the generation of random CDR loop conformations, as described below.

#### Loop Generation

For each CDR to be modeled, all side chains except proline and glycine were first "mutated" to alanine by removing all atoms beyond the  $C\beta$ . This procedure removes all atoms whose positions are not defined by setting the phi and psi angles along the backbone (note that our loop generation procedure uses heavy atoms only). Next, a library of several hundred random structures for each mutated loop was generated subject to the constraint that the N and  $C\alpha$  atom coordinates of the N-terminal residue and the  $C\alpha$  and carbonyl C atom coordinates of the C-terminal residue of the loop match those of the minimized MCPC603 structure obtained with DISCOVER as described above. The details of the methodology employed to generate these loop structures subject to the end constraints are covered in the first paper in this series.<sup>23</sup> Here, we simply note that the method is efficient compared to tree-search ring closure algorithms for generating loops with fixed-end geometries and that it seemingly covers the conformational space available to the loops uniformly (see Discussion). The method uses random numbers to set the phi and psi angles along the backbone of the loop and then subjects these angles to an iterated, linearized Lagrange multiplier procedure to produce the desired fixed-end geometry for the loops with the minimal conformational perturbation.

#### Van der Waals Screening

For most of the minimizations reported in this paper, the starting CDRs were generated with van der Waals screening, which eliminated conformations exhibiting any overlaps of fixed-diameter spheres cen-

tered on the heavy atoms. This screening considered overlaps of loop atoms with each other and with atoms in the surrounding molecule. Depending on the study being performed, screening was done either against the entire remainder of the molecule or against the framework portion of the molecule only. For screening against framework, the peptide bonds connecting the framework atoms immediately C-terminal and N-terminal to the residues at the ends of the CDRs were "cleaved," and the atoms of all CDRs except the one being modeled were removed. Most of the results presented in this paper were generated with a screening diameter of 2 Å; thus, a conformation was rejected if two heavy atoms were found with atomic centers closer than 2 Å. This value was chosen to be conservative, in the sense of allowing some high-energy structures to pass while still giving initial energies and forces low enough to avoid problems in subsequent energy minimization. Such problems can arise when atoms lie too close to one another, resulting in forces too high to allow an accurate evaluation of the direction of the net force vector on the atom in question. The screening diameter used will be given along with the results below.

### Energy Minimization and Dynamics

Each entry in a screened library for a selected loop was subjected to energy minimization or to molecular dynamics followed by energy minimization in the context of either the remainder of the molecule or the framework portions only. For both minimization and molecular dynamics, only the atoms in the randomly generated loop were free to move while the rest of the molecule was held rigid. Specifically, the backbone N, H, C, O, C $\alpha$  atoms, and C $\beta$  atoms of each amino acid in the loop (along with C $\gamma$  and C $\delta$  atoms for prolines) were considered free to move, while all other atoms, including connecting atoms at the N-terminal and C-terminal ends of the loops, were held fixed. To define a "framework" context for minimizations of molecular dynamics, the same procedure as that used for van der Waals screening was used to remove all CDRs except the one being modeled; specifically, the peptide bonds connecting the C-terminal and N-terminal amino acids of the loop to the rest of the molecule were "cleaved," and the intervening atoms of the CDRs were removed.

The software used for minimization and molecular dynamics was either the DISCOVER<sup>25</sup> package from Biosym or the CHARMM/GEMM<sup>26,27</sup> package from Harvard/NIH, with suitable modifications made to handle the format of our loop libraries. Both packages were installed on a Star ST-100 array processor hosted by a Digital Equipment Corporation MicroVAX II. In DISCOVER, all residues were treated as neutral with atomic partial charges assigned according to the present DISCOVER force field. In CHARMM/GEMM, residues were considered to carry their predominant

charge at pH 7 (the pH of the crystal), and a distance-dependent dielectric constant was used to account for solvent screening. The DISCOVER force field is an all-hydrogen force field, whereas the CHARMM/GEMM force field is a united-atom force field. The minimization algorithm used in both DISCOVER and CHARMM/GEMM was an implementation of the conjugate gradient algorithm of Polak and Ribiere.<sup>28</sup> The minimization code on the Star St-100 for CHARMM/GEMM was not selected from among those available in CHARMM for a VAX; rather, it was written and implemented by us in APCL<sup>29</sup> for the Star ST-100.

The minimizations performed with CHARMM/GEMM were run for a sufficient number of conjugate gradient search directions (1,000 for H1, L2, L3, and 2,000 for H3) to converge to typical residual energy gradients of  $5.0 \times 10^{-3}$  kcal/mol/Å. Minimizations performed in DISCOVER stopped with residual gradients of  $\sim 8.0 \times 10^{-2}$  kcal/mol/Å owing to limitations in the present software. This convergence was, however, sufficient to explore the pattern of minima found and to give a meaningful comparison with the resulted from CHARMM/GEMM. Molecular dynamics was run in CHARMM/GEMM by first subjecting each random conformation to 20 steps of conjugate gradient minimization, followed by 10 ps of dynamics at 300°K, followed in turn by 1,000 steps of conjugate gradient minimization from the end point of the trajectory. The step size for the dynamics was 1.0 fs.

The use of the Star ST-100 array processor was crucial to our ability to obtain the results presented here. The CHARMM/GEMM package gave us a relative increase in computational speed over the same software running on the the MicroVAX II of roughly a factor of 150. We have found that the MicroVAX II benchmarks at about 0.8 the speed of a VAX 780 for protein molecular dynamics. The DISCOVER package was considerably slower, owing to the use of a full hydrogen force field and to an incomplete implementation of the minimization and molecular dynamics cycles on the Star St-100. In DISCOVER, forces were calculated on the Star and passed back to the MicroVAX II, where the next step of conjugate gradient minimization was performed. The minimizations and dynamics simulations presented in this paper done with CHARMM/GEMM required about 700 hours of ST-100 computing time, while the results with DISCOVER required about 1,400 hours. The time required to tackle the two longest loops, L1 and H2, was found to be prohibitive given our present hardware and software. This will be discussed further in the Conclusions.

### Side Chain Studies

Side chains were added to the lowest-energy structures for several loops, and searches were performed for energetically favorable conformations by varying the side chain torsional angles only. Most of this work

**TABLE II. RMS Deviations Between the Crystal and Minimized Crystal Heavy Atom Positions (Å) Obtained With DISCOVER and CHARMM/GEMM**

Atoms	XTAL:DISCOVER	XTAL:CHARMM/GEMM	DISCOVER:CHARMM/GEMM
All	1.21	1.25	1.32
Backbone	0.91	1.01	1.00
BackLCDR	1.01	1.04	1.10
BackHCDR	1.00	1.08	0.79

RMS deviations were calculated after least-squares superposition of the heavy backbone atoms of the molecules to remove overall rotation and translation. XTAL refers to the coordinates taken from the 1MPC entry in the Brookhaven Protein Databank. LCDR means light chain CDR atoms. HCDR means heavy chain CDR atoms. Back refers to backbone heavy atoms. All means all heavy atoms (hydrogens excluded).

was done with Pakgraf, an interactive torsional search and minimization program developed in our laboratory.<sup>30</sup> In simple cases, two-dimensional plots of energy vs. (ch1,chi2) were generated for each amino acid exhibiting a bad contact, and the values of the torsional angles were adjusted to energetically favorable regions. In more difficult cases, amino acids nearby in space were also allowed to move, and searches were performed in the full-parameter space of the chi angles for these amino acids simultaneously in 10° increments. The final arrangements were subjected to torsional minimization in Pakgraf.

## RESULTS

In this section, we present the results obtained from minimizations of the crystal structure of the Fab fragment of MCPC603 and of the randomly generated library entries of the heavy atom backbone atoms of H1, L2, L3, and H3 in the context of the minimized molecule. Characterizations of possible side chain conformations for several of the backbone minima are also presented.

### The Minimized Crystal Structures

As described in Materials and Methods, 2,500-cycle minimizations were carried out on the 1MCP crystal coordinates in both DISCOVER and in CHARMM/GEMM with all CDRs present, prior to the generation and minimization of the random conformations of the individual CDR loops. The convergence of these minimizations was such that the final rms derivative of energy with respect to the atomic cartesian coordinates was about  $5.0 \times 10^{-2}$  kcal/mol/Å for both packages. A comparison of the minima obtained, both to each other and to the initial structure, is given in Table II. Some deviation in final coordinates is both expected and observed, owing to the differences in the force fields used by the two packages: the CHARMM/GEMM force fields uses united atoms, while the DISCOVER force field uses explicit hydrogens. The rms atomic distance of the two minima from each other (about 1 Å) is roughly the same as that of each minimum from the initial crystal structure. Figure 5a–5c

shows the superpositions of the minimized conformations of H1, L2, and L3 on those of the original crystal. The displacements of the minimized loop structures from the initial crystal coordinates are small, except for L2 minimized in CHARMM/GEMM; in this case a significant deviation is found in the first three amino acids of the loop. This will be further explored in the Discussion.

### The Minimized Poly-Ala Structures

As described in Materials and Methods, our generation and minimization procedure for the conformations of the backbone atoms of a CDR begins with the removal of all atoms in the CDR residues beyond the Cβ, effectively “mutating” the non-gly, non-pro residues in the CDR to alanine. Minimizations are then performed starting from randomly generated conformations for these mutated loops. A crucial assumption implicit in this procedure is that at least one low-energy minimum exists for the mutated loop which has atomic positions close to those in the crystal minimum for the CDR with all of its side chains atoms present. If such a minimum exists, we can reasonably expect to find it among the local minima isolated by minimization of randomly generated starting conformations, assuming only that we consider a sufficient number of starting conformations.

We have examined this assumption by removing the side chain atoms on each CDR *in situ* in the minimized crystal structure, and subjecting this mutated loop to the identical minimization procedure applied to the random loops. Minimizations were performed both with other CDRs present and with other CDRs removed (framework only). The movement of heavy backbone atoms observed in these minimizations is summarized in Table III. The resulting conformations obtained with CHARMM/GEMM are shown superimposed on the minimized crystal conformation in Figure 5d–f. For minimizations with other CDRs present, the movement of the loops are small (below 0.25 Å rms for H1 and L2; 0.55 Å for the floppier L3). For minimizations with other CDRs removed, the movements are only slightly greater, ex-

TABLE III. RMS Movements ( $\text{\AA}$ ) of Heavy Atoms During Poly-Ala Minimizations

Loop	CHARMM wCDR*	CHARMM framework†	DISCOVER wCDR*	DISCOVER framework†
H1	0.31	0.59	0.26	0.45
L2	0.14	0.20	0.25	0.26
L3	0.48	1.48	0.55	0.42
H3	0.49	—	—	—

\*wCDR: minimization performed with other CDRs present and held fixed.

†Framework: minimization performed with other CDRs removed.

H3 has been studied only in CHARMM/GEMM with other CDRs present for the present work, due to the large amount of ST-100 time required to obtain results on this loop.

cept for L3 minimized in CHARMM/GEMM: here, a significant rearrangement of the top of the loop occurs (Fig. 5f).

In all but this last case, these results confirm our expectation that minima exist for the mutated CDRs which lie close to the minimized crystal. We call these minima the poly-ala minima; their conformations gauge how closely we can expect to duplicate the minimized crystal structure with our present procedures. These poly-ala minimum-energy conformations provide convenient reference structures for describing both the energy and the spatial clustering of the minimum energy conformations found starting from our random libraries, and will be used as such throughout the remainder of this paper. Note that the appropriate poly-ala minimum-energy conformation, that obtained from minimization either in the context of the rest of the molecule or in the context of the framework parts of the molecule, will always be used when discussing the minimized random structures.

### The Minimized Random Loop Structures

The results of minimization and of molecular dynamics followed by minimization starting from randomly generated conformations of H1, L2, L3, and H3 are summarized in Figures 1–4, respectively. Although the generation of the initial random structures is covered in the first paper in this series, we have included a picture of 100 randomly generated conformations of each of the four CDRs screened against the entire rest of the molecule in Figure 6c, superimposed on the rest of the variable domains of MCPC603.

Figures 1–4 are scatter-plots, showing the correlation between energy (vertical scale) and rms atomic deviation from the appropriate poly-ala minimum-energy conformation ( $D_{\text{rms}}$ ; horizontal scale) for the minima found under identical conditions. Each plot summarizes the results of minimizations of from 100 to 1,000 random starting conformations of a loop in either CHARMM/GEMM or in DISCOVER. Specific numbers are given in the figure legends. Results are presented for each force field and for minimizations with all other CDRs both present and removed. In

each plot, the zero of energy is defined as that of the appropriate poly-ala minimum-energy conformation.

In these plots, cluster of points near zero  $D_{\text{rms}}$  and zero energy represent minimizations which converge to the poly-ala minimum-energy conformation for that CDR. Points at high  $D_{\text{rms}}$  and low energy represent alternative conformations for the loop which are energetically allowed. Points at high energy, high  $D_{\text{rms}}$  represent local minima which can be ruled out by energetic criteria. No high-energy, low-rms points were found. These would represent energetically disallowed minima with conformations spatially close to that of the poly-ala minimum. In the case of L3, points which lie below zero energy but away from zero  $D_{\text{rms}}$  represent conformations of the loop which are energetically more stable than that of the poly-ala minimum. Note that if a minimization results in energy or  $D_{\text{rms}}$  too high to be included in the plots, the point is plotted in the highest bin for that variable.

We have divided the discussion of these results into two major sections. In the first section, we discuss the results from minimizations with other CDRs present. In the second, we discuss the results from minimization with other CDRs removed. The former is of use in modeling antibody molecules where changes in sequence from a known structure are isolated to a single CDR. The latter determines the extent to which interactions with the framework portions of the molecule only can determine the structure of a particular CDR. This is of interest in helping to design a strategy for modeling the complete binding site of an antibody molecule, where changes in sequence from a known structure appear in many loops.

#### Minimizations with other CDRs present

*Minimizations of H1.* Figure 1 shows that minimization of H1 in CHARMM/GEMM, minimization in DISCOVER, and molecular dynamics followed by minimization in CHARMM/GEMM with the other CDRs present and fixed all yield clusters of energetic minima with essentially the same energy and spatial configuration as the appropriate poly-ala minimum (Fig. 1a–1c, respectively). The 20 points in Figure 1a and the 45 points in Figure 1c which lie in the lowest bin in both energy and  $D_{\text{rms}}$  all lie within  $1 \times 10^{-3}$  of

zero energy and within  $1 \times 10^{-2}$  Å of zero  $D_{rms}$ ; i.e., they are virtually indistinguishable from the poly-ala minimum-energy conformation for the loop. The 34 minima in the lowest bin in energy and  $D_{rms}$  in Figure 1b are distributed over 1 kcal/mol in energy and 0.05 Å; this distribution is due to the fact that the minimizations with DISCOVER did not converge as well as those with CHARMM/GEMM (see Materials and Methods).

The results from CHARMM/GEMM (Fig. 1a,c) show two distinct clusters of low-energy minima which lie close to one another, both spatially (0.48 Å) and energetically (0.3 kcal/m). The lower  $D_{rms}$  cluster is virtually indistinguishable in conformation from the poly-ala minimum; the higher cluster deviates slightly from it in the orientation of two peptide groups (Fig. 5g). There is no major energetic barrier separating these minima; molecular dynamics at room temperature for 10 ps is sufficient to convert one structure to the other. The higher  $D_{rms}$  minimum is not seen with DISCOVER (Fig. 1b); this difference must be due either to differences in force fields or to differences in the local environment of the loop produced in the initial minimizations of the crystal structure. Although the isolation of this higher  $D_{rms}$  structure is interesting, it should be noted that all the low-energy conformations isolated for this loop, with both DISCOVER and the CHARMM/GEMM force fields, deviate little from each other and from the two poly-ala minimum-energy conformations. Since these poly-ala minimum-energy conformations of H1 lie close to the H1 conformation in the minimized crystal, it appears that base geometry, in conjunction with backboned interactions with fixed atoms from the surrounding molecule, determines the backbone geometry of H1 independent of sequence and of interactions with other CDRs. We will return to the higher  $D_{rms}$  minimum when we discuss side chain studies below.

Comparing the results shown in Figures 1a, and 1c, it is evident that molecular dynamics followed by minimization was more effective than minimization alone in converting the initial random structures of H1 to low-energy minima. Of 100 initial conformations of H1, 70 were converted to low-energy minima after dynamics was added, while only 39 of the same initial conformations were converted by pure minimization. Evidently, many local minima on the molecular energy surface for this problem are separated by activation barriers which are readily surmounted by dynamics at room temperature. Though more effective in finding low-energy minima, dynamics followed by minimization was not more efficient; each run required approximately five times more Star ST-100 time to complete than the corresponding minimization run while yielding only twice the number of low-energy minima. On the other hand, fewer initial random starting conformations would have been re-

quired to find the poly-ala minimum had the dynamics procedure been used, and thus there is a tradeoff between time spent in loop generation and time spent in minimization or dynamics when modeling a specific loop. The first paper in this series<sup>23</sup> discusses the time required to generate these structures.

*Minimizations of L2.* For L2 in the presence of other CDRs, the results of minimization or dynamics followed by minimization with CHARMM/GEMM (Fig. 2a,c) and of minimization with DISCOVER (Fig. 2b) all show a cluster of minima near zero energy and zero  $D_{rms}$ . In the CHARMM/GEMM results, the next-lowest-energy structure lies 4.0 kcal/mol higher in energy but only 0.12 Å away. In the DISCOVER results, the next cluster lies less than 1 kcal/mol higher in energy and 0.49 Å away, and a second cluster lies ~6 kcal/mol higher in energy and ~0.7 Å away. From these data, we would select the poly-ala minimum as the most likely backbone structure with CHARMM/GEMM, though both clusters might be considered as starting points for further side chain modeling. In the DISCOVER results, the two lowest-energy minima (and possibly the third) are energetically indistinguishable and would of course be considered in further modeling of side chains. These two minima are shown superimposed in Figure 5h; one peptide unit has flipped through ~80° in the beta turn portion of the CDR (first three residues). With this exception, it once again appears that base geometry along with backbone interactions with fixed atoms from the surrounding molecule determine the backbone conformation of L2. Again the data in Fig. 2a,c show that molecular dynamics followed by minimization is more effective in converting random structures to low-energy minima: molecular dynamics followed by minimization yielded 55 virtually identical low-energy minima starting from 300 initial random structures screened at 2 Å, whereas minimization alone yielded 11.

*Minimization of L3.* For L3 with other CDRs present, minimization (Fig. 3a), or dynamics followed by minimization with CHARMM/GEMM (Fig. 3c), or minimization with DISCOVER (Fig. 3b), all fail to find simple, isolated clusters of minima at low  $D_{rms}$  which are also as low in energy as other minima found. Although at least one minimum is found in each case which is close to zero energy and zero  $D_{rms}$ , the other minima deviate significantly in conformation from the poly-ala minimum and lie considerably lower in energy (up to 8 kcal/mol in DISCOVER and up to 15 kcal/mol in CHARMM/GEMM). Dynamics followed by minimization again yields a higher concentration of low-energy minima than minimization alone. Dynamics followed by minimization found 72 negative-energy structures relative to the poly-ala minimum for L3 starting from 300 random structures generated with 2 Å van der Waals screening, while

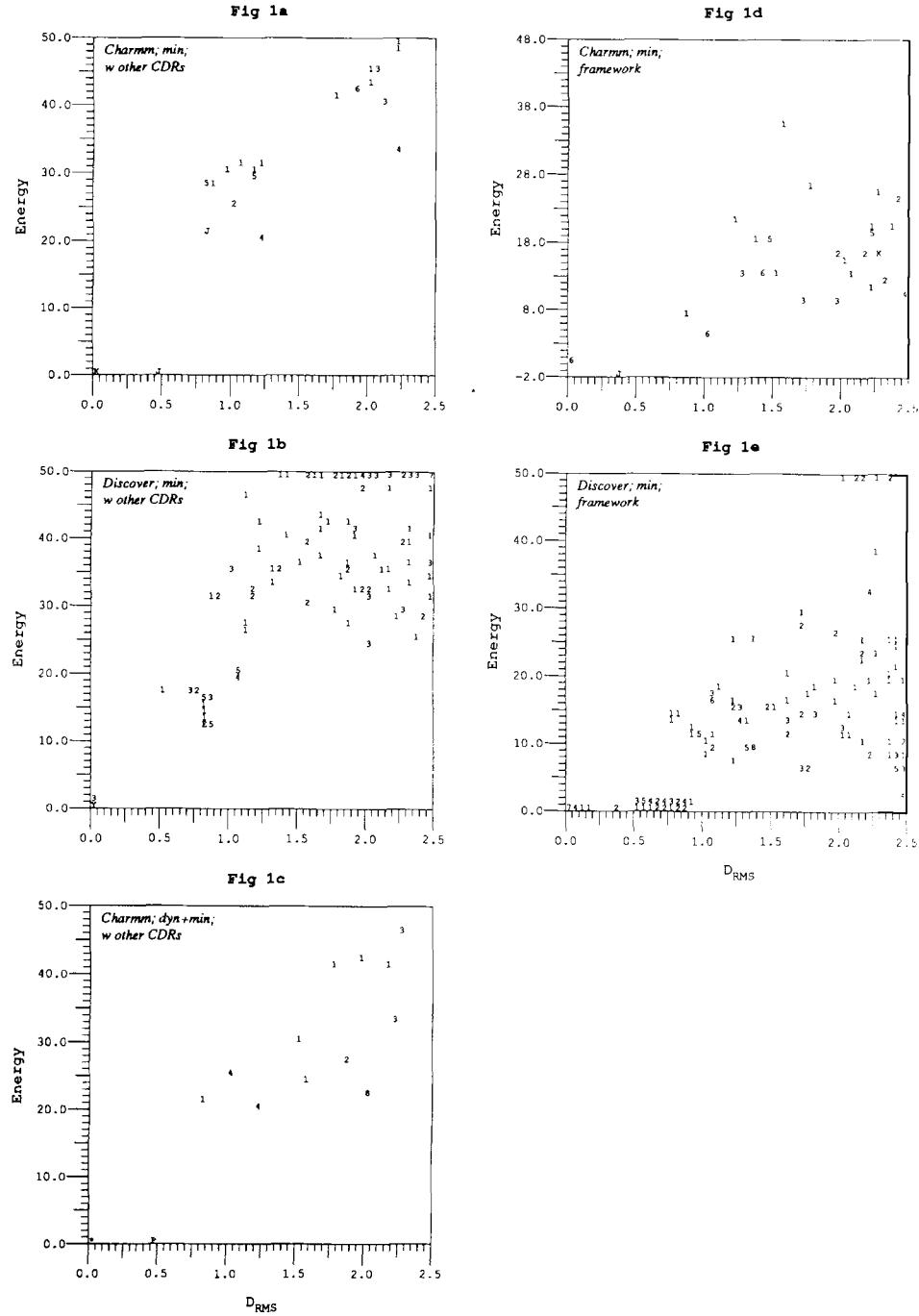


Fig. 1. Results from minimizations of randomly generated structures for H1. Results here, as well as in Figures 2 and 3, are presented as two-dimensional scatter-plots of energy above the poly-alanine minimum in kcal/mol (vertical scale) vs. atomic rms distance from the poly-alanine minimum in angstroms (horizontal scale). For each bin, the number of points is encoded as 1-9 followed by A-Z; i.e., A represents 10; Z represents 35 points within the bin. A number greater than 35 is represented by an asterisk; in these cases, the number is given in the figure caption. For all of the studies in Figure 1, the random libraries used by CHARMM/GEMM were generated with 2 Å° van der Waals screening. The random libraries used by DISCOVER were generated without van der Waals screening. **a:** Results of minimizations of 100 random structures for H1 in CHARMM/GEMM with other CDRs present and held fixed. **b:** Results of minimizations of 200 random structures of H1 in DISCOVER with other CDRs present and held fixed. **c:** Results of 10-ps molecular dynamics, followed by minimization from the trajectory end points, for 100 random structures of H1 in CHARMM/GEMM with all other CDRs present and held fixed. The asterisk represents 45 structures in the lowest-energy, rms bin (all lie within 1E-3 kcal/mol of zero energy and 1E-2 Å° of zero rms). The same initial structures were used here as in panel a. **d:** Results of minimizations of 100 random structures for H1 in CHARMM/GEMM with all other CDRs removed. The library was generated with 2 Å° screening against the framework portions of the molecule only. **e:** Results of minimizations of 100 random structures of H1 in DISCOVER with other CDRs removed.

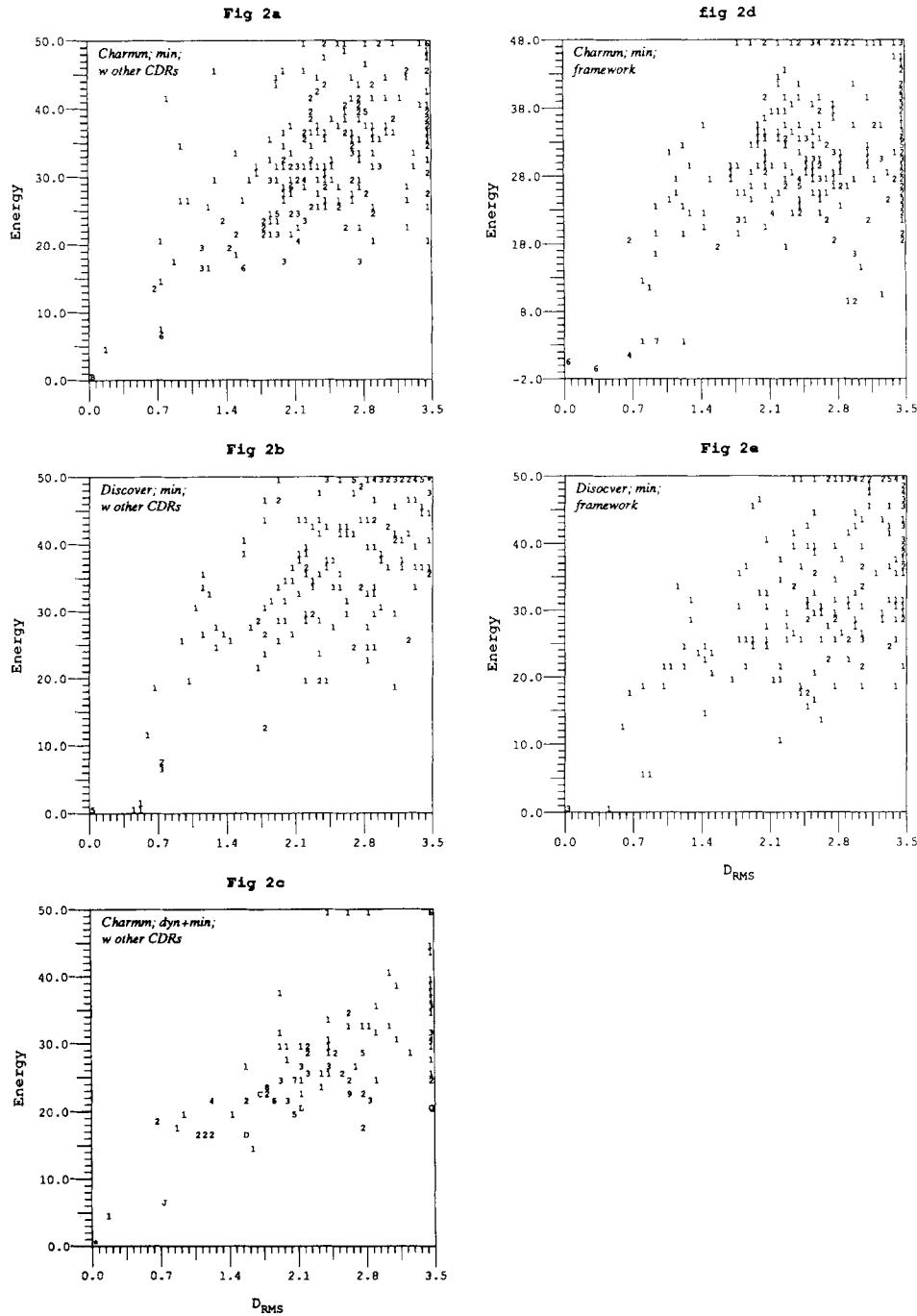


Fig. 2. Results from minimizations of randomly generated structures for L2. As for H1, the random libraries used by CHARMM/GEMM were generated with 2 Å van der Waals screening. The random libraries used by DISCOVER were generated without screening. **a:** Results of minimizations of 300 random structures for L2 in CHARMM/GEMM with all other CDRs present and held fixed. **b:** Results of minimizations of 250 random structures of L2 in DISCOVER with other CDRs present and held fixed. **c:** Results of 10-ps molecular dynamics, followed by minimization from the trajectory end points, for 300 random structures of H1 in CHARMM/GEMM with all other CDRs present and held fixed. The asterisk corresponds to 55 minima in the lowest-energy, lowest-rms bin. **d:** Results of minimizations of 300 random structures for L2 in CHARMM/GEMM with all others CDRs removed. **e:** Results of minimizations of 250 random structures of L2 in DISCOVER with other CDRs removed.

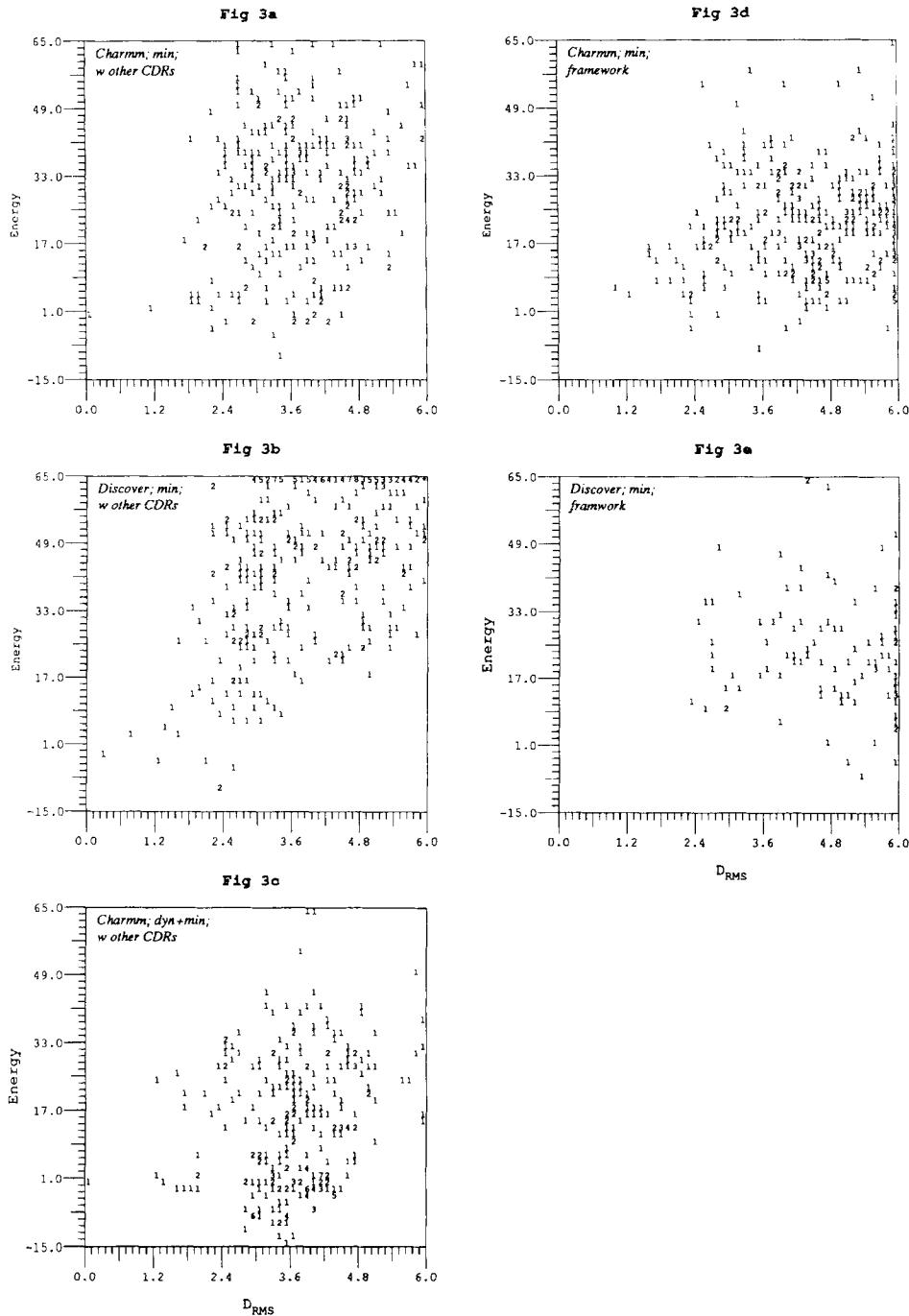


Fig. 3. Results from minimizations of randomly generated structures for L3. The random libraries used by DISCOVER were generated without van der Waals screening. The random libraries used by CHARMM/GEMM were generated with 2 Å° screening. **a:** Results of minimizations of 300 random structures for L3 in CHARMM/GEMM with all other CDRs present and held fixed. **b:** Results of minimizations of 400 random structures of L3 in DISCOVER with other CDRs present and held fixed. 200 of the initial structures were generated with 2 Å° van der Waals screening. 200 were generated with 1.0 Å° screening. **c:** Results of 10-ps molecular dynamics, followed by minimization from the trajectory end points, for 300 random structures of L3 in CHARMM/GEMM with all other CDRs present and held fixed. **d:** Results of minimizations of 350 random structures for L3 in CHARMM/GEMM with all other CDRs removed. **e:** Results of minimizations of 300 random structures of L3 in DISCOVER with other CDRs removed. The initial structures were generated with 2 Å° van der Waals screening.

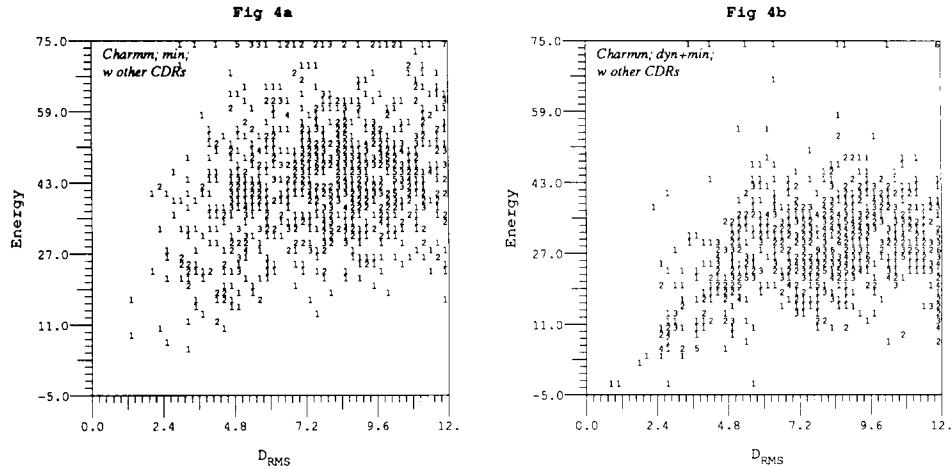


Fig. 4. Results of minimization studies on H3. Because of the computer time involved, studies on H3 were performed only with all other CDRs present and only in CHARMM/GEMM. The library generated used a combination of  $1\text{ \AA}^\circ$ ,  $1.5\text{ \AA}^\circ$ , and  $2.0\text{ \AA}^\circ$  Van der Waals screening. **a:**Results of 1,050 minimizations with other CDRs present of H3 in CHARMM/GEMM. **b:**Results of 950 10-ps molecular dynamics simulations followed by minimization from the end point of each trajectory in CHARMM/GEMM.

minimization alone found 14 negative energy structures starting from the same 300 random structures. The seven distinct lowest-energy conformations from Figure 3a found by minimization in CHARMM/GEMM are shown superimposed on the poly-alanine minimum in Figure 5i. The five lowest-energy conformations are shown in the context of the rest of the molecule in Figure 6d. The existence of these new low-energy minima for L3 indicate that the loop has considerable conformational flexibility, and that the backbone conformation depends on more than simple base geometry and interactions of backbone atoms with fixed atoms from the surrounding molecule. We will consider this flexibility further below in Side Chain Studies and in the Discussion.

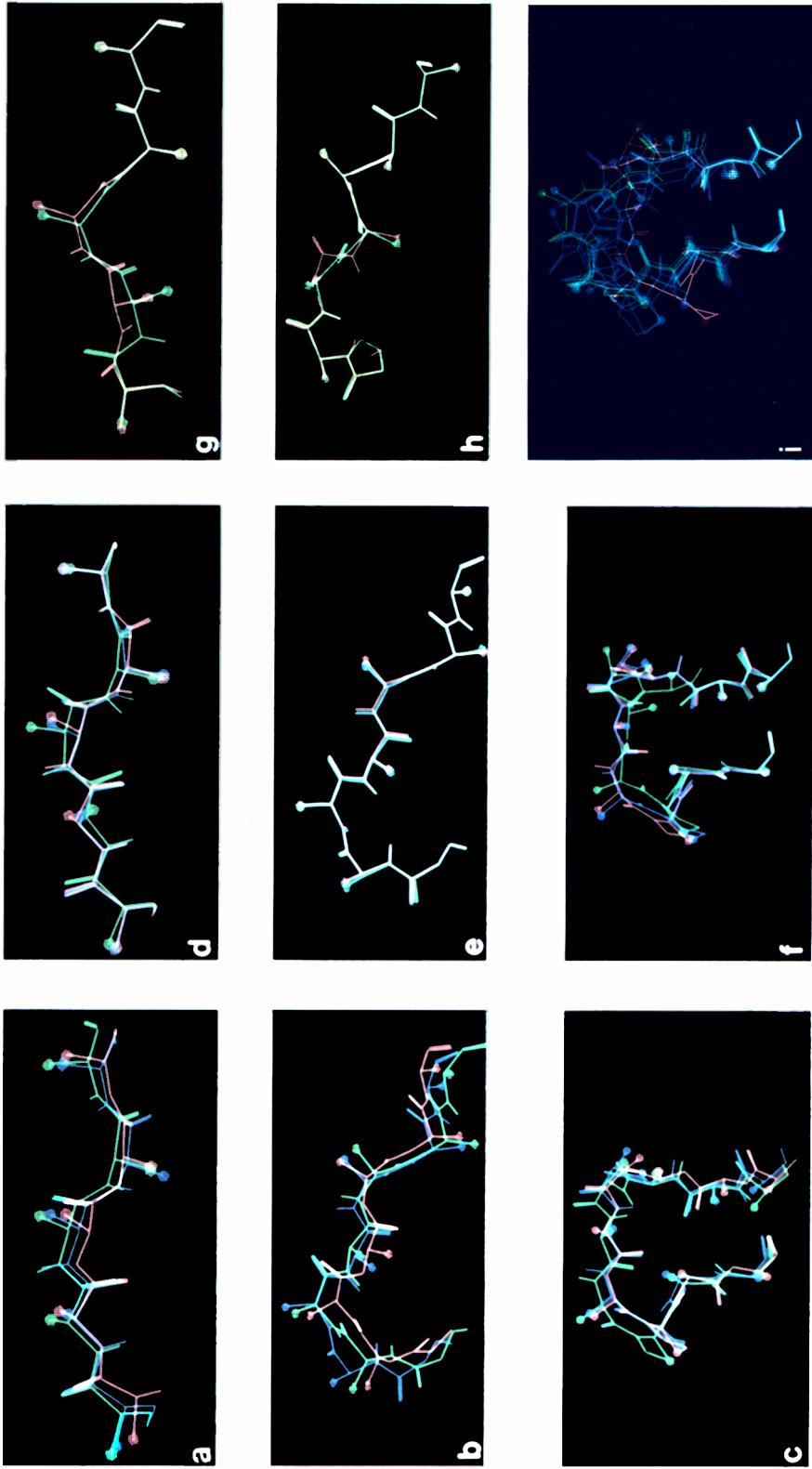
There is a proline in the seventh position of L3 which is *cis* in the crystal structure. Since our procedure for generation of the random loop conformation for the CDRs began with the backbone atomic coordinates in the crystal, the random libraries were generated with a *cis* proline in this position. As a check on possible differences, and for completeness since in the absence of crystal data we would most likely not have assumed a *cis* form for the proline, a full set of random conformation was generated with a *trans* proline in this position and minimized in CHARMM/GEMM. The same pattern of low-energy minima deviating from the crystal and poly-alanine minimum-energy conformation of L3 near the top of the loop was found with either the *trans* and *cis* forms for the proline.

**Minimizations of H3.** Figure 4a,4b shows scatter-plots of energy vs.  $D_{rms}$  for minimization and for molecular dynamics followed by minimization of H3 with CHARMM/GEMM with the other CDRs present

(we did not study H3 with DISCOVER). H3 is the longest loop we have studied to date, and the isolation of minima as low in energy as the poly-alanine minimum required considerably more starting conformations than the other CDRs (1050 in Fig. 4a; 950 in Fig. 4b). Once again, molecular dynamics followed by minimization isolated more low-energy structures than minimization alone. Only in the thousand runs of dynamics followed by minimization did we find four minima which are within 1 kcal/mol of each other and about 5 kcal/mol lower than the poly-alanine minimum. The CHARMM/GEMM poly-alanine minimum is shown in Figure 6a superimposed on the 1MCP loop coordinates before and after the initial minimizations in CHARMM/GEMM. The four lowest-energy minima from Figure 4b are shown superimposed on each other and on the poly-alanine minimum in Figure 6b. From this figure, it is clear that the general fold of the lower  $D_{rms}$  minima loosely follow the fold of the poly-alanine minimum (and of the crystal). The highest  $D_{rms}$  structure does not; we consider this further in Side Chain Studies.

#### Minimizations in the absence of other CDRs

We expect minimization of random starting conformations to be less effective in locating minima close to the native structure when other CDRs are removed, owing to the relaxation of constraints on the conformation of the single moving CDR. However, we can reasonably scale our expectations by considering the interactions which the CDRs make in the native crystal structure with framework and with other CDRs. For example, LZ exhibits a larger number of contacts with the framework than with other CDRs (Table 1); for such a loop, we might expect to find nearly the same clusters of minima with and without



**Fig. 5.** Results of minimization and dynamics studies presented pictorially. **a–c:** Superpositions of the minimized crystal forms of H1, L2, and L3, respectively, on the original crystal forms: red, 1MCP crystal conformation; blue, CHARMM/GEMM minimized crystal conformation; green, DISCOVER minimized crystal conformation. **d–f:** superpositions of poly-alanine minima on the minimized crystal conformations for H1, L2, and L3, respectively, in CHARMM/GEMM: red, minimized crystal conformation; blue, poly-alanine minima with other CDRs removed. **g:** Superposition of two low-energy minima found in CHARMM/GEMM for H1 with other CDRs present: red, low  $D_{rms}$  conformation; green, high  $D_{rms}$  conformation. **h:** Superposition of two low-energy minima found in DISCOVER for L2 with other CDRs present: red, low  $D_{rms}$  conformation; green, high  $D_{rms}$  conformation. **i:** Superposition of the seven lowest-energy distinct conformations of L3 found using CHARMM/GEMM with other CDRs present on the minimized crystal conformation: red, minimized crystal; green, lowest-energy minimum; blue, other low-energy minima.

other CDRs present. Figure 2d,2e clearly show clusters near zero energy and zero  $D_{rms}$ . In DISCOVER, the same two lowest-energy clusters are present both with and without other CDRs (Fig. 2b,e). For CHARMM/GEMM, a second low-energy minimum appears which is spatially close to the new minimum found in DISCOVER (Fig. 2a,d). L2 makes contact with other CDRs only with its first two residues, and it is in the first three residues, which form its turn portion, that the deviations occur.

H1 is characterized by approximately equal numbers of contacts with framework and with other CDRs in the crystal structure (Table I), and forms one backbone hydrogen bond to the backbone of a framework residue (H35–H93) and one to the backbone of the first residue of H3 (H33–H95). Further, as shown in the first paper in this series,<sup>23</sup> the long base extent of H2 severely restricts the possibilities for its conformations independent of its surroundings. For such a geometry, removal of the CDRs might not be critical, and in Figures 1d,e, clusters of minima are evident at zero energy and zero  $D_{rms}$  in both the CHARMM/GEMM and DISCOVER results. Recall in making such comparisons that the movement of the poly-alanine minimum-energy conformations away from the minimized crystal structure is important in gauging how close the minima are to the native crystal structure; this is shown in Table III and Figure 5d. With both programs, a second cluster of minima at low energy and about 0.5 Å  $D_{rms}$  is evident; these structures are close to the second minimum isolated with CHARMM/GEMM in the presence of other CDRs (the latter was shown in Fig. 5g). Note that the smear of minima in Figure 1e is most likely due to poor convergence in the DISCOVER results (see Materials and Methods). However, that this smear exists indicates that there is a broad range in  $D_{rms}$  available to the conformations of this loop with little energetic penalty.

For L3, the minimizations we have performed with other CDRs removed have failed to isolate a structure close to the native. However, the same pattern of minima clustered at high  $D_{rms}$  and negative energies is evident with and without other CDRs in Figure 3a,e; the only additional trend in the data in Figure 3d,e is a shift of high-energy minima to slightly lower (but still high) energies in the absence of the other CDRs. This is not unreasonable, since the removal of the other CDRs from the environment of L3 offers the loop more room to relax away from bad van der Waals overlaps, either within the loop or with surrounding framework atoms. The loss of minima near zero energy and zero  $D_{rms}$  in Figure 3d in the DISCOVER results is probably due to the smaller number of starting conformations used. Since L3 required 45 minutes per minimization in DISCOVER compared to 3 minutes per minimization in CHARMM/GEMM, we only performed 100 minimizations to obtain the results in Figure 3d.

### Side Chain Studies

Our procedures have isolated more than one energetically feasible local minimum for the backbone conformations of H1, L2, L3 and H3 under various conditions and it was our hope that ambiguous minima could be sorted further by consideration of the actual sequence. We are presently engaged in the development of a comprehensive program to determine the set of energetically allowed side chain conformations for any given backbone conformation. However, for the work reported here we have used our interactive torsional modeling package, Pakgraf, to examine the side chain conformations available to selected backbone minima (see Materials and Methods).

For H1 with other CDRs present, minimizations with CHARMM/GEMM isolated two backbone conformations which differ from each other by 0.48 Å (Figures 1a,5g). For the conformation with no deviation from the poly-alanine minimum, we found acceptable side chain orientations of the CDR side chains with only minor adjustment of surrounding side chain degrees of freedom. Here, we define "acceptable" to mean that there are no van der Waals contacts with energies greater than about 1 kcal/mol, and that the overall energy is negative according to the Pakgraf force field.<sup>30</sup> For the higher  $D_{rms}$  structure we found unacceptable van der Waals overlaps for all orientations of the phenylalanine residue in the second position of the loop and of the residues which it could contact. These could only be relieved by subsequent relaxation of backbone atom positions with CHARMM/GEMM, which forced the backbone minimum closer to the poly-alanine minimum. Although this is a pleasing result, it should be noted that the two minima found in CHARMM/GEMM certainly lie within the expected resolution of our present procedures, both from each other and from the poly-alanine minimum structure.

For L2, two distinct minima were isolated in DISCOVER with the other CDRs present. It was possible to build side chains into acceptable positions for both of these. As such, it would be impossible in the context of our present energy calculations to select between these two backbone conformations.

For L3, our procedures found a collection of backbone minima lying lower in energy than the poly-alanine minimum with both CHARMM/GEMM and DISCOVER. To investigate whether side chain interactions could play a role in determining the structure of the loop, side chains were built onto the lowest-energy minimum found with DISCOVER by means of Pakgraf, and the resulting molecule was minimized with both DISCOVER and CHARMM/GEMM, holding no atoms fixed. In the latter case, minimizations were performed with both a constant and a distance-dependent dielectric and with the histidine in the loop charged and uncharged. For comparison

with the resulting conformations, which we will call the *minmin* conformations below, the 1MCP coordinates were subjected to minimizations under identical conditions.

The results of these studies are summarized in Table IV. The energies shown have been calculated as the sum of internal energy for L3 plus the energy of interaction of the loop with the surrounding molecule, under the various force field conditions listed. The first thing to note from this table is that the differences in energy between the minimized crystal structure and the *minmin* structure are largely ameliorated by including side chains. In all but the case of CHARMM/GEMM with constant dielectric and an uncharged loop histidine, the differences are reduced to less than 6 kcal/mol. The second thing to note from the table is that, in all cases except CHARMM/GEMM with a distance-dependent dielectric and a neutral histidine, the minimized crystal conformation ranks lower in energy than the *minmin* conformation. Thus, for L3, the inclusion of side chain interactions has proven to be critical to the energetic selection of conformation for the loop.

For L3, our assumption that a search for minimum-energy conformations of only the backbone atoms of the CDR will yield a collection of low-energy conformations which includes the native structure is thus correct only if we consider conformations of that backbone which range up to about 15 kcal/mol higher in energy than the lowest energy minimum found with our poly-ala representation of the loops. We will explore the implications of this further in the Discussion.

For H3, the highest  $D_{rms}$  conformation shown in Figure 6b is considerably different from that of the poly-ala minimum. We have used Pakgraf to inquire whether the side chains of the MCPC603 sequence generate sufficient overlaps to eliminate this structure. The answer is yes; however, the lowest  $D_{rms}$  structure is consistent with the sequence. Thus, our ability to select conformations for this loop is limited to selecting low-energy conformations which lie within about 1 Å rms of the poly-ala minimum.

## DISCUSSION

This is the second paper in a series in which we intend to study the structure of the combining sites of antibody molecules whose structure is known, and to use the resulting methods to predict the conformations and binding properties of antibodies whose three-dimensional structure is unknown but whose amino acid sequence has been determined. Our overall approach is to examine a large set of conformations of the CDRs which are compatible with the constraints imposed by interactions with the rest of the molecule and with each other. A complete energetic conformational search procedure would entail generation of a library of conformations for each loop, which we have done; selection of all mutually com-

patible conformations for each of the CDRs; and exploration of all possible orientations of the side chains for each of the CDR combinations generated. Such an exhaustive conformational search is computationally demanding but is under way in our laboratory. In this paper, we have reported the results of a preliminary investigation, in which we investigate the low-energy conformations for a "minimal volume" representation of the CDRs generated by the removal of all atoms beyond the  $C\beta$  on the side chains. This "minimal volume" representation explores the largest set of possible conformations available to any particular sequence of amino acids for the loop. The results found from minimizations and molecular dynamics starting from random conformations of this representation for the CDR can show us the extent to which the base geometry, length, and backbone hydrogen bonding patterns of the CDR determine its conformation. In this context it is worth noting that, based on superpositions of conformations of hypervariable loops in crystallographically determined structures, de la Paz et al. have suggested that the length of a loop is primary and its sequence secondary in determining its conformation.<sup>20,21</sup> We have performed our minimizations both in the context of the molecule with other CDRs present and with other CDRs removed. The latter determine the extent to which interactions of a CDR with only framework portions of the molecule determine its conformation. Such knowledge is useful in designing a progressive strategy for modeling a complete combining site by methods similar to those used in this paper. Those CDRs should be modeled first whose low-energy minima with other CDRs removed are similar to those with other CDRs present. These CDRs can be added to the "framework" for the remaining CDRs, and the process iterated.

### H1

The order of presentation in the Results (H1, L2, L3, H3) was in fact the order in which we expected increasing difficulty in using our present methodology to predict the conformation of the CDRs. H1 is a short CDR of five amino acids with an extended base, whose backbone forms one hydrogen bond to the backbone of a framework residue (H35–H93) and one hydrogen bond to the backbone of the first residue of H3 (H33–H95). As such, we might expect its conformation to be primarily determined by these interactions, and only weakly determined by side chain interactions. The fact that minimizations with other CDRs present isolate only one low-energy minimum for H1 with DISCOVER and two spatially similar minima which interconvert by molecular dynamics with CHARMM/GEMM confirms this expectation. The further fact that a similar pattern of minima is found for minimizations with other CDRs present and other CDRs removed suggests that its interaction with framework is sufficient to determine its structure.

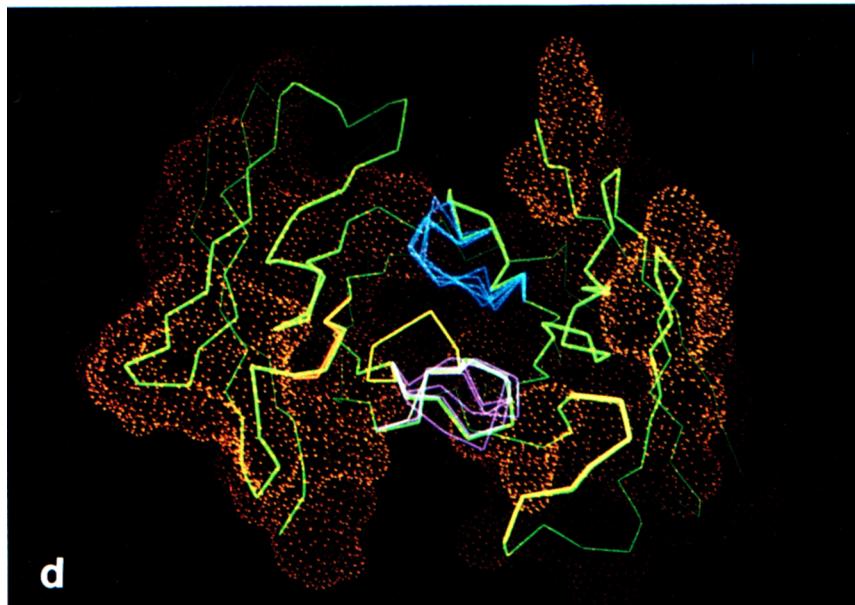
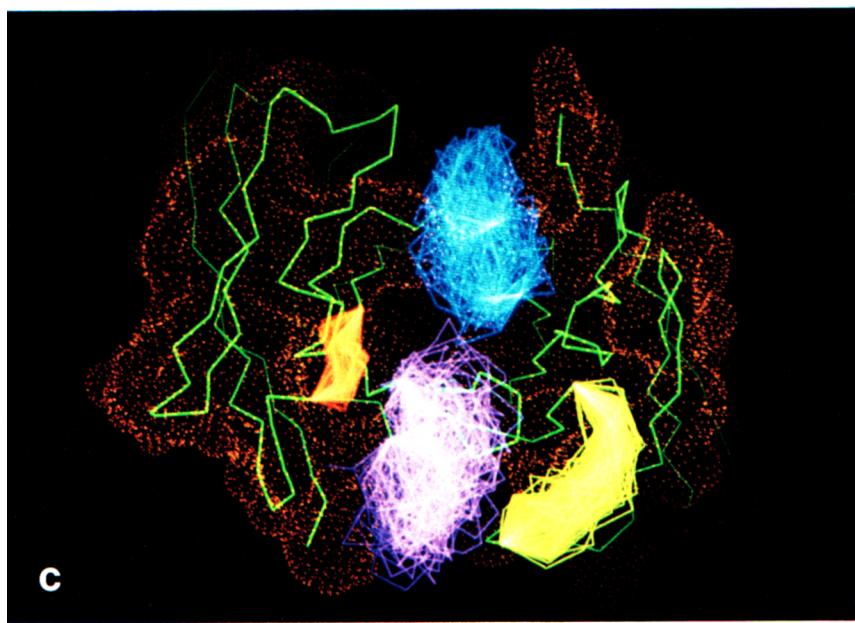
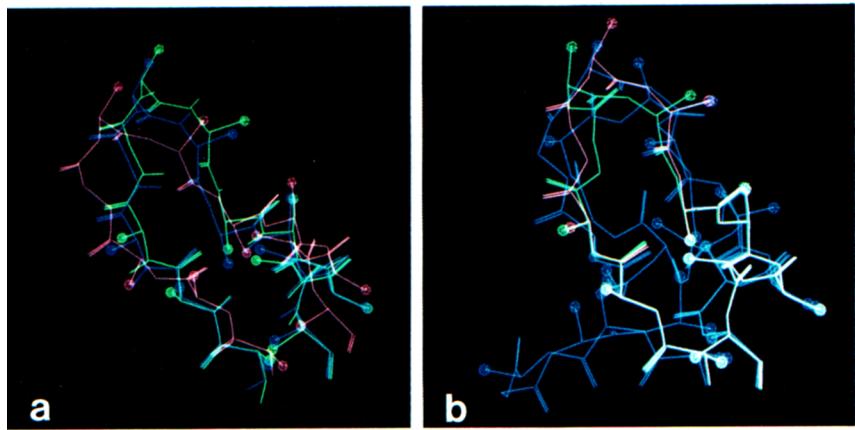


TABLE IV. Energy of Minimum Structures of L3 Found Under Various Force Field Assumptions\*

Dielectric His charge	CHARMM/GEMM				DISCOVER
	$\epsilon = r$		$\epsilon = 1$		$\epsilon = 1$
	+	0	+	0	0
Minimized crystal	-236.1	-218.9	-476.4	-478.1	-81.5
<i>Minimin</i>	-233.1	-221.3	-474.5	-465.0	-75.9
<i>Minmin-minimized</i> crystal	+3.0	-2.4	+1.9	+13.1	+5.6

\*Minimized crystal refers to results of minimizations performed from the original crystal coordinates. *minmin* refers to results of minimizations performed from the lowest-energy backbone conformation isolated by random minimizations in DISCOVER, after building side chains onto this structure by means of torsional minimization procedures.

Our method has therefore both found the correct conformation for H1, and suggested that this structure will be sequence independent for all H1 loops with the same length provided that no glys or pros occur in the new sequence. Of course, our present procedures are spatially conservative in the sense that the side chains of the loops we have investigated have been replaced by side chains which occupy less volume. As such, when considering a new sequence with a bulky side chain such as phenylalanine in a position where a less bulky side chain exists in MCPC603, the ability of the side chain to fit in the available space should be checked. Further, the introduction of new charged amino acids may result in large forces, both from other nearby charges (accounted for in our present force fields) and from interactions with solvent (not accounted for in our present force fields). Those effects which are modeled by our present force fields can be quickly evaluated for a new sequence by building the side chains in place, and performing further minimizations to determine whether the backbone conformation moves.

In explaining the lack of dependence of the conformation of H1 on side chain interactions, we should note that the conformations available to H1 are severely restricted by the long extent of its base, as clearly seen in figures which depict random conformations for H1 with and without van der Waals screening in the first article in this series. A  $C\alpha$  superposition of H1 from the three crystallographically determined light-chain heavy-chain dimers KOL,<sup>4</sup> NEW,<sup>6</sup> and MCPC603<sup>3,3a</sup> can be found in de

la Paz et al.,<sup>20</sup> which illustrates the virtual identity of conformation of H1 in these proteins. In predicting the conformation of H1 for the then unsolved structure of the immunoglobulin D1.3, Chothia et al.<sup>22</sup> suggested that the structure of H1 would be similar to that observed in KOL (and hence MCPC603), which did not in fact turn out to be the case. However, as described in Materials and Methods, the hypervariable loop definition of Chothia et al. for H1 is not co-extensive with that of Kabat et al., and the deviations in the conformation of the H1 structure in D1.3 are N-terminal to the region which we have considered in this paper. The two residues in common in the two definitions, H31 and H32, have the same backbone conformation in D1.3 and KOL.

## L2

L2 is a slightly longer loop of seven amino acids, which is composed of a turn region of three amino acids and an extended region of four amino acids. The backbone of the extended region forms hydrogen bonds with the backbone of framework: L50-L39, L53-L49 twice, and L55-L47. It is constrained to fold in a restricted region of space, between a wall of conserved framework amino acids on one side and a conserved aromatic residue (Tyr L49) on the other. This conserved residue forms part of the interface between the light and heavy chain variable domains. For such a loop, we might reasonably expect our procedures to isolate the crystallographic structure, and in general, the results confirm these expectations. In minimizations with other CDRs present, a

Fig. 6. Further results of minimizations presented pictorially. **a:** The H3 conformation in 1 MCP before minimization (red) and after minimization in CHARMM/GEMM (blue), along with the poly-alanine minimum for the loop (green). **b:** The four lowest-energy conformations of H3 found in CHARMM/GEMM with other CDRs present, superimposed on the poly-alanine minimum: red, poly-alanine minimum; green: lowest-energy conformation; blue: others. **c:** One hundred randomly generated structures for each of the CDRs studied in this paper. The surface of the protein shown has been generated as a solvent-accessible surface, after removal of all CDRs from the molecule. The green trace shows the minimized crystal structure from DISCOVER as a  $C\alpha$  plot. Clockwise from the upper left, the CDRs are shown color-coded yellow for L2; purple for H3, orange for H1, and blue for L3. **d:** The five lowest-energy structures found in CHARMM/GEMM for each of the CDRs studied in this paper. Note that for H1 and L2, these superimpose to the line resolution of the figure. Orientation and color-coding is the same as for panel c except for H3, for which one low-energy conformation has been colored yellow. This is the low-energy structure described in the text which can be eliminated by bad van der Waals contacts once side chains are built onto the loop. The deviation of the blue low-energy structures for L3 from the green minimized crystal conformation near the top of the loop can be clearly seen.

single well-isolated low-energy minimum is found in CHARMM/GEMM, and two minima are found in DISCOVER. Each package isolated one minimum which is virtually indistinguishable from the poly-ala minimum for the loop. In minimizations performed with other CDRs removed, a structure close to the conformation of the poly-ala minimum for the loop was always found, which is expected since the contacts made by L2 are almost entirely with framework (Table I).

A closer look at the results on L2 reveals more variability in the structures found by our methods for this CDR than for H1. In DISCOVER, two minima were found with a significant ( $\sim 80^\circ$ ) flip of one of the peptide groups in the turn portion of the CDR (Fig. 5h); side chain studies revealed that both conformations are compatible with side chain interactions for the actual sequence of L2. Furthermore, there was structural movement in this loop during the initial minimization of the crystal coordinates with CHARMM/GEMM. The subsequent poly-ala minimizations and random library minimizations reproduce the minimized crystal conformation rather well (Fig. 5e), but deviate from the conformation in the 1MCP coordinates in the first three amino acids of the CDR (Fig. 5b). The removal of side chain atoms beyond the  $C\beta$  should not be as critical for modeling the first three residues as for other residues in L2, since the first two residues are glycine and alanine and hence are identical in sequence in the protein and in the random library entries. In terms of bulky residues, only the orientation of the third  $C\alpha$ - $C\beta$  bond (serine) is affected in the deviations shown in Figure 5b,h. The glycine in the first position offers considerable flexibility to L2 in MCP603, and indeed, the ( $\phi$ , $\psi$ ) values of this gly in the 1MCP coordinates ( $66^\circ$ ,  $193^\circ$ ) are not normally accessible to non-gly residues. Although gly is the most frequently found residue in this position of L2, many known sequences exhibit non-gly residues in this position, and hence structural variation is expected with sequence variation in this portion of the loop.

### L3

There are two aspects to the results for L3 which we wish to note. First, since minima are found with both CHARMM/GEMM and DISCOVER which lie considerably lower in energy than the poly-ala conformation, we cannot *a priori* expect minimization of loops without their side chains to isolate conformations close to the native for long, floppy loops like L3 which lack hydrogen bonds to anchor them to their surroundings. Here and below, we are using the term "floppy" in the sense introduced in the first paper in this series,<sup>23</sup> to mean a loop whose base vector length is a small fraction of the maximum possible extent for the loop. We have examined a few of the low-energy backbone minima isolated in CHARMM/

GEMM to determine the source of the large energy differences between them and the poly-ala minimum for the loop, and two contributions were found. The first is a larger number of van der Waals contacts made to surrounding residues by the low-energy minimum conformation than by the poly-ala minimum. This contribution is artificial, since van der Waals contacts made by side chains in the original crystal conformation are broken when the side chains are removed. The second is a new interaction between a backbone hydrogen bonding dipole and a charged residue (Arg H52, which is in H2). This interaction is particularly strong when evaluated by means of CHARMM/GEMM with a distance-dependent dielectric, and may be overemphasised in our present work owing to the lack of explicit solvent. We should remark that the clustering of minima whose conformations deviate from the poly-ala minimum and the crystal minimum conformation near the top of the loop is further supported by a 1-ns simulation of L3 without side chains with CHARMM/GEMM. Starting from the lowest energy minimum from Figure 3c with the remainder of the molecule held fixed, there was no significant movement away from these conformations. Of course, this might be anticipated from the large energy difference between these minima and the poly-ala minimum.

The second notable aspect of the L3 results is that since the large energy differences between the minimized crystal structure and the low-energy minima are reduced to a few kcal/mol by including side chains, any future procedure incorporating side chains which uses a force field similar to those used here is likely to find a variety of low-energy conformational states for L3 which differ significantly from the native crystal structure. From visual examination of the intermolecular contacts in the crystal structure of 1MCP on an Evans and Sutherland PS300, it seems unlikely that these contacts are responsible for the observed conformation. The observed mobility in this region of the crystal is low (B factors of 5–6); as such, theoretical considerations not included in our present energy calculations will be needed to select among the low-energy conformations for L3 including side chains. One candidate for such a consideration is solvation energy, which has been included here only through the use of a distance-dependent dielectric in CHARMM/GEMM. The distance-dependent dielectric accounts only crudely for the screening of Coulomb interactions between charges and dipoles at the surface and the interior of the protein, and ignores completely the Born energy association with the total or parital burial of a charge beneath the protein surface.<sup>31,32</sup> Such considerations can be included either by using explicit water molecules during simulations or by incorporating continuum electrostatic calculations into the minimization or molecular dynamics procedures.<sup>33</sup> Another contribution to the solvation

energy of a protein is related to the surface area of the protein which is buried during the folding process. This has been shown empirically to be correlated with the energy of transfer of amino acids from water to organic solvents,<sup>34</sup> and conjectured to be important to protein stability.<sup>34-36</sup> Whereas this contribution may play a role for other CDRs, a calculation of the accessible surface of L3 for each of the low-energy minima including side chains (*minmin* structures) and for the minimized crystal structure from CHARMM/GEMM by means of the method of Shrike and Rupley<sup>37</sup> shows little difference in accessible area (several square angstroms).

### H3

H3 is the longest loop which we have tackled to date (11 amino acids), and as such required considerably more starting conformations and hence Star ST-100 time to find minima with as low an energy as the poly-alanine minimum. H3 is a loop which varies significantly in length in known sequences. In MCPC603, it forms an extended structure which is a twisted extension of the beta barrel of the heavy chain domain. It forms hydrogen bonds with itself but not with its surroundings except in the first residues at its base. Its sequence has a large number of bulky aromatic residues, which near its base, are packed tightly against surrounding residues. As such, there is little to dictate its structure when represented by our minimum volume representation for the CDR, and we might expect to find low-energy minima which deviate significantly from the poly-alanine minimum for the CDR.

The four low-energy minima found with the use of dynamics followed by minimization from 950 starting conformations in CHARMM/GEMM do show considerable variation in rms deviation from the poly-alanine minimum for loop, the largest deviation being ~5.8 Å  $D_{rms}$  (Fig. 4b,6b). Note that, unlike the results presented for H1 and L2, the number of minimizations performed has not been sufficient to multiply populate the low-energy minima, and therefore minimizations are likely to yield additional low-energy minima over this range of  $D_{rms}$ . We were not able to build side chains onto the conformation at ~5.8 Å  $D_{rms}$ ; however, the difficulty we found was with side chain overlaps near the base of the loop and not at the top of the loop where large deviations occur from both the 1MCP conformation and the poly-alanine minimum. This leaves open the possibility that additional minima might exist which are compatible with the MCPC603 sequence and which show at least as much deviation as does this structure from the poly-alanine minimum or the minimized crystal. Although the minimum with the lowest  $D_{rms}$  exhibited positive van der Waals energies of ~1 kcal/mol to the Pakggraf force field when side chains were added and minimized holding the backbone of the loop fixed, the

movement of the C $\beta$ s which would be required to alleviate the overlaps was less than 0.2 Å average which we considered insignificant. Thus, the lowest  $D_{rms}$  structure, which roughly follows the poly-alanine minimum and the minimized crystal conformation for H3 (Fig. 5a,b), is compatible with the MCPC603 sequence. We have not attempted to build side chains onto the other two low-energy minima.

### A Related Procedure

A technique for the prediction of the conformation of loop structure in proteins which is somewhat similar to ours has recently been described by Moult and James.<sup>30</sup> In their procedure, a large number of possible loop conformations is generated by a tree-search algorithm, which selects from a predetermined set of allowed values for the phi, psi, and side-chain chi angles for the loop in question. These authors do not rely on minimization to refine the loop conformations further; rather, for the two short loops in *Streptomyces griseus* trypsin which they have modeled, they find that a screen against bad van der Waals contacts followed by a single ranking of the structures on the basis of electrostatic energy is sufficient to select conformations close to the crystallographic native. Two remarks are in order in comparing their work to that presented here. First, tree-search algorithms for generating backbone conformations are slow in producing an acceptable structure for long loops. Though we started with procedures similar to those described by Moult and James, we found their effectiveness to be practically limited to short loops of roughly seven amino acids or less (the loops modeled by Moult and James involved backbone degrees of freedom for three and five amino acids). The second point is that their examination of many side chain conformations on the distinct acceptable backbone conformations goes beyond the scope of our present paper. In this paper, we have relied on the backbone interactions of the loops as described by present-day force fields to delimit a set of low-energy loop conformations which are further annealed by energy minimization; Moult and James rely primarily on considerations of electrostatic energy, by using an image charge representation of the effect of solvent on the interaction between charged groups in the protein.

Solvation and electrostatic effects are clearly important, as evidenced in our results by the changes in relative energies of the *minmin* vs. minimized crystal conformations of L3 on substitution of a distance-dependent dielectric for a constant dielectric. In continuing our efforts in modeling antibodies, the next step is the introduction of many possible side chain conformations onto the backbone structures isolated by means of our present procedures. When charged side chains are introduced, we intend to consider electrostatic solvation effects, including Born self-energy effects, into our minimization procedures.

## CONCLUSIONS

We have proposed a method for exploring a large number of local minima for the backbone conformations of loops in proteins and applied this method to four of the CDRs of the known crystal structure MCPC603 by means of two different programs for minimization and molecular dynamics. Since the method we introduce in this paper initially ignores side chains, the goal of these efforts is to find energetically allowable conformations for the backbones of the CDRs for further characterization of possible side chain orientations. Our results have met with some success for each of the CDRs studied, in that distinct local minima or clusters of local minima have been isolated which lie close to the minimized crystal structure. For the two shortest loops (H1 and L2), the minimization procedures introduced have been, perhaps surprisingly, sufficient to select either a single (L2 in CHARMM/GEMM) or two spatially similar (L2 in DISCOVER and H1 in CHARMM/GEMM) low-energy minima which lie close to the conformation of the corresponding CDRs in the minimized crystal. This result has allowed us to address those factors which determine the structure of these loops in MCPC603, and to suggest that their conformation will be sequence independent in other antibodies in which the loops have the same length. These conclusions support the assertion of de la Paz et al. based on a comparison of conformations of the CDRs in crystallographically solved structures that length and base geometry are primary factors in determining CDR conformations (at least for short base-extended loops) and that side chain interactions are secondary.<sup>20,21</sup> Our further result that similar minima are found when other CDRs are either present or absent allows us to conclude that interactions with the framework portion of the molecule along with loop length and base geometry are sufficient to define narrowly the possible structures of both H1 and L2.

For the two longer loops examined (L3 and H3), our procedures have isolated a number of low-energy backbone conformations which deviate significantly from the observed crystal structure. With side chains removed, minima for one of the loops (L3) are found which are energetically favorable compared to the minimized crystal conformation by up to 15 kcal/mol. There are two conclusions which seem inevitable from these observations. The first is that the future investigation of these longer, floppier loops will have to include side chains at an early stage in the modeling process. The second is that the number of conformational states available to these loops is large, and that side chains and perhaps interactions with substrate can play a critical role in determining the conformation of the loops. In the case of H3, considerations of side chains were sufficient to eliminate the backbone conformation which deviated the most from the crystal structure. However, in the case of L3, further minimizations after incorporation of side

chains on the backbone did not allow such a selection, and in fact decreased the energy differences between the backbone minima. As such, results on L3 also point to inadequacies in the application of simple minimization and dynamics procedures to proteins in the absence of solvent or of a surrounding crystalline environment.

The results for all of the CDRs are informative, in that the clustering of local minima found in energy and in space tells us when the application of the present procedure is likely to yield a robust prediction for the conformation of the backbone of loop being studied. As such, in spite of the variety of energy minima found for the backbone conformation of L3 and H3, all of the results we have presented in this paper encourage the application of this or similar techniques to hypervariable loops in antibodies whose sequence is known but whose three-dimensional structure is unknown. In this paper, we did not consider the conformations available to the two longest hypervariable loops in MCPC603 (L1 and H3, which have 17 and 19 amino acids, respectively), owing to the prohibitive computational requirements of minimizing of a sufficiently large number of starting conformations to span the conformational space available to these loops even on a modern array processor such as the Star ST-100. However, as future supercomputers and special-purpose processors such as the FASTRUN<sup>39</sup> device presently under construction for our laboratory come on line, it should be possible to apply techniques such as those presented here to longer loops, and to consider alternative side chain conformations earlier in the search for energetically allowable conformations. A pictorial summary of our present methods and results is given in Figure 6c, d, which shows 100 randomly generated starting conformations with van der Waals screening against the entire molecule for each of the four CDRs studied, as well as the five lowest-energy structures for each loop found by means of molecular dynamics followed by minimization in CHARMM/GEMM.

## ACKNOWLEDGMENTS

This work was supported by NIH grant RR-00442 and an NSF equipment grant DMB-84-0296. We wish to thank Dr. Martin Karplus of Harvard for CHARMM; Dr. Bernard Brooks of the NIH for GEMM; the Biosym corporation for DISCOVER; and Dr. Ray Hagstrom of Argonne National Laboratory for several microcoded subroutines for the STAR ST-100 which are presently installed in the DISCOVER program. We also thank Mark Reboul for programming assistance and Nicholas Necles for photographic assistance at Columbia.

## NOTE ADDED IN PROOF

Recently Brucoleri and Karplus<sup>40</sup> reported a method for predicting protein loop conformations which, like that of Moult and James, relies on a grid

search for most of the backbone dihedral angles of the loop. The last six dihedrals are, however, determined analytically using a method based on Go and Sheraga<sup>41</sup>. The initial grid points are chosen from allowed regions of the Ramachandran map. Side chain torsional conformations are further considered. The loop generation method is rapid, and the authors show that it is possible to find native conformations of various protein fragments in this manner; however, they do not perform exhaustive minimization on large numbers of generated structures. In contrast, we have attempted to demonstrate that it is possible to saturate the conformational space available to the backbone degrees of freedom of a loop, at least to the extent that low energy minima are multiply populated. We have found that in some cases, conventional force-fields predict the native conformation using this approach even in the absence of side chains.

In addition, a grid search grows exponentially in the number of degrees of freedom of the system, and therefore, as these authors themselves point out, their method (like the method of Moult and James) is applicable to relatively short loops only. Our strategy for generation is in principle applicable to fragments of any length, although of course the number of conformations which must be generated and minimized in order to multiply populate the low energy minima increases dramatically regardless of generation strategy as the number of degrees of freedom in the system increases.

## REFERENCES

- Wu, T.T, Kabat, E.A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* 132:211-250, 1970.
- Kabat, E., Wu, T., Bilofsky, H., Reid-Miller, M., Perry, H.: "Sequence of Proteins of Immunological Interest." Washington, DC: U.S. Department of Health and Human Services, National Institute of Health, 1983.
- 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. USA* 71:4298-4302, 1974.
- Yoshinori, S., Cohen, G.H., Padlan, E.A., Davies, D.R. Phosphocholine binding immunoglobulin fab McPC603: An x-ray diffraction study at 2.7 Å. *J. Mol. Biol.* 190:593-604, 1986.
- 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.
- 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.
- 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.
- Epp, O., Latham, 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.
- Amit, A.G., Mariuzza, R.A., Phillips, S.E.V., Poljak, R.J. The three dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* 233:747-753, 1986.
- Suh, S.W., Bhat, T.N., Navia, M.A., Cohen, G.H., Rao, D.N., Rudikov, S., Davies, D.R. The galactan-binding immunoglobulin FabJ539: An x-ray diffraction study at 2.6 Å° resolution. *Proteins* 1:74-79, 1986.
- Richardson, J. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34:167-339, 1981.
- Amzel, L.M., Poljak, R.J. Three dimensional structure of immunoglobulins. *Annu. Rev. Biochem.* 48:961-997, 1979.
- Davies, D.R., Metzger, H.A. Structural basis of antibody function. *Annu. Rev. Immunol.* 1:87, 1983.
- Lesk, A.M., Chothia, C. Evolution of proteins formed by beta-sheets II. The core of immunoglobulin domains. *J. Mol. Biol.* 160:325-342, 1982.
- Chothia, C., Novotny, J., Brucolari, R., Karplus, M. Domain association in immunoglobulin molecules: The packaging of variable domains. *J. Mol. Biol.* 186:651-663, 1985.
- Novotny, J., Brucolari, R., Newell, J., Murphy, D., Haber, E., Karplus, M. Anatomy of the antibody binding site. *J. Biol. Chem.* 258:14433-14437, 1983.
- Padlan, E.A., Davies, D.R., Pecht, I., Givol, D., Wright, C. Model-building studies of antigen-binding sites: The haptoglobin-binding site of MOPC-315. *Cold Spring Harbor Symp. Quant. Biol.* 41:627-637, 1976.
- 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.
- 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.
- Pawlita, N., Mushinkska, E., Feldman, R.J., Potter, M. A monoclonal antibody that defines an idiotope with two subsites in galactan-binding myeloma proteins. *J. Exp. Med.* 154:1946-1956, 1981.
- de la Paz, P., Sutton, B.J., Darsley, M.J., Rees, A.R. Modeling of the combined sites of anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. *EMBO J.* 5:415-425, 1986.
- Rees, A.R., de la Paz, P. Investigating antibody specificity using computer graphics and protein engineering. *Trends Biol. Sci.* 11:144-148, 1986.
- Chothia, C., Lesk, A., Levitt, M., Amit, A., Mariuzza, R., Phillips, V., Poljak, R. The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. *Science* 233:756-758, 1986.
- Shenkin, P.S., Yarmush, D.L., Fine, R.M., Levinthal, C. Predicting antibody hypervariable loop conformation I. Ensembles of random conformations for ring-like structures. (submitted).
- 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.
- DISCOVER is the name of a program available from the Biosym corporation, San Diego, California. The force field is described in Hagler, A. Potential functions in proteins, molecular dynamics and drug design. In: "Molecular Dynamics and Protein Structure." Hermans, J., ed. Western Springs, IL. 1984. pp. 133-139.
- Brooks, B.R., Brucolari, 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.
- 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).
- Polak, E., Ribiere, G. Note sur la Convergence des Directions Conjugue'es. *Rev. Fr. Inf. Rech. Oper.* 16:35-43, 1969.
- Array Processor Control Language (APCL) is a proprietary high level language from Star Technologies, Inc. used to program the Star ST-100 array processor.
- Katz, L., Levinthal, C. Interactive computer graphics and representation of complex biological structures. *Ann. Rev. Biophys. Bioeng.* 1:465-504, 1972. The force field parameters are those of Hagler, A.T., Huler, E., Lifson, S. Energy

- functions for peptide and proteins. *J. Am. Chem. Soc.* 96:5319–5327, 1974.
31. Gilson, M., Rashin, A., Fine, R., Honig, B. On the calculation of electrostatic interactions in proteins. *J. Mol. Biol.* 183:503–516, 1985.
  32. Klapper, I., Hagstrom, R., Fine, R., Sharp, K., Honig, B. focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino acid modification. *Proteins* 1:47–59, 1986.
  33. Rashin, A., Gilson, M., Fine, R., Honig, B. On the treatment of electrostatic interactions in molecular dynamics. In: "Molecular Dynamics and Protein Structure." Hermans, J., ed. Western Springs, IL: Polycrystal Book Service. 1984:26–30.
  34. C. Chothia. Hydrophobic bonding and accessible surface area in proteins. *Nature* 248:338–339, 1974.
  35. Rashin, A. Buried surface area, conformational entropy, and protein stability. *Biopolymers* 23:1605–1620, 1984.
  36. Vita, C., Palvoppo, D., Fontana, A., Rashin, A. Independent folding of carboxyl terminal fragment of thermolysin; Appendix: Prediction of the stabilities of thermolysin. *Biochemistry* 23:5512–5519, 1985.
  37. Shrake, A. and Rupley, J.A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79:351–371, 1973.
  38. Moult, James, M.N.G. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins* 1:146–163, 1986.
  39. Levinthal, C., Fine, R., Dimmler, G. FASTRUN: A special-purpose hard-wired computing device for molecular mechanics. In: "Molecular Dynamics and Protein Structure." Jan Hermans, ed. Western Springs, IL: Polycrystal Book Service. 1984:126–132.
  40. Brucoleri, R.E., and Karplus, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 26:137–196, 1987.
  41. Go, N. and Sheraga, H.A. Ring closure and local conformational deformation of chain molecules. *Macromolecules* 3:178–187, 1970.