Multiple copy sampling in protein loop modeling: Computational efficiency and sensitivity to dihedral angle perturbations

QIANG ZHENG,¹ RAKEFET ROSENFELD,² CHARLES DELISI,² AND DONALD J. KYLE¹

¹ Division of Structural Biology, Scios Nova Inc., 6200 Freeport Center, Baltimore, Maryland 21224

² Department of Biomedical Engineering and Biomolecular Engineering Research Center, College of Engineering,
Boston University, Boston, Massachusetts 02215

(RECEIVED October 7, 1993; ACCEPTED December 17, 1993)

Abstract

Multiple copy sampling and the bond scaling-relaxation technique are combined to generate 3-dimensional conformations of protein loop segments. The computational efficiency and sensitivity to initial loop copy dispersion are analyzed. The multicopy loop modeling method requires approximately 20–50% of the computational time required by the single-copy method for the various protein segments tested. An analytical formula is proposed to estimate the computational gain prior to carrying out a multicopy simulation. When 7-residue loops within flexible proteins are modeled, each multicopy simulation can sample a set of loop conformations with initial dispersions up to $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side-chain rotatable dihedral angles. The dispersions are larger for shorter and smaller for longer and/or surface loops. The degree of convergence of loop copies during a simulation can be used to complement commonly used target functions (such as potential energy) for distinguishing between native and misfolded conformations. Furthermore, this convergence also reflects the conformational flexibility of the modeled protein segment. Application to simultaneously building all 6 hypervariable loops of an antibody is discussed.

Keywords: antibody modeling; loop closure; multicopy sampling; scaling-relaxation

The prediction of the 3-dimensional conformation of protein loop segments is of wide interest in the study of protein structure and function. The objective of loop modeling is to find low-energy conformations for segments subject to a distance constraint between terminal atoms, within a framework of known protein structure (Go & Scheraga, 1970; Weiner et al., 1983; Brucoleri & Karplus, 1985; Shenkin et al., 1987). Examples include the deletions and insertions in homologous proteins (Alwyn & Thirup, 1986; Greer, 1990; Summers & Karplus, 1990; Palmer & Scheraga, 1991), hypervariable loops of immunoglobulins (Chothia & Lesk, 1987; Brucoleri et al., 1988), and the intra- and extracellular loops of G-protein coupled receptors (Gether et al., 1993; Kyle et al., 1994). A loop-closure method usually treats the geometric and energetic requirements separately by first generating conformations with the proper end-to-end distance and subsequently evaluating their energy.

The recently proposed bond scaling-relaxation method for loop closure meets the geometric and energetic requirements simultaneously (Zheng et al., 1993b, 1993c). The interaction be-

tween the loop segment and its protein environment is included throughout the scaling-relaxation process, resulting in accurate predictions. The conformational sampling of this method has been enhanced (Rosenfeld et al., 1993) by combining it with multiple copy sampling techniques (Elber & Karplus, 1990; Miranker & Karplus, 1991), i.e., multiple copies of a loop segment with random initial conformations were generated and subsequently scaled and relaxed simultaneously. During relaxation, the protein atoms in the framework (i.e., outside the modeled loop segment) respond to the average force from all loop copies, whereas each loop copy responds to the full force from the framework but feels no forces from other copies. In this way, m simulations of a loop-framework complex are approximated by a single simulation of a complex containing m loop copies, attached to a single framework. The computational efficiency of the method is achieved by calculating the framework internal interactions only once for all m loop copies, thereby replacing m different frameworks with a single average framework. This procedure is extended in this work to model proteins with multiple loop regions where, in each region, multiple copies of loops are generated simultaneously. The copies in a given region are transparent to one another, but each feels the average force exerted by copies in each of the other regions.

Reprint requests to: Qiang Zheng, Division of Structural Biology, Scios Nova Inc., 6200 Freeport Center, Baltimore, Maryland 21224; e-mail: qiang@jhunix.hcf.jhu.edu.

Applications of multicopy sampling techniques include studies on reaction pathways of ligand diffusion (Elber & Karplus, 1990; Czereminski & Elber, 1991; Verkivker et al., 1992), protein side-chain placement (Roitberg & Elber, 1991), receptor mapping (Miranker & Karplus, 1991; Caflisch et al., 1993), and peptide docking to cellular receptors (Rosenfeld et al., 1993). In addition, the theoretical basis of the dynamics (Ulitsky & Elber, 1993) and the statistical mechanics (Straub & Karplus, 1991; Zheng et al., 1993a) of multicopy method have been studied.

Multiple copy sampling techniques were initially introduced to increase computational efficiency by simultaneously sampling a wide range of ligand conformations in complex with a single protein (Elber & Karplus, 1990). This can be achieved if the protein is not affected, or if it is only slightly perturbed, by interactions with the ligand. To prevent the protein from being distorted by interactions with the ligand copies having very different conformations and locations, the protein is often treated as rigid (Miranker & Karplus, 1991; Caflisch et al., 1993). However, by using a rigid protein framework, there is no inherent advantage of multicopy sampling because the framework internal interactions are excluded. Furthermore, rigidity can also have undesirable effects because ligand-protein and protein-protein interactions can be important in determining the local conformation of the ligand binding site on the protein (Shoichet & Kuntz, 1991; Rosenfeld et al., 1993).

In multicopy loop modeling the loop is treated as a ligand, and the remaining protein as framework. When several loop regions in a protein are modeled simultaneously, the interactions between them can be important. Consider, for example, the antigen binding site of an antibody, which consists of 6 hypervariable loop regions interspersed within a conserved framework. The effective framework for each loop region consists of the conserved framework plus the other 5 loop regions, which generally can be treated as neither rigid nor as a very small perturbation from a rigid structure. Because the multiple loop regions respond to the average force exerted by all copies in other regions, inaccuracies in one region can amplify inaccuracies in the others. The problem that arises in such multiple loop region situations is essentially one of optimization: on the one hand, loop dispersions within a given region must be small enough to correspond to a physically meaningful average conformation for the loop region (Zheng et al., 1993a); on the other hand, limiting the loop dispersion decreases sampling efficiency, thereby diminishing the advantages of the multicopy method. The average conformation of copies is only physically meaningful when all copies converge during a simulation. The major objective of this work is to determine the maximum dispersions of initial conformations attainable before convergence diminishes.

In the following a systematic analysis of the computational efficiency of multicopy loop closure is first presented, and a method for calculating the optimal number of copies for a given system is proposed. The following sections are aimed at determining the maximum conformational dispersion of initial conformations for which convergence is observed during multicopy relaxation when only 1 loop region in the protein is being modeled and the remainder is treated as framework (this single loop region problem is in contrast to the multiple-loop region problem introduced above). We then show that the degree of convergence of multiple loop copies can be used, in conjunction with energy functions, as an effective target with which the experimentally observed conformation can be identified. In the last

section a preliminary analysis of the applicability of the proposed method for simultaneous modeling of the 6 hypervariable loops of an immunoglobulin is presented. Although the present work focuses on protein loop modeling, the results are generally applicable to a range of problems in molecular modeling and drug design.

Methods

The multicopy scaling-relaxation method for loop modeling consists of 3 steps. (1) For each loop region, multiple initial conformations are generated by assigning random values to all rotatable backbone and side-chain dihedral angles in a loop. All bond lengths of the copies are subsequently scaled to meet the distance constraint between the first N-terminal amine nitrogen and the last C-terminal carbonyl carbon in the loop segment, as required for proper connection to the framework. The scaled copies of loop regions are then patched to the framework via terminal peptide bonds, resulting in a loop-framework complex, as shown in Figure 1A. The equilibrium bond lengths for each loop are scaled accordingly. (2) The complex is relaxed via en-

B

Multicopy modifications to the force and energy calculation

- a_i^p : an atom in the i^{th} loop region and p^{th} copy.
 $N(i)$: number of copies in loop region i ; $N=1$ for the framework.
 $F(a_i^p, b_j^q)$: normal interaction force on atom a_i^p from atom b_j^q .
 $E(a_i^p, b_j^q)$: normal interaction energy between atoms a_i^p and b_j^q .

$$\begin{aligned}
 F(a_i^p, b_j^q) &\rightarrow \lambda_F \cdot F(a_i^p, b_j^q) \\
 E(a_i^p, b_j^q) &\rightarrow \lambda_E \cdot E(a_i^p, b_j^q)
 \end{aligned}$$

with $\lambda_F = 1$ and $\lambda_E = 1/N(i)$ if $i=j$ and $p=q$
 $\lambda_F = \lambda_E = 0$ if $i=j$ and $p \neq q$
 $\lambda_F = 1/N(j)$ and $\lambda_E = 1/(N(i) \cdot N(j))$ if $i \neq j$

Fig. 1. A: Backbone trace of a sample loop-framework complex of protein Zif268. The complex contains 2 loop regions (red), each containing 10 loop copies. **B:** Force and energy modifications in the multicopy scaling-relaxation method.

ergy minimization. During relaxation, the equilibrium bond lengths are gradually restored while the terminal loop atoms as well as the framework atoms are constrained to their known positions with a force constant of 5 kcal/mol·Å. When more than 1 loop region is modeled, the copies in a given region are transparent to one another, but each feels the average force exerted by copies in each of the other regions. Thus the loop-loop interaction force on a copy in region i is obtained by averaging the interactions contributed by region j , and then summing over j . The forces and energies are modified as illustrated in Figure 1B. (3) Energy and convergence of the relaxed copies are analyzed.

The multicopy scaling-relaxation method has been incorporated into CHARMM (version 22 from Molecular Simulation Inc.; Brooks et al., 1983). All calculations were carried out with the following specifications. The All Hydrogen parameters were used with a nonbonded cutoff distance of 10 Å and a distance-dependent dielectric of $4R$, where R is the distance between 2 interacting atoms. Minimizations were performed using the adopted basis Newton Raphson routine in CHARMM. A minimization procedure was terminated when the energy drop in 50 steps of minimization was less than 2.0 kcal/mol. The native conformations used in this work are the crystal conformations refined by this minimization procedure. The RMS deviation (RMSD) between 2 loop conformations was calculated for backbone heavy atoms *without rotation and translation*. All calculations were carried out on Silicon Graphics workstations either 4D/480 or 4D/240.

Results

Computational efficiency

The computational time required for energy minimization, the most time-consuming part of the multicopy scaling-relaxation method, can be approximately estimated to be proportional to the product of 2 factors, (1) the number of minimization steps $S_{\min}$, and (2) the number of operations in an energy calculation. For the minimization procedure used in this work, $S_{\min}$ depends on the total number of atoms in a complex. The most costly step in an energy evaluation is the calculation of the nonbonded interaction energy. In what follows, the number of nonbonded interactions of m single-loop simulations performed sequentially is compared with the corresponding number of a

single m -loop simulation. For ease of explanation the discussion in this section is confined to proteins containing a single loop region, though the generalization to multiple loop regions is straightforward.

For a multicopy complex consisting of m copies of a single loop region having n atoms and a framework of N atoms, the number of nonbonded interactions is approximately $N^2/2 + mNn + mn^2/2$. Similarly, the corresponding number for a normal (single copy) complex is $N^2/2 + Nn + n^2/2$. The computational efficiency of the multicopy method is measured by the ratio η of the simulation time required by an m -copy complex and that required by m normal complexes:

$$\eta = \frac{S_{\min}(mn + N)(N^2/2 + mnN + mn^2/2)}{mS_{\min}(n + N)(N^2/2 + nN + n^2/2)} \sim \frac{S_{\min}(mn + N)}{S_{\min}(n + N)} \left(\frac{1}{m} + \frac{2n}{N} \right) \quad (1)$$

for $N \gg n$ and $m \gg 1$.

Because $S_{\min}$ is almost a constant (150–200 steps) for all the loop-framework complexes studied in this section (Table 1) and $m \leq 20$, the above equation can be approximately reduced to

$$\eta \sim \frac{1}{m} + \frac{2n}{N}. \quad (2)$$

The above estimation of η implicitly assumes that all possible pairwise interactions are counted. With a nonbonded cutoff usually used in molecular modeling the numbers of protein-protein and ligand-protein interactions are NM and nM , respectively, where M is the average number of atoms within the cutoff radius. However, the final approximation of η in Equation 2 remains the same.

Two conclusions can be drawn from Equation 2: (1) For a given loop-protein complex, the efficiency of the multicopy method cannot be indefinitely improved by increasing the number of copies, because the efficiency is bound by $2n/N$. As a corollary, the optimal number of copies is approximately $N/2n$. (2) The multicopy method is most efficient when the loop segment is small compared to the framework. These conclusions are not restricted to loop-framework complexes and should apply to any system in which a small subsystem is multiple copied, such as ligand-protein complexes.

Table 1. Protein segments modeled

Segment class	Protein name	PDB name	Resolution (Å)	Loop residue range	Ratio of atoms $2n/N$	Sequence (1-letter code)
α -Helix	Zif268 zinc finger domain	1zaa	2.1	16–22	0.48	RSDELTR
	Myoglobin	1mba	1.6	91–97	0.11	QFAKEHV
β -Strand	Zif268 zinc finger domain	1zaa	2.1	11–17	0.51	DRRFSRS
	Bovine pancreatic trypsin inhibitor (BPTI)	8pti	1.8	18–24	0.37	IIRYFYN
Surface loops	Rubredoxin	7rxn	1.5	16–22	0.25	AEGDPDN
	Immunoglobulin (MCPC603)	1mcp	2.7	56–62	0.12	GASTRES
	Immunoglobulin (Bence-Jones)	2rhe	1.6	51–57	0.14	YNDLLPS
	Myoglobin	1mba	1.6	119–125	0.08	SVAAPPA

Fig. 2. Computational efficiency of the multicopy method as a function of loop copy number. Modeled loops are listed in Table 1. η is the ratio of the computational times required by a single m -copy simulation and by m single-copy simulations. The ratio of the numbers of atoms in the loop and in the framework, n/N is given in parentheses for each loop-framework complex.

To verify the above conclusions, a series of calculations was carried out for the loop-framework complexes in Table 1. For each loop, initial conformations were generated by randomly perturbing the native conformations within $\pm 30^\circ$ for backbone and $\pm 60^\circ$ for side-chain rotatable dihedral angles. Multicopy simulations with varying copy numbers were carried out and compared to corresponding single-copy simulations (Fig. 2). The multicopy simulations show a 2–5-fold decrease in computational time in comparison to the single-copy simulations, and large decreases were observed for small ratios of $2n/N$. When the number of copies was increased from 10 to 20, little additional gain in efficiency was observed. This is in agreement with Equation 2, because $1/m = 0.1$ and 0.05 , respectively, which is smaller or approximately equal to $2n/N$ ($=0.08$ – 0.51 for the loop-framework complexes modeled).

When m was further increased, the efficiency began to decline. An attempt to model 99 copies of 1zaa 11–17 took about twice the time required to model 99 corresponding single-loop complexes. This is because $S_{\min}$ cannot be approximated as a constant for large m 's: $S_{\min}(m=99) \sim 6 \cdot S_{\min}(m=1)$. For this reason, all multicopy simulations hereafter were carried out with 10 copies. Our studies on peptide docking to cellular receptors (Rosenfeld et al., 1993) were also performed with 10 peptide copies, which is in accord with the above estimation (in this case, $N \sim 3,000$, $n \sim 150$, and $N/2n = 10$).

The relationship between the efficiency and the ratio of atom numbers of the loop and the framework, n/N , is plotted in Figure 3. The asymptotic relation when $n/N \ll 1$ can be clearly approximated by a linear function of n/N ($\eta \sim 1.93n/N$).

These results suggest that Equations 1 and 2 provide a simple way to estimate the computational gain in using the multicopy method prior to carrying out the simulations. For example, the use of 100 copies of a 2-atom ligand, CO, by Czerminski and Elber (1991) to simulate its diffusion through the protein leg-hemoglobin ($\sim 1,500$ atoms) is computationally 100 times more efficient than 100 single-ligand simulations, because $\eta = 1/m + 2n/N \sim 0.01$.

Perturbations of native conformations of single loop regions

To determine the maximum dispersions of initial loop conformations attainable before convergence starts to diminish, 8–7-residue loops were modeled (Table 1). These were chosen to represent a wide variety of protein conformational types in medium-size loop segments. The dispersion in the initial loop

copies is defined in terms of the range of angles allowed for rotatable backbone and side-chain dihedral angles (illustrated in Fig. 4 for backbone dihedral angles). Ten random initial copies were generated for each loop by perturbing the native conformation, followed by scaling and relaxation. The convergence and energy of loop copies were then analyzed. A relaxed copy is considered to be in the same cluster as the unperturbed conformation (either a native or a predicted conformation), if the backbone RMSD between the two is less than $1.0 \text{ \AA}$. The procedure was repeated with increasing perturbations until no convergence of the copies was observed.

Although the choice of a $1.0\text{-\AA}$ cutoff indicates when a particular copy is in the same cluster as the original unperturbed conformation, one will see below that in order to use convergence as a target function for evaluating the resulting conformations, one needs to be specific about when the group of loop copies as a whole is considered to form a cluster. The decision is obvious at the 2 extremes, when either all 10 copies are within $1 \text{ \AA}$ of the unperturbed conformation, or none is within $1 \text{ \AA}$. The preliminary analysis in this work will follow a conservative course and consider a cluster to be present after relaxation only if all members of the group are within $1 \text{ \AA}$ of the unperturbed conformation.

By varying the amplitude of conformational perturbations in the initial copies and monitoring convergence by energy minimization, one can probe the width of the energy minimum. The greater the dispersion that will lead the copies to converge, the wider the minimum. Knowing this maximum dispersion attain-

Fig. 3. Computational efficiency of the multicopy method as a function of the ratio of the numbers of atoms in the loop and framework, n/N . Loops used are listed in Table 1. η is the ratio of the computational times required by a single 10-copy simulation and by 10 single-copy simulations. Maximal efficiency was observed for minimal n/N .

Fig. 4. Illustration of the backbone dihedral sampling enhancement in multicopy sampling. A single sampling point (ϕ, ψ) by the single-copy method is expanded to an area $(\phi \pm \Delta\phi, \psi \pm \Delta\psi)$; $\pm\Delta\phi$ and $\pm\Delta\psi$ are the dihedral angle dispersions. In this work, $\Delta\phi$ and $\Delta\psi$ are set to be equal over the entire modeled loop segment.

able before convergence starts to diminish is important, because it provides valuable information on the conformational entropy of the various minima found (Rosenfeld et al., 1993).

The results of 1 series of perturbations for the 1zaa 11–17 loop are shown in Table 2. All pairwise RMSDs of initial, or relaxed copies form a symmetric matrix, of which only the upper triangle is shown. The numbers in the top row of each matrix are the RMSDs of each copy from the native conformation. The energies of each of the relaxed copies are listed above the corresponding matrices. This energy is the sum of the normal (i.e., without using the scalings in Fig. 1B) framework internal energy, the loop–framework interaction energy and the loop internal energy. The number in the i^{th} row from top and the j^{th} column from left ($2 \leq i, j \leq 10$) in all matrices is the backbone RMSD between the $(i-1)^{\text{th}}$ and j^{th} copy. For dispersions up to $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side-chain dihedral angles, all 10 copies fell into the same cluster with the native conformation. For dispersions of $\pm 20^\circ$ for backbone and $\pm 40^\circ$ for side chains, 2 copies are very far from the native conformation after relaxation. Therefore, in this sample of copies the maximum dispersions attainable before convergence starts to diminish are $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side chains.

Although energy minimization procedures are very local, averaging the interaction energies of the loop copies with the framework results in some smoothing of the energy surface (Roitberg & Elber, 1991). This leads to larger movement of the copies than would be expected from a single-copy minimization. For example, the RMSD between the third and sixth copies of $\pm 15^\circ$ — $\pm 30^\circ$ dispersions was 4.4 Å before relaxation and only 0.8 Å after.

Of additional importance is an energy–RMSD correlation of the relaxed copies within each matrix. The copies that did not converge, or converged but with large RMSD tend to have higher energy. The 2 unconverged copies with dispersions of $\pm 20^\circ$ for backbone and $\pm 40^\circ$ for side chains have significantly high energies (2,761 and 161 kcal/mol) than the converged ones. The energy of the copy with the largest RMSD (0.8 Å) from the native conformation with dispersions of $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side chains is -176 kcal/mol, about 25 kcal/mol higher than the average energy of other 9 copies. The implication is that for low RMSD conformations, the width and depth of the associated energy minima appear to be correlated.

In contrast, no energy–RMSD correlation was observed between relaxed copies drawn from different dispersion groups (i.e., different matrices). For example, one of the copies with dispersions of $\pm 30^\circ$ for backbone and $\pm 60^\circ$ for side chains (data not shown) and the sixth copy with dispersions of $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side chains have energy values of -17 and -176 kcal/mol, respectively, whereas their RMSDs from the native conformation are 0.2 and 0.8 Å, respectively. Moreover, whereas the average energy of the converged copies increases from -203 kcal/mol for dispersions of $\pm 10^\circ$ for both backbone and side chains to -124 kcal/mol for dispersions of $\pm 20^\circ$ for backbone and $\pm 40^\circ$ for side chains, the average RMSD from the native conformation only rises from 0.1 to 0.2 Å. This demonstrates that although the energy function is sensitive to small distortions of covalent bonds and angles, as well as to atomic overlaps, the tendency to converge is relatively robust against such differences.

The convergence results of all 8 modeled loops are shown in Table 3. For the 4 internal loops (1zaa 11–17, 1mba 91–97, 1zaa 11–17, and 8pti 18–24) the maximum dispersions are $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side chains. The results for the 4 surface loops are less definitive. Two of them (7rxn 16–22 and 2rhe 51–57) show the same maximum dispersions as the internal loops (Table 3). For 1mba 119–125, the maximum dispersions are approximately $\pm 10^\circ$ for backbone and $\pm 30^\circ$ for side chains. The relaxed RMSD of the unconverged copy in this case is 1.03 Å, only marginally above the arbitrary cutoff of 1.0 Å. For 1mcp 56–62 the maximum dispersions are approximately $\pm 10^\circ$ for both backbone and side chains. The results are approximate for a number of reasons including coarse graining of dihedral angles sampling. The initial and relaxed RMSDs of the unconverged copy in this interval are 1.95 and 1.13 Å, respectively.

The convergence of side chains as well as the overall loop copies of 1zaa 11–17 are shown in Figure 5. Most of side chains converged well, except the guanadinium groups of the 3 arginines, suggesting that these groups are flexible. This observation is consistent with our previous results on the modeling of the same loop by using the single-copy scaling-relaxation method (Zheng et al., 1993c). Because the native positions of these guanadinium groups were determined from a protein–DNA complex, in which they are in close contact with the DNA, these groups may not adopt to any particularly favored positions in the absence of DNA. This result, together with the relatively poor convergence of surface loops of 1mcp 56–62 and 1mba 119–125, suggest that the degree of convergence reflects the conformational flexibility of a native conformation.

Perturbation of predicted conformations of single loop regions

The results in the previous section assess stability against perturbation of native loop conformations. If the concept of convergence is to be useful as a predictive tool, however, one needs to carry out a similar analysis for predicted conformations. The above procedure was therefore repeated for 8 predicted conformations of the same loops used in the previous section, each chosen as the lowest-energy conformation from loops generated by using the single-copy scaling-relaxation method (Zheng et al., 1993c).

Table 2. Pairwise RMSDs (Å) of the loop copies of Zif268 11–17

Initial copies										Relaxed copies									
Backbone $\pm 10^\circ$ – sidechain $\pm 10^\circ$										–199	–207	–204	–204	–206	–205	–198	–195	–204	–203
1.0	0.8	0.8	1.7	1.1	0.8	1.4	1.7	1.2	0.7	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.2	0.1
	0.9	0.7	2.6	2.0	1.7	0.7	0.8	2.1	0.4		0.2	0.1	0.2	0.2	0.2	0.1	0.1	0.2	0.1
		1.0	1.9	1.4	1.3	0.9	1.3	1.6	0.7			0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
			2.3	1.9	1.5	1.2	1.2	1.9	0.8				0.2	0.1	0.1	0.1	0.1	0.2	0.1
				0.7	1.0	2.8	3.1	0.6	2.2					0.2	0.1	0.2	0.1	0.1	0.2
					0.6	2.2	2.6	0.4	1.6						0.1	0.2	0.1	0.1	0.2
						2.1	2.4	0.5	1.4							0.1	0.1	0.0	0.2
							0.5	2.4	0.8								0.1	0.2	0.1
								2.8	1.1									0.1	0.1
									1.8										0.1
Backbone $\pm 10^\circ$ – sidechain $\pm 30^\circ$										–199	–204	–204	–202	–206	–204	–198	–197	–203	–200
1.0	0.8	0.8	1.7	1.2	0.8	1.4	1.7	1.2	0.7	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	0.9	0.7	2.6	2.0	1.7	0.7	0.8	2.1	0.4		0.1	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.1
		1.0	1.9	1.4	1.3	0.9	1.3	1.6	0.7			0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1
			2.3	1.9	1.5	1.2	1.2	1.9	0.8				0.1	0.1	0.1	0.1	0.1	0.1	0.2
				0.7	1.0	2.8	3.1	0.6	2.2					0.1	0.1	0.1	0.1	0.1	0.1
					0.6	2.2	2.6	0.4	1.6						0.1	0.2	0.2	0.1	0.1
						2.1	2.4	0.5	1.4							0.2	0.1	0.1	0.2
							0.5	2.4	0.8								0.1	0.2	0.2
								2.8	1.1									0.1	0.1
									1.8										0.1
Backbone $\pm 15^\circ$ – sidechain $\pm 30^\circ$										–197	–207	–195	–205	–204	–176	–202	–197	–202	–207
1.5	1.1	2.4	1.7	1.2	2.4	1.8	1.0	2.6	0.8	0.1	0.1	0.2	0.1	0.2	0.8	0.2	0.1	0.3	0.1
	1.3	3.7	2.9	2.6	1.1	3.1	0.6	3.9	2.0		0.2	0.3	0.2	0.2	0.8	0.2	0.1	0.4	0.2
		2.7	1.9	2.0	1.8	2.3	0.9	2.8	1.1			0.2	0.1	0.1	0.8	0.1	0.1	0.2	0.2
			1.1	1.5	4.4	0.9	3.2	0.6	1.8				0.2	0.2	0.8	0.2	0.3	0.3	0.2
				0.9	3.6	0.6	2.4	1.1	1.2					0.1	0.8	0.1	0.1	0.2	0.1
					3.5	0.8	2.1	1.8	1.0						0.8	0.1	0.2	0.2	0.1
						4.0	1.5	4.5	2.7							0.7	0.7	0.7	0.8
							2.6	1.1	1.4								0.1	0.2	0.1
								3.5	1.5									0.3	0.1
									2.0										0.2
Backbone $\pm 20^\circ$ – sidechain $\pm 40^\circ$										–120	–127	2,761	–124	–114	161	–128	–123	–123	–129
2.0	2.2	3.1	2.4	1.3	3.4	1.1	0.5	1.2	1.4	0.2	0.2	2.2	0.2	0.2	4.9	0.3	0.1	0.1	0.1
	3.6	1.4	4.0	0.8	4.8	2.6	1.8	2.9	0.9		0.3	2.2	0.4	0.2	4.9	0.3	0.3	0.3	0.3
		4.5	0.8	3.0	1.4	1.5	2.4	1.0	2.9			2.2	0.1	0.1	4.8	0.3	0.1	0.1	0.2
			5.0	2.0	5.5	3.4	3.0	3.9	1.9				2.2	2.2	6.3	2.2	2.2	2.2	2.2
				3.3	1.5	1.8	2.6	1.3	3.3					0.2	4.8	0.3	0.1	0.2	0.1
					4.2	1.9	1.1	2.2	0.6						4.9	0.3	0.1	0.2	0.2
						2.6	3.7	2.3	4.0							4.9	4.9	4.9	4.9
							1.4	0.9	1.8								0.2	0.3	0.3
								1.5	1.3									0.1	0.1
									2.1										0.1
Backbone $\pm 30^\circ$ – sidechain $\pm 30^\circ$										–5	102	720	252	–18	4,295	348	–36	–34	–33
2.8	3.1	4.3	3.4	1.9	4.8	1.5	0.7	1.7	1.3	1.0	2.0	5.1	5.3	0.3	5.4	0.9	0.2	0.2	0.1
	4.8	2.0	5.5	1.2	6.1	3.5	2.5	3.9	2.0		1.9	5.0	5.3	0.9	5.3	1.6	1.0	1.0	1.0
		5.7	1.3	4.0	2.0	2.1	3.4	1.5	3.0			5.3	4.3	1.9	4.1	1.8	1.9	1.8	2.0
			6.5	2.7	6.6	4.5	4.2	5.2	3.3				8.7	5.1	7.6	5.0	5.1	5.1	5.2
				4.6	2.3	2.5	3.7	1.8	3.6					5.2	2.4	5.1	5.2	5.2	5.3
					5.4	2.6	1.6	3.0	1.2						5.3	1.0	0.2	0.3	0.3
						3.5	5.2	3.4	4.5							5.0	5.3	5.2	5.3
							2.1	1.2	1.6								0.9	0.8	0.9
								2.1	1.4									0.1	0.2
									2.0										0.2

As shown in Table 4, the overall convergence is similar to that observed previously for the native conformations. The maximum dispersions for 1zaa 11–17, 1zaa 16–22, 8pti 18–24, 7rxn 16–22, and 1mcp 56–62 are identical to those found by perturbing the native conformations. The maximum dispersions for

1mcp 56–62 can also be approximated as $\pm 10^\circ$ for both backbone and side chains, as before, because the initial and relaxed RMSDs are 1.83 and 1.15 Å, respectively.

Although the overall results obtained with the predicted conformations are similar to those with the native conformations,

Table 3. Perturbations of the native loop conformations

Loop segment	Dihedral dispersions	Copies converged	RMSD ^a (Å)	RMSD ^b (Å)	RMSD ^c (Å)	RMSD ^d (Å)
1zaa 11–17	±10° — ±10°	10	1.12	0.11	1.47	0.12
	±10° — ±30°	10	1.12	0.11	1.47	0.22
	±15° — ±30°	10	1.65	0.18	2.09	0.29
	±20° — ±40°	8	1.49	0.22	1.98	0.20
	±30° — ±30°	5	1.41	0.34	1.88	0.49
	±30° — ±60°	4	1.31	0.46	1.73	0.58
1zaa 16–22	±10° — ±10°	10	0.72	0.19	0.96	0.26
	±10° — ±30°	10	0.72	0.18	0.96	0.25
	±15° — ±30°	10	1.02	0.19	1.39	0.28
	±20° — ±40°	8	1.26	0.23	1.75	0.34
	±30° — ±30°	6	1.43	0.40	1.56	0.56
	±30° — ±60°	5	1.40	0.42	1.51	0.60
8pti 18–24	±10° — ±10°	10	0.62	0.09	0.89	0.09
	±10° — ±30°	10	0.62	0.10	0.89	0.10
	±15° — ±30°	10	0.86	0.23	1.18	0.31
	±20° — ±40°	10	1.02	0.16	1.34	0.20
	±30° — ±30°	8	1.25	0.35	1.44	0.49
	±30° — ±60°	8	1.25	0.42	1.44	0.56
7rxn 16–22	±10° — ±10°	10	0.67	0.33	0.90	0.43
	±10° — ±30°	10	0.67	0.25	0.90	0.31
	±15° — ±30°	10	0.93	0.40	1.22	0.51
	±20° — ±40°	8	1.05	0.60	1.33	0.75
	±30° — ±30°	7	1.37	0.54	1.25	0.51
	±30° — ±60°	6	1.38	0.53	1.14	0.54
1mcp 56–62	±10° — ±10°	9	0.88	0.17	0.76	0.24
			(1.95)	(1.13)		
	±10° — ±30°	8	0.82	0.18	0.76	0.23
	±15° — ±30°	8	1.07	0.32	0.89	0.28
	±20° — ±40°	7	1.32	0.55	1.18	0.53
	±30° — ±30°	4	1.68	0.55	0.88	0.64
2rhe 51–57	±10° — ±10°	10	0.49	0.16	0.67	0.22
	±10° — ±30°	10	0.49	0.11	0.67	0.16
	±15° — ±30°	10	0.74	0.23	0.94	0.33
	±20° — ±40°	9	0.83	0.33	1.07	0.42
	±30° — ±30°	7	0.93	0.32	1.26	0.42
	±30° — ±60°	8	1.02	0.38	1.41	0.55
1mba 91–97	±10° — ±10°	10	0.61	0.20	0.85	0.28
	±10° — ±30°	10	0.61	0.19	0.85	0.28
	±15° — ±30°	10	0.92	0.40	1.29	0.60
	±20° — ±40°	9	1.08	0.38	1.49	0.57
	±30° — ±30°	4	1.09	0.26	1.52	0.39
	±30° — ±60°	5	1.18	0.53	1.55	0.82
1mba 119–125	±10° — ±10°	9	0.77	0.43	1.05	0.62
	±10° — ±30°	9	0.77	0.39	1.05	0.56
			(1.27)	(1.03)		
	±15° — ±30°	8	0.90	0.51	1.23	0.70
	±20° — ±40°	6	0.97	0.46	1.42	0.68
	±30° — ±30°	4	1.09	0.46	1.42	0.69
	±30° — ±60°	4	1.09	0.51	1.42	0.75

^{a–d} RMSD^a and RMSD^b are the averages of the RMSDs of initial and relaxed copies from a native conformation, respectively. RMSD^c and RMSD^d are the averages of all pairwise RMSDs of initial and relaxed copies that converged, respectively. The RMSDs of the unconverged copies discussed in the text are given in parentheses.

Fig. 5. Initial (A) and relaxed (B) copies of the loop 1zaa 11–17 with dispersions of $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side-chain rotatable dihedral angles. Significant convergence was observed for the backbone atoms, the hydrophobic side chain of Phe 14, and most of the acidic groups (red).

the extent of similarity appears to depend on the accuracy of prediction (Table 4). Thus the maximum dispersions for 1mba 91–97, the least accurate prediction among the internal loops, are $\pm 10^\circ$ and $\pm 30^\circ$ for backbone and side chains, respectively, in contrast to maximum dispersions of $\pm 15^\circ$ and $\pm 30^\circ$ when the native conformation was perturbed. Similarly, the maximum dispersions of 2rhe 51–57 decrease to $\pm 5^\circ$ for backbone and $\pm 10^\circ$ for side chains (data not shown). The unconverged copy of 2rhe 51–57 with dispersions of $\pm 10^\circ$ for both backbone and side chains showed little tendency to move toward the unperturbed conformation during relaxation: its initial and relaxed RMSDs are 1.23 and 1.21 Å, respectively. This was also observed for the unconverged copy of 1mba 119–125 with dispersions of $\pm 10^\circ$ for both backbone and side chains: initial and final RMSDs are 1.63 and 1.39 Å, respectively.

Another informative variable for comparing native with predicted conformations is the degree of convergence, which is characterized in this work as the percent difference between the RMSDs before and after relaxation, averaged over all pairs of the converged copies. Despite the similarity in the maximum dispersions obtained with native and predicted conformations the percent difference in the degree of convergence generally decreases as the RMSD from the native increases (Fig. 6). For predicted loops that are within approximately 0.5 Å from the native, the degree of convergence is similar to that of the native conformations and then drops noticeably as the RMSD increases to 1.2 Å. The average degrees of convergence of the loops gen-

erated from the predicted conformations of 2rhe 51–57 and 1mba 119–125, both having RMSDs of greater than 1 Å, are approximately 40–60% lower than that of the native conformations. This shows that the degree of convergence may be a useful indicator in distinguishing native and nearby minimum-energy conformations.

Dependence of maximum dispersions on the number of loop copies

The maximum dispersions defined in the previous sections were determined with rather sparse sampling (10 loop copies) for optimal computational efficiency. The objective of this section is to examine in detail whether or not these dispersions were obtained by chance and to test their dependence on the number of loop copies. Loop 1zaa 11–17 was used for this study.

To address the first question, 10 separate runs with 10 copies were performed with dispersions of $\pm 10^\circ$, $\pm 15^\circ$, and $\pm 30^\circ$ for backbone and $\pm 30^\circ$ for side chains. The results are shown in Table 5. For both the native and predicted conformations, the smallest perturbations consistently resulted in full convergence. For the larger perturbations, in 4 of the 10 runs, 1 copy did not converge, whereas the other 6 runs fully converged.

To address the second question, the same loop was modeled with increasing number of loop copies. For dispersions of $\pm 10^\circ$ for backbone and $\pm 30^\circ$ for side chains, all copies in all runs converged without exception (Table 5), suggesting that the maxi-

Table 4. *Perturbations of the predicted loop conformations*

Loop segment	RMSD ^a (Å)	Dihedral dispersions	Copies converged	RMSD ^b (Å)	RMSD ^c (Å)	RMSD ^d (Å)	RMSD ^e (Å)
1zaa 11–17	0.47	±10° — ±10°	10	1.15	0.14	1.45	0.15
		±10° — ±30°	10	1.15	0.26	1.45	0.22
		±15° — ±30°	10	1.76	0.34	2.20	0.29
		±20° — ±40°	8	1.68	0.35	2.32	0.30
		±30° — ±30°	6	1.57	0.44	2.72	0.55
		±30° — ±60°	5	1.33	0.35	1.85	0.27
1zaa 16–22	0.12	±10° — ±10°	10	0.69	0.17	0.93	0.19
		±10° — ±30°	10	0.69	0.21	0.93	0.26
		±15° — ±30°	10	0.96	0.36	1.30	0.45
		±20° — ±40°	8	1.11	0.32	1.49	0.39
		±30° — ±30°	6	1.44	0.40	1.52	0.42
		±30° — ±60°	5	1.39	0.37	1.51	0.45
8pti 18–24	0.19	±10° — ±10°	10	0.62	0.08	0.90	0.09
		±10° — ±30°	10	0.62	0.10	0.90	0.10
		±15° — ±30°	10	0.86	0.10	1.19	0.14
		±20° — ±40°	9	0.91	0.17	1.27	0.26
		±30° — ±30°	7	1.05	0.33	1.21	0.51
		±30° — ±60°	8	1.22	0.55	1.39	0.76
7rxn 16–22	0.54	±10° — ±10°	10	0.66	0.42	0.88	0.38
		±10° — ±30°	9	0.58	0.41	0.76	0.52
		±15° — ±30°	10	0.97	0.51	1.29	0.58
		±20° — ±40°	9	1.06	0.66	1.38	0.78
		±30° — ±30°	8	1.45	0.74	1.57	0.69
		±30° — ±60°	4	1.25	0.78	1.18	0.57
1mcp 56–62	0.36	±10° — ±10°	9	0.86	0.16	0.70	0.20
				(1.83)	(1.15)		
		±10° — ±30°	8	0.80	0.14	0.67	0.16
		±15° — ±30°	8	1.13	0.18	0.90	0.20
		±20° — ±40°	6	1.35	0.34	0.71	0.35
		±30° — ±30°	5	1.62	0.48	1.16	0.37
2rhe 51–57	1.09	±30° — ±60°	4	1.59	0.65	1.06	0.67
		±10° — ±10°	9	0.55	0.40	0.76	0.56
				(1.23)	(1.21)		
		±10° — ±30°	9	0.55	0.40	0.76	0.55
		±15° — ±30°	7	0.69	0.42	0.96	0.57
		±20° — ±40°	8	0.82	0.51	1.14	0.67
1mba 91–97	0.82	±30° — ±30°	7	0.94	0.49	1.30	0.60
		±30° — ±60°	7	0.94	0.59	1.30	0.72
		±10° — ±10°	10	0.61	0.26	0.84	0.35
		±10° — ±30°	10	0.61	0.29	0.84	0.37
		±15° — ±30°	9	0.84	0.51	1.20	0.75
		±20° — ±40°	7	1.05	0.54	1.46	0.70
1mba 119–125	1.19	±30° — ±30°	3	1.08	0.57	1.10	0.44
		±30° — ±60°	3	1.08	0.57	1.10	0.53
		±10° — ±10°	9	0.80	0.45	1.04	0.65
				(1.63)	(1.39)		
		±10° — ±30°	8	0.66	0.39	0.86	0.60
		±15° — ±30°	7	0.84	0.51	0.92	0.70
		±20° — ±40°	8	1.10	0.54	1.18	0.65
		±30° — ±30°	4	1.48	0.60	1.50	0.85
		±30° — ±60°	2	1.38	0.78	1.35	0.71

^{a–c} RMSD^a is between a predicted and its corresponding native conformations. RMSD^b and RMSD^c are the averages of the RMSDs of initial and relaxed copies from a predicted conformation, respectively. RMSD^d and RMSD^e are the averages of all pairwise RMSDs of initial and relaxed copies that converged, respectively. The RMSDs of the unconverged copies discussed in the text are given in parentheses.

Fig. 6. Percent decrease of the degree of convergence for perturbations about the native and the predicted conformations for the loops described in Table 1. The degree of convergence is defined as the percent difference between the RMSDs before and after relaxation, averaged over all pairs of the converged copies. The percent decrease of the degree of convergence for a predicted loop conformation is defined as the percent difference between the degree of convergence of the predicted and the native conformation. Less well-predicted (large RMSD) conformations corresponded to larger decrease of convergence.

imum dispersions are at least $\pm 10^\circ$ for backbone and $\pm 30^\circ$ for side chains. For dispersions of $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side chains, most of the copies converge ($\geq 91\%$).

These results demonstrate that the maximum dispersions for loop 12aa 11–17 can be approximated as $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side chains as determined with 10-copy perturbations in the previous sections. This conclusion is further supported by the significantly lower convergence (22–61%) for larger dispersions ($\pm 30^\circ$ for both backbone and side chains, Table 5). Moreover, because the percentage of unconverged copies does not vary consistently with the number of copies (Table 5), the maximum dispersions determined with 10-copy perturbations in the previous section are not fortuitous.

It is important to note that the determined maximum dispersions are at best approximate. There are numerous uncertainties involved in the measurement of divergence. Given the complexity of the CHARMM energy function and the modeled segments containing over 100 atoms, it is difficult to measure accurately the “size” of an energy minimum in conformational space with finite copy samplings. Such uncertainties are observed from the incomplete convergence in 4 of the 10 runs of the 10-copy perturbations (Table 5). These results suggest that our criteria for defining a cluster are perhaps too stringent. Moreover, the maximum dispersions may also be affected by the em-

pirical nature of the CHARMM energy function and the distant-dependent dielectric used in this work.

Perturbations of single loop regions with various lengths

The 6 hypervariable loops of the immunoglobulin MCPC603 (Table 6) were chosen to study how the maximum dispersions vary with the sequence length of a loop. The above procedure was applied to each loop, and the results are shown in Table 7. The shortest loop, H1 (5 residues), has the largest maximum dispersions: $\pm 20^\circ$ for backbone $\pm 40^\circ$ for side chains. The maximum dispersions for H2 (9 residues) are $\pm 10^\circ$ for both backbone and side chains. The longest loop L1 (12 residues) showed little convergence for dispersions as small as $\pm 1^\circ$ for both backbone and side chains. These results demonstrate that the maximum dispersions decrease with increasing loop size.

A closer examination of the L1 loop copies revealed that it contains a region (the first 4–5 residues) that converged well, followed by a hair-pin loop in which no convergence was observed (Fig. 7). The side chain of Leu 29 in all copies converged to the native position as expected from its role in stabilizing the conserved part of the L1 loop (Chothia & Lesk, 1987). In order to test whether the failure to converge is due to the length of L1 loop, or to the inherent flexibility in the 7-residue hair-pin segment, this segment was modeled separately while keeping the other 5 residues at their native positions. After relaxation, 5 out of 10 copies did not converge when the initial dispersions were only $\pm 10^\circ$ for both backbone and side chains (Table 8). In contrast, there was no more than 1 unconverged copy when the same dispersions were used to perturb the native conformation of all 7-residue loops modeled in previous sections. This shows that the hair-pin loop is inherently flexible, in agreement with previous studies of the canonical structures of antibody hypervariable loops (Chothia & Lesk, 1987).

Perturbations of multiple loop regions

The objective of this section is to demonstrate that the maximum dispersions defined by modeling individual loop regions can also be applied to multiple loop regions in a protein when all the loop regions are simultaneously modeled. The set of initial conformations with the maximum dispersions generated for MCPC603 in the previous section were used. The resulting loop–framework complex containing 10 copies for each of the 6 hypervariable loop regions was then relaxed (Fig. 8). The overall convergence

Table 5. Convergence for various number of loop copies

Number of copies	Number of converged copies (percent)					
	Native conformation			Predicted conformation		
	$\pm 10^\circ$ — $\pm 30^\circ$	$\pm 15^\circ$ — $\pm 30^\circ$	$\pm 30^\circ$ — $\pm 30^\circ$	$\pm 10^\circ$ — $\pm 30^\circ$	$\pm 15^\circ$ — $\pm 30^\circ$	$\pm 30^\circ$ — $\pm 30^\circ$
10 × 10	100 (100)	96 (96)	59 (59)	100 (100)	96 (96)	61 (61)
10	10 (100)	10 (100)	5 (50)	10 (100)	10 (100)	6 (60)
25	25 (100)	24 (96)	11 (56)	25 (100)	23 (92)	10 (60)
50	50 (100)	48 (96)	29 (58)	50 (100)	49 (98)	28 (56)
99	99 (100)	90 (91)	55 (56)	99 (100)	97 (98)	22 (22)

Table 6. Hypervariable loops of MCPC603

Loop	Sequence length	Residue range	Sequence (1-letter code)
L1	12	Light chain: 26–37	SQSLNLSGNQKN
L2	7	Light chain: 56–62	GASTRES
L3	6	Light chain: 97–102	DHSYPL
H1	5	Heavy chain: 28–32	TFSDF
H2	9	Heavy chain: 50–58	ASRNKGNKY
H3	8	Heavy chain: 102–109	YYGSTWYF

is similar to that observed when each loop region was modeled separately (Table 9). With 2 exceptions (H1 and H2), all copies in each region converged to the corresponding native conformation. The RMSDs of the 2 unconverged copies are only slightly higher than the cutoff of 1.0 Å (Table 9). It is important to note that most of the Tyr and Phe side chains of L3, H1, H2, and H3, which are involved in antigen binding, converged well.

These results demonstrate that the interactions between 2 multicopied loop regions can be indeed approximated by the average interaction between the loop regions—the key assumption of the multiple copy method for loop modeling.

Discussion

A new method for constrained minimization, which combines bond scaling-relaxation with multiple copy techniques, has been analyzed. Although bond scaling-relaxation is efficient in finding minimum energy conformations of protein loops, its effectiveness in correctly identifying the native conformation is sometimes limited by a common problem; viz. the lack of reliability of an approximate energy function for choosing the native from among several low energy conformations (Novotny et al., 1988, 1989; Shoichet & Kuntz, 1991). The combined method can be used to explore efficiently the basin of energy minima, and the presented analysis suggests that the breadth of the basin provides information that can aid in identifying the native conformation.

There are 2 central considerations that bear on the effectiveness of the combined method, one related to the parameters that determine the optimal number of copies, and the other related to the parameters that affect stability against conformational variation. The question of optimal copy number has been addressed relatively unambiguously. The optimal number of copies is approximately $N/2n$, where N is the size of the protein framework, and n the size of the loop. This expression has been analytically derived and numerically verified in this work.

The concept of stability is somewhat more subtle, but it can be viewed in the following way. When a range of conformations is generated from a given conformation and is then energy minimized by multicopy method, the copies will—under certain conditions related to the magnitude of the perturbation and the stability of the unperturbed conformation—relax to the unperturbed conformation, or converge closely about it. The question of interest is how much conformational variation is allowed before convergence starts to diminish. The stability has been analyzed in a variety of contexts, including (1) native internal and

Table 7. Perturbations of the native conformations of MCPC603 loops

Loop	Dihedral dispersions	Copies converged	RMSD ^a (Å)	RMSD ^b (Å)	RMSD ^c (Å)	RMSD ^d (Å)
L1	$\pm 1^\circ$ — $\pm 1^\circ$	10	1.17	0.51	0.54	0.46
	$\pm 2^\circ$ — $\pm 5^\circ$	8	1.79	0.62	0.88	0.76
	$\pm 5^\circ$ — $\pm 5^\circ$	4	1.73	0.53	1.36	1.05
	$\pm 10^\circ$ — $\pm 10^\circ$	0	—	—	—	—
L2	$\pm 10^\circ$ — $\pm 10^\circ$	9	0.78	0.31	0.70	0.35
	$\pm 10^\circ$ — $\pm 30^\circ$	9	0.78	0.26	0.70	0.32
	$\pm 15^\circ$ — $\pm 30^\circ$	9	1.05	0.49	0.93	0.52
	$\pm 20^\circ$ — $\pm 40^\circ$	7	1.22	0.68	1.13	0.63
	$\pm 30^\circ$ — $\pm 30^\circ$	4	1.40	0.62	0.77	0.52
	$\pm 30^\circ$ — $\pm 60^\circ$	3	1.48	0.67	0.73	0.60
L3	$\pm 10^\circ$ — $\pm 10^\circ$	10	0.73	0.22	1.02	0.31
	$\pm 10^\circ$ — $\pm 30^\circ$	10	0.73	0.28	1.02	0.38
	$\pm 15^\circ$ — $\pm 30^\circ$	9	1.25	0.44	1.32	0.64
	$\pm 20^\circ$ — $\pm 40^\circ$	8	1.24	0.31	1.24	0.43
	$\pm 30^\circ$ — $\pm 30^\circ$	6	1.52	0.49	0.93	0.70
	$\pm 30^\circ$ — $\pm 60^\circ$	7	1.70	0.59	1.25	0.85
H1	$\pm 10^\circ$ — $\pm 10^\circ$	10	0.52	0.13	0.50	0.13
	$\pm 10^\circ$ — $\pm 30^\circ$	10	0.52	0.22	0.50	0.26
	$\pm 15^\circ$ — $\pm 30^\circ$	10	1.06	0.22	0.66	0.22
	$\pm 20^\circ$ — $\pm 40^\circ$	10	1.28	0.29	1.00	0.31
	$\pm 30^\circ$ — $\pm 30^\circ$	7	1.56	0.68	0.97	0.83
	$\pm 30^\circ$ — $\pm 60^\circ$	6	1.54	0.63	0.94	0.71
H2	$\pm 5^\circ$ — $\pm 5^\circ$	10	0.69	0.34	0.96	0.48
	$\pm 10^\circ$ — $\pm 10^\circ$	10	1.65	0.48	1.04	0.57
	$\pm 10^\circ$ — $\pm 30^\circ$	9	1.63	0.41	0.94	0.71
	$\pm 15^\circ$ — $\pm 30^\circ$	8	1.75	0.65	1.08	0.76
	$\pm 20^\circ$ — $\pm 40^\circ$	3	2.07	0.86	0.83	0.58
	$\pm 30^\circ$ — $\pm 30^\circ$	0	—	—	—	—
H3	$\pm 5^\circ$ — $\pm 5^\circ$	10	1.21	0.63	0.54	0.36
	$\pm 10^\circ$ — $\pm 10^\circ$	9	1.36	0.64	0.76	0.59
			(1.83)	(1.06)		
	$\pm 10^\circ$ — $\pm 30^\circ$	9	1.36	0.64	0.76	0.65
	$\pm 15^\circ$ — $\pm 30^\circ$	4	2.03	0.90	1.00	0.80
	$\pm 20^\circ$ — $\pm 40^\circ$	2	2.47	0.98	1.24	0.81
	$\pm 30^\circ$ — $\pm 30^\circ$	0	—	—	—	—

^{a-d} RMSD^a and RMSD^b are the averages of the RMSDs of initial and relaxed copies from the native conformation, respectively. RMSD^c and RMSD^d are the averages of all pairwise RMSDs of initial and relaxed copies that converged, respectively. The RMSDs of the unconverged copies discussed in the text are given in parentheses.

surface single-loop regions, (2) predicted internal and surface single-loop regions, and (3) native multiple-loop regions.

The results with single region loops demonstrate that stability depends on the size and location of the loop, the latter being most easily classified as internal or surface. Thus for the 8 7-residue loops tested, the 4 interior loops are all stable for perturbations up to $\pm 15^\circ$ for backbone and $\pm 30^\circ$ for side-chain dihedral angles. For the surface loops, 2 fell in the $\pm 15^\circ$ – $\pm 30^\circ$ category and 2 did not.

Further perspective on these results was obtained from the analysis of the 6 hypervariable loops in MCPC603. The shortest loop H1, which contains 5 residues, shows the largest maxi-

Fig. 7. Initial (A) and relaxed (B) copies of the L1 loop of the immunoglobulin MCPC603 with dispersions of $\pm 10^\circ$ for both backbone and side chains: the native conformation (red), loop copies (blue), and copies of Leu 29 side chain (yellow). The conserved region of L1 (the first 4–5 residues from the left) converged well, whereas the flexible hairpin loop did not.

imum dispersions: $\pm 20^\circ$ for backbone and $\pm 40^\circ$ for side-chain dihedral angles. The maximum dispersions for the H2 loop, which contains 9 residues, are $\pm 10^\circ$ for both backbone and side-chain dihedral angles. In general, one expects flexibility to increase with loop size, and hence the size of the maximum dispersion to decrease. These observations take on greater significance when viewed against the background of predicted loop stability, and multiple loop regions, both of which are discussed in the following sections.

The central consideration for determining whether conformational stability will be useful as a target function is the extent to which the native conformations are more stable than nearby low-energy conformations. It was suggested (Rosenfeld et al., 1993) that clustering (convergence) is sometimes a more effective target than energy when searching for the native within a set of low-energy conformations. The extent to which such a procedure is useful will depend on predicted conformations near

the native retaining a wide basin (corresponding to large maximum dispersions), which diminishes with deviation from the native. This is precisely what has been shown in the present work. The most accurate predictions have wide basins of attraction, and the less accurate predictions have slightly narrower basins of attraction. These results suggest that the degree of convergence can aid in distinguishing between native and misfolded loop conformations.

For multiple-loop regions the analysis is more complicated. Rather than a single slightly perturbable framework responding to the average force of multiple copies from a single loop region, portions of the framework will now be flexible, consisting of multiple copies of other loop regions, and each of those copies will respond to the average force of copies from other re-

Table 8. *Perturbations of the partial L1 loop: Residues 31–37*

Dihedral dispersions	Copies converged	RMSD ^a (Å)	RMSD ^b (Å)	RMSD ^c (Å)	RMSD ^d (Å)
$\pm 5^\circ - \pm 5^\circ$	10	1.17	0.51	0.69	0.57
$\pm 10^\circ - \pm 10^\circ$	5	1.79	0.62	0.86	0.65
$\pm 10^\circ - \pm 30^\circ$	4	1.73	0.53	0.95	0.67
$\pm 15^\circ - \pm 30^\circ$	1	2.03	0.69	—	—

^{a–d} RMSD^a and RMSD^b are the averages of the RMSDs of initial and relaxed copies from the native conformation, respectively. RMSD^c and RMSD^d are the averages of all pairwise RMSDs of initial and relaxed copies, respectively. The relaxed RMSDs are only for converged copies.

Table 9. *Comparison of individually and simultaneously modeled loop regions of MCPC603*

Loop	Dihedral dispersions	Copies ^a converged	Copies ^b converged
L1	$\pm 1^\circ - \pm 1^\circ$	10	10
L2	$\pm 10^\circ - \pm 30^\circ$	9 (1.13 Å)	10
L3	$\pm 10^\circ - \pm 30^\circ$	10	10
H1	$\pm 20^\circ - \pm 40^\circ$	10	9 (1.06 Å)
H2	$\pm 10^\circ - \pm 10^\circ$	10	9 (1.19 Å)
H3	$\pm 5^\circ - \pm 10^\circ$	10	10

^a Individual loop modeling. The RMSDs of the unconverged copies to the corresponding native conformations are given in parentheses.

^b Simultaneous loop modeling. The RMSDs of the unconverged copies to the corresponding native conformations are given in parentheses.

Fig. 8. Initial (A) and relaxed (B) copies of all 6 hypervariable loops of the immunoglobulin MCPC603. Shown is the backbone and the loop aromatic side chains. The framework atoms are in blue, and the colors of the loops are: L1 (red), L2 (yellow), L3 (white), H1 (green), H2 (purple), and H3 (brown). The backbone (with the exception of the flexible region of L1) and the aromatic side chains of the modeled loops converged well.

gions. When this full complexity was taken into account, small but significant differences in stability were observed, indicating isolated instabilities in certain directions within the attraction basin. Indeed, because of sparse sampling, one cannot be certain that such instabilities do not exist, even when all copies in a region fall into a single cluster. The presented analysis therefore cannot be viewed as rigorous at this stage of its development, but as a heuristic means of assessing stability and its relation to native and other nearby conformations.

The failure of flexible loop regions to converge (hypervariable loop L1 and Arg side chains of Izaa 11–17) suggests that multiple loop modeling can be utilized to assess flexibility. The alternative, and commonly used method, for assessing conformational flexibility is normal mode analysis. It is based on small perturbations and only applies in the immediate vicinity of a minimum-energy conformation, because the energy surface is usually very rough. In contrast, multicopy perturbation can explore a large region in conformational space as shown by the convergence in Figures 5 and 7. This is because the effective energy surface for a multicopy complex is smoothed by averaging ligand–protein interaction over many copies (Roitberg & Elber, 1991; Zheng et al., 1993a).

One potential application of the proposed methods is to assist experimental structure determination. For example, poor X-ray diffraction or weak NOE signals may be a result of inherently flexible regions in a protein, or artifacts. The multiple copy perturbation suggests a new approach to differentiate between these causes. Failure of the copies to converge may indicate inherent flexibility. On the other hand, convergence can be viewed as evidence for the existence of a stable conformation.

Acknowledgments

We thank Jackie Sinsko for technical assistance. Helpful comments of the reviewers are appreciated. R.R. and C.D. were supported in part by NIAID grant AI30535 award to C.D.

References

- Alwyn T, Thirup S. 1986. Using known substructures in protein model building and crystallography. *EMBO J* 5:819–822.
- Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
- Brucoleri RE, Haber E, Novotny J. 1988. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature* 335:564–568.
- Brucoleri RE, Karplus M. 1985. Chain closure with bond angle variations. *Macromolecules* 18:2767–2773.
- Cafilisch A, Miranker A, Karplus M. 1993. Multiple copy simultaneous search and construction of ligands in binding sites: Application to inhibitors of HIV-1 aspartic proteinase. *J Med Chem* 36:2142–2164.
- Chothia C, Lesk AM. 1987. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 196:901–917.
- Czermanski R, Elber R. 1991. Computational studies of ligand diffusion in globins: I Leghemoglobin. *Proteins Struct Funct Genet* 10:70–80.
- Elber R, Karplus M. 1990. Enhanced sampling in molecular dynamics: Use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J Am Chem Soc* 112:9161–9175.
- Gether U, Johansen TE, Snider RM, Lowe JA III, Nakanishi S, Schwartz TW. 1993. Different binding epitopes for substance P and the non-peptide antagonist, CP96,345, on the NK₁ receptor. *Nature* 362:345–348.
- Go N, Scheraga HA. 1970. Ring closure and local conformational deformations of chain molecules. *Macromolecules* 3:178–187.
- Greer J. 1990. Comparative modeling methods: Application to the family of the mammalian serine proteases. *Proteins Struct Funct Genet* 7:317–334.
- Kyle DJ, Sinsko JA, Stormann TM, Chakravarty S. 1993. A proposed model of bradykinin bound to the rat B2 receptor and its utility for drug design. *J Med Chem*. Forthcoming.
- Miranker A, Karplus M. 1991. Functionality maps of binding sites: A multicopy simultaneous search method. *Proteins Struct Funct Genet* 11:29–34.
- Novotny J, Brucoleri RE, Saul FA. 1989. On the attribution of binding energy in antigen–antibody complexes MCPC603, D1.3, and HyHEL-5. *Biochemistry* 28:4735–4749.
- Novotny J, Rashin AA, Brucoleri RE. 1988. Criteria that discriminate between native proteins and incorrectly folded models. *Proteins Struct Funct Genet* 4:19–30.
- Palmer KA, Scheraga HA. 1991. Standard-geometry chains fitted to X-ray derived structures: Validation of the rigid-geometry approximation. I. Chain closure through a limited search of “loop” conformations. *J Comput Chem* 12:505–526.
- Roitberg A, Elber R. 1991. Modeling side chains in peptides and proteins:

- Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *J Chem Phys* 95:9277-9287.
- Rosenfeld R, Zheng Q, Vajda S, DeLisi C. 1993. Computing the structure of bound peptides: Application to antigen recognition by class I MHCs. *J Mol Biol* 234:515-520.
- Shenkin PS, Yarmush DL, Fine RM, Wang H, Levinthal C. 1987. Predicting antibody hypervariable loop conformation. I: Ensembles of random conformations for ringlike structures. *Biopolymers* 26:2053-2085.
- Shoichet BK, Kuntz ID. 1991. Protein docking and complementarity. *J Mol Biol* 221:327-346.
- Straub JE, Karplus M. 1991. Energy equipartitioning in the classical time-dependent Hartree approximation. *J Chem Phys* 94:6737-6739.
- Summers N, Karplus M. 1990. Modeling of globular proteins. A distance-based data search procedure for the construction of insertion/deletion regions and Pro $\leftrightarrow$ non Pro mutations. *J Mol Biol* 216:991-1016.
- Ulitsky A, Elber R. 1993. The thermal equilibrium aspects of the time dependent Hartree and the locally enhanced sampling approximations: Formal properties, a correction, and computational examples for rare gas clusters. *J Chem Phys* 98:3380-3388.
- Verkhivker G, Elber R, Gibson QH. 1992. Microscopic modeling of ligand diffusion through the protein leghemoglobin: Computer simulations and experiments. *J Am Chem Soc* 114:7866-7878.
- Weiner PK, Profeta S Jr, Wipff G, Havel T, Kuntz ID, Langridge R, Kollman PA. 1983. A distance geometry study of ring systems. *Tetrahedron* 39:1113-1121.
- Zheng Q, Rosenfeld R, Kyle DJ. 1993a. Theoretical analysis of the multi-copy sampling method in molecular modeling. *J Chem Phys* 99:8892-8896.
- Zheng Q, Rosenfeld R, Vajda S, DeLisi C. 1993b. Loop closure via bond scaling and relaxation. *J Comput Chem* 14:556-565.
- Zheng Q, Rosenfeld R, Vajda S, DeLisi C. 1993c. Determining protein loop conformation using scaling-relaxation techniques. *Protein Sci* 2:1242-1248.