Protein Science (1993), *2*, 1502–1510. Cambridge University Press. Printed in the USA.

Modeling of protein loops by simulated annealing

V. COLLURA,[1] JUNICHI HIGO,[2] AND J. GARNIER
Unité d'Ingénierie des Protéines, Biotechnologies, INRA, 78352, Jouy en Josas, France

(RECEIVED April 6, 1993; REVISED MANUSCRIPT RECEIVED June 4, 1993)

Abstract

A method is presented to model loops of protein to be used in homology modeling of proteins. This method employs the ESAP program of Higo et al. (Higo, J., Collura, V., & Garnier, J., 1992, *Biopolymers 32*, 33–43) and is based on a fast Monte Carlo simulation and a simulated annealing algorithm. The method is tested on different loops or peptide segments from immunoglobulin, bovine pancreatic trypsin inhibitor, and bovine trypsin. The predicted structure is obtained from the ensemble average of the coordinates of the Monte Carlo simulation at 300 K, which exhibits the lowest internal energy. The starting conformation of the loop prior to modeling is chosen to be completely extended, and a closing harmonic potential is applied to N, CA, C, and O atoms of the terminal residues. A rigid geometry potential of Robson and Platt (1986, *J. Mol. Biol. 188*, 259–281) with a united atom representation is used. This we demonstrate to yield a loop structure with good hydrogen bonding and torsion angles in the allowed regions of the Ramachandran map. The average accuracy of the modeling evaluated on the eight modeled loops is 1 Å root mean square deviation (rmsd) for the backbone atoms and 2.3 Å rmsd for all heavy atoms.

Keywords: loop prediction; modeling of proteins; Monte Carlo; protein structure prediction; Robson–Platt potential; simulated annealing

At present, modeling of protein structure by homology to a protein of known structure is an effective way to obtain the correct topology of the polypeptide chain fold. Currently this comprises two essential steps. The first step is the alignment of amino acid sequences of the protein to be modeled with those from protein(s) of known structure. In cases where the sequences have weak homologies and unequal lengths this step is subject to local misalignments when compared to crystallographic alignments. This can be complemented by biochemical information relating activity to a modified amino acid sequence and may sometimes be corrected by secondary structure prediction based on sequence similarity (Levin & Garnier, 1988; Garnier & Levin, 1991). In a second step the coordinates of the residues of the known protein are transferred to residues of the unknown protein that match in the alignment. This can be done through the use of software with a relatively good accuracy (Srinivasan et al., 1993). However, there are segments of the polypeptide chain that do not match at all because they are of different lengths and sequences. Usually, but not always, they correspond to peptide segments of aperiodic structures joining the secondary structural elements forming the conserved core of the homologous proteins (see the survey made by Chothia & Lesk [1986]). Some of these variable segments may have important biological functions with important therapeutic applications, such as the recognition of antigens by antibodies (Foote & Winter, 1992).

Modeling of protein loops has previously included either a search of similar loops in a protein data base (Chothia & Lesk, 1987; Sutcliffe et al., 1987; Chothia et al., 1989) or a selection of loops after a more or less extensive conformational search (Fine et al., 1986; Shenkin et al., 1987; Bruccoleri et al., 1988; Martin et al., 1989) followed by energy minimization. We want to report here the results of a Monte Carlo method (Higo et al., 1992) for the modeling of loops of medium size, six to nine amino acid residues from an initial extended conformation. These loops were taken from different proteins in order to ascertain the generality of a method that has previously only been applied to the modeling of hypervariable loops of immunoglobulins (Gibrat et al., 1992; Higo et al., 1992). The

Reprint requests to Jean Garnier: Unité d'Ingénierie des Protéines, Biotechnologies, INRA, 78352, Jouy en Josas, France.
[1] Present address: Proteus Molecular Design, Ltd. Lyme Green Business Park, Macclesfield SK11 0JL, United Kingdom.
[2] Present address: Nagayama Protein Array Project, ERATO, JDRDC, Tsukuba Research Consortium, 5-9-1 Tokodai, Tsukuba, 300-26, Japan.

results shown below indicate an accuracy very similar to the accuracy of modeling of the immunoglobulin loops without the difficulties linked to conformational search algorithms such as that used by Bruccoleri et al. (1988) in which the search space increases exponentially with the number of amino acids in the loop.

Monte Carlo–Metropolis (MC) simulations (Metropolis et al., 1953) have been considered as less efficient than molecular dynamics (MD) for nonhomogeneous and anisotropic systems such as macromolecules (Northrup & McCammon, 1980). This point of view has been challenged by Bouzida et al. (1992). Furthermore, a number of algorithms have been developed to increase the efficiency of MC simulations in exploring conformational space (Pangali et al., 1978; Rossky et al., 1978; Goldman, 1983; Noguti & Gō, 1985; Cao & Berne, 1990; Garel & Orland, 1990; Shin & Jhon, 1991; Bouzida et al., 1992; Kotelyanskii & Suter, 1992). An MC simulation offers, like MD, the possibility of calculating equilibrium properties by computing appropriate averages over a set of generated conformations in order to represent the thermal fluctuations at the temperature at which the experimental data are obtained or at which the system performs its biological function. If the MC simulation is long enough, the same average value as that obtained experimentally at equilibrium should be calculated. This is a priori a more realistic representation of the properties of a peptide loop including its conformation than a single minimized conformation at 0 K.

The simulated annealing that we developed, ESAP (Higo et al., 1992), starts with an MC simulation at high temperature from the extended conformation of the peptide loop. The loop is considered to be flexible, the other atoms of the protein being conformationally fixed. The loop closure is achieved by applying an harmonic potential to the four backbone atoms of the terminal residues. The process is then annealed to the final temperature, a process found to be most effective at finding the global minimum, avoiding as far as possible the system becoming trapped in local minima (Kirkpatrick et al., 1983). From the MC simulation having the lowest calculated internal energy value, a structure can be predicted by calculating the Boltzmann average of the coordinates of the sampled conformations of that simulation at 300 K. The accuracies of the predictions were 1 Å root mean square deviation (rmsd) for the backbone atoms and 2.3 Å rmsd for all the atoms on average for the eight loops tested. Hydrogen bond networks and torsion angle distributions of the loops were also analyzed.

Materials and methods

Protein loops

For a comparison between the modeled loops and the fixed part of the proteins, the coordinates were taken from the Brookhaven Protein Data Bank (Bernstein et al., 1977). For the antibody McPC603 (1mcp at 2.7 Å resolution), data were taken from the study of Satow et al. (1987) with the hypervariable loops L2 (G_{56}ASTRES$_{61}$) and H3 (Y_{102}YGSTWYF$_{109}$) used. For BPTI (4pti at 1.5 Å resolution), data were from Marquart et al. (1983) with the following loops used: B1 (L_6EPPYTG$_{12}$), B2 (C_{14}KARIIR$_{20}$), B3 (F_{22}YNAKAGL$_{29}$), B4 (G_{36}GCRAKR$_{42}$), and B5 (N_{44}FKSAED$_{50}$). For the purpose of comparison the BPTI loops corresponded to those already modeled by Dudek and Scheraga (1990). For trypsin (2ptn, 1.55 Å resolution), data were from Walter et al. (1982) with the loop T1 (N_{123}TKSSGTSY$_{131}$) used.

Target function

The target function, E_t, was a linear combination of two kinds of terms, an harmonic distance constraint, E_d, and the nonbonded atomic interactions, E_i:

$$E_t = \lambda_d E_d + \lambda_i E_i, \tag{1}$$

with

$$E_d = \sum_i (r_i - r_{i,x})^2, \tag{2}$$

where r_i and $r_{i,x}$ are vectors representing the position of atom i of the N- and C-terminal residues for, respectively, the flexible peptide segment and the corresponding position from X-ray data. The summation for E_d is taken over the main-chain atoms (N, CA, C, and O) of the terminal residues.

The nonbonded interactions E_i were calculated with the force field developed by Robson and Platt (1986). This force field makes use of rigid geometry reducing significantly the number of variables to the torsion angles of the backbone (ϕ and ψ) and χ for the side chains. It contains pairwise atomic interactions, van der Waals and electrostatic, for the nonbonded atoms, and an intrinsic rotation potential for the side-chain dihedral angles. It uses a united atom representation in which hydrogens of nonpolar atoms are treated implicitly and hydrogens of polar atoms that may form hydrogen bonds are added, in both flexible and fixed parts, at the stereochemically standard positions as determined by the positions of nonhydrogen atoms connected to the hydrogen atoms. The peptide bond dihedral angle ω was fixed at 180° and proline residues in trans or cis conformation. The dielectric constant was set to 5 and a cutoff distance of 10 Å was used for the calculation of pairwise atomic interactions. The pairlist was renewed every 10 Monte Carlo steps. Because the coordinates of the fixed part were taken from raw X-ray data without correction for the rigid geometry used for the flexible peptide segments, the pairwise atomic interactions between the fixed part and the two side chains

of the N- and C-terminal residues of each segment were neglected.

Monte Carlo simulation

We used the Monte Carlo method of Metropolis et al. (1953) for the conformational sampling of the flexible peptide segments, adopting the scaled collective variables developed by Noguti and Gō (1985). The collective variables were the eigenvectors obtained by diagonalizing the second derivative matrix of the target function with respect to the torsion angles. An orthogonal and normalized set of eigenvectors in the torsion angle space was calculated, and a vector step from the currently accepted conformation to a trial one was constructed by a linear combination of the eigenvectors. The second derivative matrix was numerically calculated every 2,000 steps of the simulation with a cutoff distance of 5 Å for the calculation of the nonbonded atomic interactions. The coefficient k_i assigned to the ith eigenvector in the linear combination was

$$k_i = cp_i(RT/N\mu_i)^{1/2}, \tag{3}$$

where R, T, N, and μ_i are the gas constant, the absolute temperature, the number of torsion variables, and the ith eigenvalue, respectively. The p_i is a random number assigned to the ith eigenvector, which is uniformly distributed from -1 and $+1$ during the simulation. The constant c was used to modulate the acceptance ratio around 0.5–0.6 during the simulation. A typical value was 5.0 rad^{-1}. When an eigenvalue was negative, the value was replaced by a small fixed positive value (1.0 kcal/mol/rad^2).

The vector step $\Delta\theta$ in conformational space had elements $\Delta\theta_j$:

$$\Delta\theta_j = \sum_{i=1,N} x_{i,j}k_i, \tag{4}$$

where N is the total number of variable angles and $\Delta\theta_j$ the step taken and added to the torsion angle θ_j of the last accepted conformation; $x_{i,j}$ is an element of the eigenvector matrix and k_i the coefficient defined above in Equation 3.

Simulated annealing

The simulated annealing was performed as described by Higo et al. (1992). The gradual decrease of temperature was obtained by increasing the λ values of Equation 1, keeping the temperature at 300 K. The effective temperatures were $300/\lambda$. First an extended conformation for each flexible peptide segment was generated by setting all the dihedral angles to 180°, and the N, HN, and CA atoms of the N-terminal residues of the peptide segments were moved by appropriate translations and rotations to their reference position from the X-ray data of the fixed part. Next an MC simulation as above was done with λ_d and λ_i equal to 0.01 but counting only nonbonded interactions between atoms of the loop and the fixed part. This allowed the proper closing of the loop within 5,000 MC steps. Then the last accepted conformations every 2×10^4 steps were taken as starting conformations for an annealing process that is shown in Table 1. Usually five of those starting conformations were found to be sufficient. The lowest energy conformation from each annealing stage was used as the starting conformation for the next annealing stage. The ensemble of conformations sampled at 300 K from the last MC simulation, stage 5, was characterized by the internal energy, U, and the average coordinates of the constituents, R_i:

$$U = \langle E_t \rangle = N_{\text{total}}^{-1} \sum_{j=1,m} (n_j + 1)E_{t,j} \tag{5}$$

$$R_i = \langle \boldsymbol{r}_i \rangle = N_{\text{total}}^{-1} \sum_{j=1,m} (n_j + 1)\boldsymbol{r}_{i,j}, \tag{6}$$

where $E_{t,j}$, n_j, m, and N_{total} are the energy function (Equation 1) of the jth accepted conformation, the number of rejected conformations after the jth accepted conformation, the total number of accepted conformations, and the total number of MC steps. Values $\boldsymbol{r}_{i,j}$ are the coordinates of atom i from the jth accepted conformation.

The five simulations for the modeling of a seven-amino acid loop took 7 h on an Alliant FX80 configured with two 4 ACE complexes. No attempt was made to improve further the parallelism of 66% found for the FORTRAN 77 program that we wrote to execute the calculations.

Analysis of the modeled loops

Comparisons with X-ray data were done by calculating the rmsd from the X-ray coordinates with those coordinates calculated from Equation 6 for the backbone three or four heavy atoms, rmsdb (N, CA, C, and O) and for all the heavy atoms including the side-chain atoms, rmsda. Although the force field involved a standard ge-

Table 1. *Annealing schedule*

	Stage				
	1	2	3	4	5
λ_d[a]	0.1	0.5	2.5	10.0	10.0
λ_i[a]	0.01	0.05	0.25	1.0	1.0
T[b] $= 300/\lambda_i$	3×10^4	6×10^3	1.2×10^3	300	300
No of steps (steps/10^4)	8	8	8	4	4

[a] From Equation 1.
[b] Effective temperature for the nonbonding interactions in degrees K.

ometry different from the one used to refine the X-ray data, the comparison with the X-ray coordinates was done without regularization of the X-ray structure and without superposition algorithms. The torsion backbone angles ϕ and ψ, calculated from the coordinates of Equation 6 were also compared to observed values calculated from the X-ray coordinates. The hydrogen bond network was calculated using the algorithm of Kabsch and Sander (1983) adapted to include the side-chain atoms.

Results and discussion

Ensemble averages of the energies and coordinates have been considered by Higo et al. (1992) as the best approximation to compare the modeled hypervariable loops of antibodies with X-ray data. Comparison was made between the X-ray coordinates without regularization of the X-ray structure with the rigid geometry and without a least-squared superposition of the modeled loop. This direct comparison with X-ray data usually increases the rmsd by about 20% (Gibrat et al., 1992) but has the advantage of being closer to the expected error of the modeling of an unknown loop structure.

Rigid geometry approximation and ensemble averages of coordinates

In some instances we have noticed that the use of Boltzmann averages of the coordinates could alter some bond lengths of side-chain atoms, due to conformational fluctuations during the MC simulation. This can be corrected by fitting a rigid geometry structure on the average coordinates. The corrections from the average coordinates are usually very small, less than 0.1 Å for the backbone rmsd and 0.2 Å for all the heavy atoms. This is negligible compared to crystallographic errors, estimated for the best-resolved structures to be 0.5 Å for all heavy atoms (see discussion in Gibrat et al., 1992). The use of a rigid geometry was also found to be in the range of the lowest crystallographic errors. For instance, regularization, which is distance minimization of the rigid geometry used with X-ray coordinates for the six loops of J539 antibody at 1.95 Å resolution, gave an average rmsd for all heavy atoms of 0.5 Å (Gibrat, pers. comm.).

Internal energy and lowest rmsd

Previous results (Higo et al., 1992) have shown that internal energy obtained from the last MC simulation (Equation 5) could be used as a criterion for evaluating the accuracy of the modeling, i.e., the lower the internal energies, the closer is the structure of the modeled loops to those from X-ray analyses. For half of the loops (Table 2) the lowest internal energy corresponded to the lowest rmsd, backbone or all atoms. For the other four loops,

Table 2. *Results of the Monte Carlo (MC) simulations for each tested loop*[a]

	U	rmsdb	rmsda	*U*	rmsdb	rmsda
		B1			B2	
$\tau 1$	−7.46	0.89	2.65	**4.89**	**0.87**	**1.97**
$\tau 2$	−12.38	1.01	1.63	10.03	1.35	4.49
$\tau 3$	**−13.19**	**0.92**	**2.59**	34.42	1.79	4.43
$\tau 4$	−12.00	1.00	1.46	7.34	0.69	2.43
$\tau 5$	−9.09	1.10	1.64	8.79	1.48	4.39
		B3			B4	
$\tau 1$	18.79	2.29	5.52	18.14	0.76	2.46
$\tau 2$	26.18	2.87	5.42	**7.88**	**0.93**	**2.31**
$\tau 3$	18.03	1.90	4.78	40.34	3.09	4.91
$\tau 4$	22.94	2.98	5.06	30.15	3.29	4.26
$\tau 5$	**−0.80**	**1.54**	**2.67**	23.36	0.82	2.02
		B5			T	
$\tau 1$	−4.32	0.49	1.59	15.19	6.51	6.43
$\tau 2$	30.41	3.88	5.58	**−3.30**	**1.38**	**2.70**
$\tau 3$	20.72	1.73	4.34	2.78	0.84	2.26
$\tau 4$	11.37	1.11	1.76	6.30	5.96	5.81
$\tau 5$	**−4.67**	**0.54**	**1.75**	216.00	3.73	5.25
		L2			H3	
$\tau 1$	12.31	1.48	3.62	25.46	3.44	5.18
$\tau 2$	11.01	0.91	1.65	22.80	3.80	5.52
$\tau 3$	**8.54**	**0.83**	**1.46**	19.73	2.41	5.88
$\tau 4$	11.46	1.47	4.48	**6.13**	**0.93**	**3.25**
$\tau 5$	11.06	0.92	1.85	15.78	2.11	5.30

[a] $\tau 1$-$\tau 5$ correspond to five different annealings. *U* is given in kcal/mol; rmsd is in Å, rmsdb for backbone atoms N, CA, and C, and rmsda for all heavy atoms. B1–B5 correspond to BPTI (bovine pancreatic trypsin inhibitor) loops, T to the trypsin loop, and L2 and H3 to hypervariable loops of McPC603 (see Materials and methods for definition of the loops). Bold values correspond to the lowest internal energy.

the rmsds of the lowest energy structures were very close to the lowest rmsd. Modeling based solely on the conformations of lowest internal energy yielded an average rmsd backbone (three atoms) for the eight loops of 0.99 Å compared to 0.87 Å for the lowest rmsd found during all the MC simulations. For all the heavy atoms the average rmsd was 2.34 Å compared with 2.09 Å for the lowest rmsd. The loss of accuracy by selecting the lowest internal energy simulations is about 10%, or 0.1–0.2 Å. These differences between the lowest internal energies and lowest rmsd probably reflect the fact that we did not calculate the free energy change or the explicit contribution of the solvent.

The ESAP method of Higo et al. (1992) has already been found to be more accurate for modeling hypervariable loops of immunoglobulins than the conformational search methods (Gibrat et al., 1992), and the results for

the two hypervariable loops of McPC603 confirmed these findings. The accuracy of the modeling of the bovine pancreatic trypsin inhibitor (BPTI) and trypsin loops in Table 2 were very similar to the accuracy found for the hypervariable loops even if some of them contain residues in regular secondary structure such as those of B2, B3, and B5. Those results are also close to those found from the modeling of the same loops of BPTI made by Dudek and Scheraga (1990) if one takes into account the fact that their comparisons were made with regularized loops and not with direct X-ray coordinates. However for loops B1, B4, and B5, they obtained a significantly better backbone rmsd although the all-atoms rmsd was not reported. It is worth noting that Dudek and Scheraga (1990) started all their simulations not from a completely random or extended structure of the loops, as we did, but from a collection of backbone structures obtained by deforming the regularized X-ray backbone structure. One may question the efficiency of this process to generate structures that are not related to the initial X-ray structure. In practical loop modeling, the Dudek and Scheraga method is not applicable because their algorithm requires knowledge of the X-ray structure for the generation of a starting conformation.

Stereo drawings of the modeled loops of BPTI and trypsin are presented in Figure 1 and compared to X-ray data. A more detailed description of their individual residue rmsds is given in Table 3 for the MC simulations of lowest internal energy. The closure of the loops with the simple harmonic potential was quite effective, with an rmsd of 0.1–0.2 Å for the backbone atoms of the terminal residues.

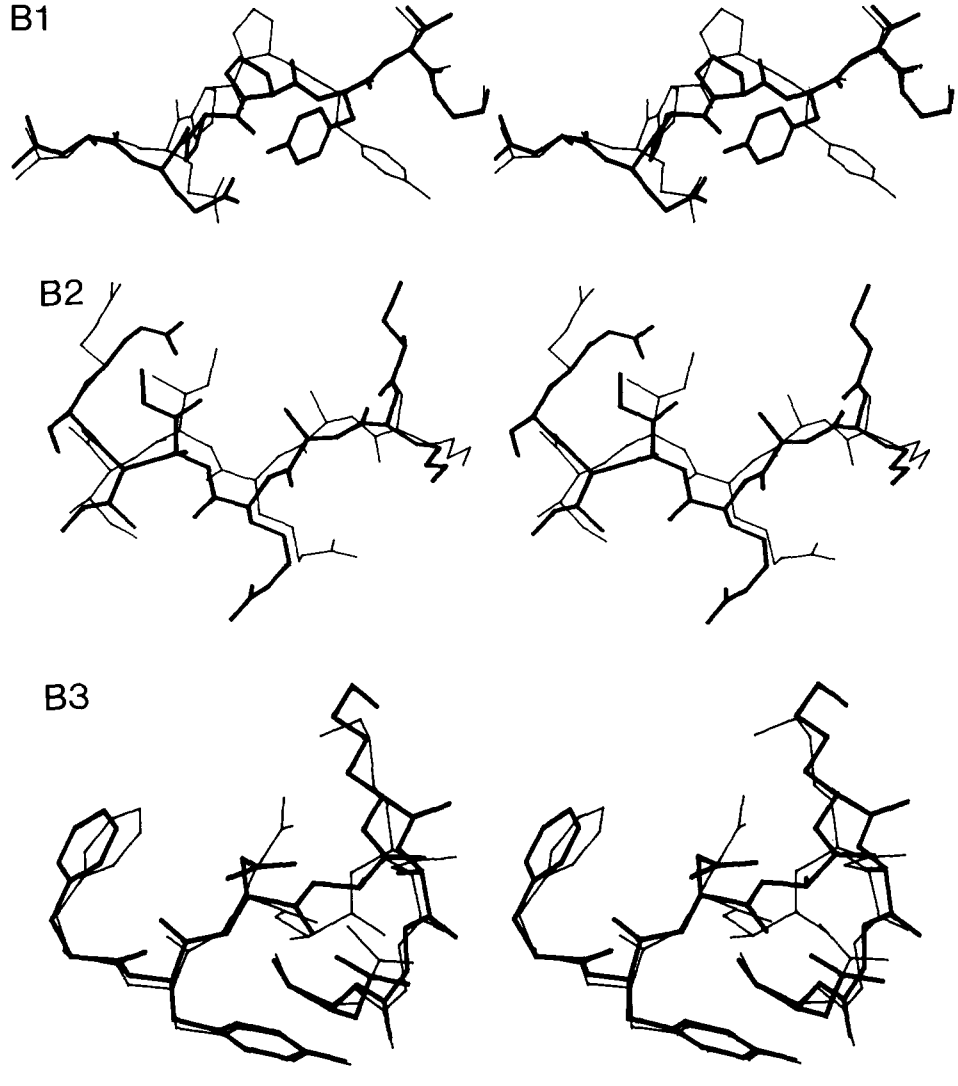

Fig. 1. Stereo drawings of the loops from BPTI and trypsin. These correspond to the MC simulations of lowest internal energy in Table 2. The X-ray structures are drawn with thin lines and the simulated structures with thick lines. The labeling of the loops is the same as in the Materials and methods, B1-B5 for BPTI, and T for trypsin. (*Continues on facing page.*)

It is becoming increasingly accepted that the distributions of the observed torsion angles, ϕ and ψ, in the Ramachandran map, are localized in specifically defined regions for all residues except glycine (Morris et al., 1992). There are a very few exceptions to this when the X-ray structure is well refined at high resolution. We plotted in Figure 2 the observed torsion angles and the simulated ones from the MC simulation of the lowest internal energy for the BPTI and trypsin loops. The torsion angles calculated from the X-ray structure were found in the allowed regions, and all the observed positive ϕ torsion angles corresponded to Gly residues except R39 of loop B4 of BPTI, which is in the allowed region of a left α-helix. Remarkably, all the simulated torsion angles were found to be in the correct regions of the map except one, residue Cys-38 of loop B4 of BPTI at 65° and −99°. Its observed value was −146° and 156°. The other positive values of ϕ corresponded to Gly residues. The means of absolute variations of the predicted torsion angles with the observed angles were 30 and 32° for the ϕ and ψ angles, respectively, with a standard deviation of 42°.

Analysis of the simulated hydrogen bonds

It is well recognized that hydrogen bonds are a characteristic feature of protein structures. The analysis of the hydrogen bonds for the lowest internal energy simulations were compared with the X-ray data, and some of the results are presented in Table 4. The predicted hydrogen bonds that match the observed ones have a remarkably similar calculated energy; however, a certain number of bonds are not observed or are incorrectly calculated (part B of Table 4). Noticeably on an analysis made on the five loops of BPTI and the trypsin loop, 43 hydrogen bonds were predicted for 44 bonds observed, and among them 26 or 45% of the predicted hydrogen bonds were correct. Among the observed hydrogen bonds, 70% were bonds including backbone atoms but only 60% for the predicted bonds. The backbone hydrogen bonds were twice as accurately predicted as those of the side-chain hydrogen bonds: 78% of the predicted bonds were correct for the backbone atoms and only 31% for the side chains. This reflects the greater rmsds of the side-chain atoms com-

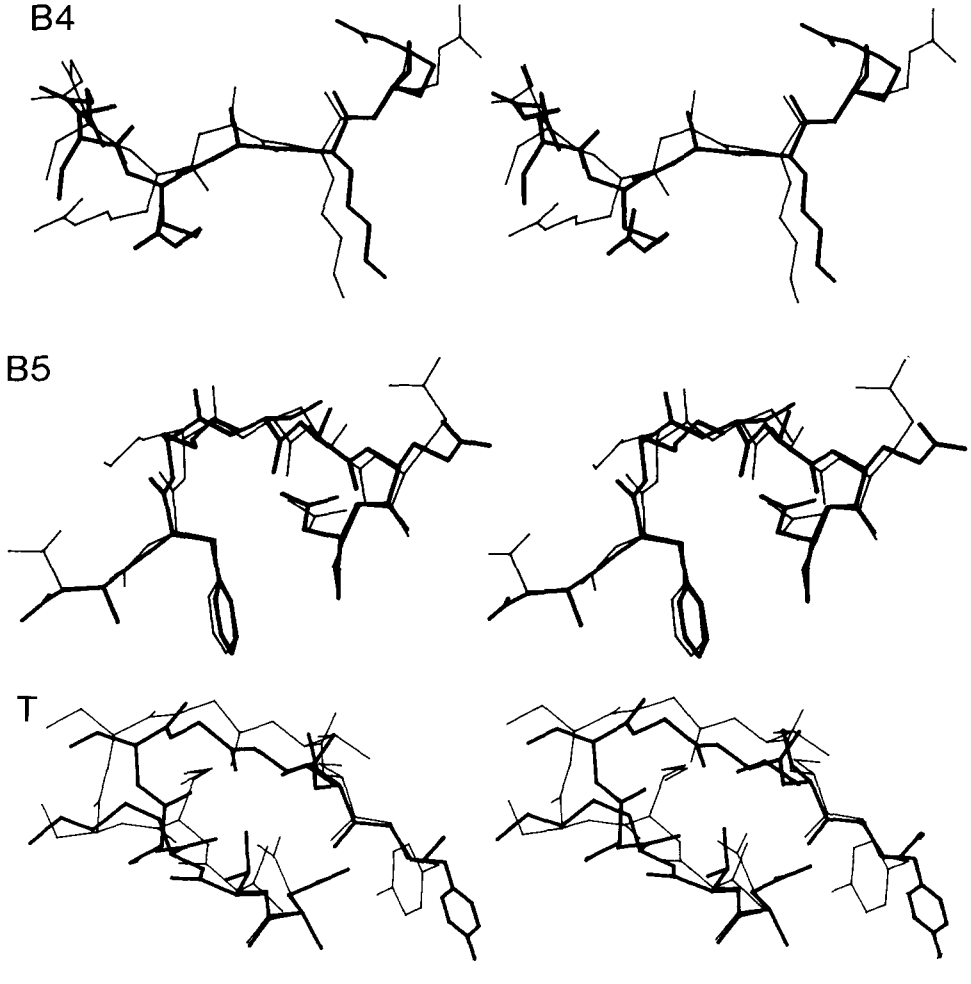

Fig. 1. *Continued.*

Table 3. *Root mean square deviations in Å between the model loops and their X-ray structure on a residue basis for BPTI and trypsin loops*

Loop	Residue	rmsdb[a]	rmsdb[b]	rmsda[c]
B1	L6	0.09	0.15	0.51
	E7	0.89	1.05	1.97
	P8	1.23	1.40	1.34
	P9	1.56	1.41	2.15
	Y10	1.04	0.98	4.82
	T11	0.31	0.35	0.41
	G12	0.11	0.16	0.16
	Total	0.92	0.94	2.59
B2	C14	0.03	0.03	0.10
	K15	0.60	1.29	1.33
	A16	1.17	1.27	1.37
	R17	1.25	1.45	3.06
	I18	1.33	1.34	2.07
	I19	0.44	0.47	1.65
	R20	0.08	0.09	1.94
	Total	0.87	1.03	1.97
B3	F22	0.10	0.09	0.76
	Y23	0.39	0.45	1.23
	N24	0.59	0.54	1.76
	A25	2.29	2.58	2.69
	K26	3.02	3.08	6.09
	A27	1.88	1.69	1.82
	G28	0.81	0.77	0.77
	L29	0.16	0.17	0.58
	Total	1.55	1.59	2.67
B4	G36	0.16	0.33	0.33
	G37	1.42	2.31	2.31
	C38	1.23	1.99	1.97
	R39	0.97	1.40	2.90
	A40	1.01	0.98	1.06
	K41	0.74	0.72	1.54
	R42	0.17	0.15	3.06
	Total	0.93	1.35	2.31
B5	N44	0.08	0.31	2.07
	F45	0.30	0.34	0.45
	K46	0.58	0.63	2.48
	S47	0.89	0.86	1.01
	A48	0.82	0.76	0.87
	E49	0.40	0.38	2.74
	D50	0.15	0.13	0.39
	Total	0.54	0.55	1.75
T2	N123	0.07	0.09	1.62
	T124	0.38	0.48	1.54
	K125	1.34	1.38	4.58
	S126	2.81	3.03	4.01
	S127	1.39	1.76	1.62
	G128	2.03	2.33	2.33
	T129	1.03	1.02	1.31
	S130	0.36	0.34	0.96
	Y131	0.14	0.13	2.83
	Total	1.38	1.53	2.70

[a] Includes atoms N, CA, and C.
[b] Includes atoms N, CA, C, and O.
[c] Includes all heavy atoms, backbone, and side chain.

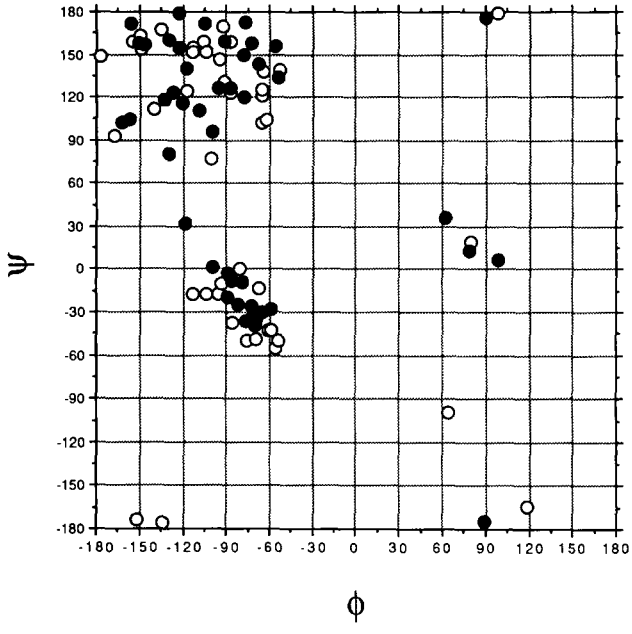

Fig. 2. Ramachandran map of observed and simulated torsion angles. The symbol ● corresponds to the torsion angles calculated from the X-ray coordinates and ○ to those calculated from the ensemble average coordinates of the MC simulations of lowest internal energy. All the torsion angles of the residues of the five loops of BPTI and of trypsin are represented.

pared to the backbone atoms. Possibly the difficulty in simulating more accurately the side chains of the amino acids depends, in part, on the solvent approximation made here. There are no explicit water molecules in the simulation, only a global solvent effect through a dielectric constant of 5 and the use of the "s" function, which smoothes the electrostatic interactions at long distance (Robson & Platt, 1986). However, the total number of side-chain hydrogen bonds predicted, 16, does not exceed by much the observed number, 13. These results, including the overall number of predicted hydrogen bonds (43 for 44 observed), greatly favor the quality of the force field used here to simulate surface loops.

Conclusion

The ESAP method developed by Higo et al. (1992) for modeling hypervariable loops of immunoglobulins has been found to be a method applicable to loops or peptide segments of similar length from different proteins with a comparable accuracy, about an rmsd backbone of 1 Å and an all-atom rmsd of 2.3 Å. Could one expect better accuracy? The solvent effect was not explicitly included in this simulation, but the relative quality of hydrogen bond prediction suggests that the force field used is suitable for the modeling of solvent-accessible loops. It has the merit of the simplicity related to the use of a rigid geometry. This geometry is close to the observed geometry

Table 4. *Predicted and observed hydrogen bonds of BPTI loops B1 and B5*

Loop[a]	Loop amino acid		Other amino acid		Hydrogen bond energy[b] (kcal/mol) Predicted	Observed
B1						
A	L6	N-HN	O-C	D3	−2.04	−1.88
	E7	C-O	HND1-ND2	N43	−1.87	−2.95
	T11	C-O	HN-N	G36	−1.76	−2.35
B	E7	N-HN	O-C	D3	−0.68	>
		N-HN	O-C	F4	>	−1.29
		CD-OE1	HNZ1-NZ	K41	−0.95	>
		CD-OE2	HN-N	N43	>	−1.67
	P8	C-O	HNZ1-NZ	K41	−0.76	>
	P9	C-O	HN-N	T11	>	−0.55
B5						
A	N44	N-HN	O-C	R42	−2.18	−2.23
	F45	N-HN	O-C	Y21	−2.62	−2.56
		C-O	HN-N	Y21	−2.55	−2.94
	S47	C-O	HN-N	C51	−1.09	−2.02
	A48	C-O	HN-N	M52	−1.90	−1.13
	E49	C-O	HN-N	R53	−1.62	−1.93
	D50	C-O	HN-N	T54	−2.16	−2.39
B	N44	CG-OD1	HN11-NH1	R20	>	−3.34
		CG-OD1	HN-N	Y10	−0.52	>
		ND2-HD1	O-C	K41	−1.79	>
	K46	NZ-HNZ1	OD1-CG	D50	−0.69	>
	S47	N-H	OD1-CG	D50	−0.57	>
	E49	CD-OE1	HN12-NH1	R53	−1.1	>
	D50	CG-OD1	HN21-NH2	R53	>	−2.9
		CG-OD2	HN21-NH2	R53	−1.19	>

[a] Entries A correspond to predicted hydrogen bonds that match the observed ones and entries B to those that mismatch observed hydrogen bonds.

[b] >, Energy is higher than −0.5 kcal/mol.

of well-refined and resolved X-ray structures and simulated torsion angles clustered in low minimum regions of the Ramachandran map as is observed from X-ray data. The calculated internal energies, with a few exceptions and without excessive errors, seemed to be a good index of the lowest rmsd, avoiding the need for free energy calculations.

Thus for the time being progress is not expected by use of a new force field but rather from a better simulated annealing process. On a series of five different MC simulations, there were always several simulations of high energy and rmsd conformations, indicating that the loops were trapped in local minima. Work is in progress to overcome this problem.

Acknowledgments

We acknowledge the postdoctoral fellowship from INRA to J.H. and the predoctoral fellowship from the MRT to V.C. (MRT no. 88178). We thank Dr. Roger Mee from Proteus for carefully reading the manuscript.

References

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol. 112*, 535–542.

Bouzida, D., Kumar, S., & Swendsen, R.H. (1992). Efficient Monte Carlo methods for the computer simulation of biological molecules. *Phys. Rev. 45*, 8894–8901.

Bruccoleri, R.E., Haber, E., & Novotny, J. (1988). Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature 335*, 564–568.

Cao, J. & Berne, B.J. (1990). Monte Carlo methods for accelerating barrier crossing: Anti-force-bias and variable step algorithms. *J. Chem. Phys. 92*, 1980–1985.

Chothia, C. & Lesk, A.M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J. 5*, 823–826.

Chothia, C. & Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol. 196*, 901–917.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M., & Poljak, R.J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature 342*, 877–883.

Dudek, M.J. & Scheraga, H.A. (1990). Protein structure prediction using a combination of sequence homology and global energy minimization. I. Global energy minimization of surface loops. *J. Comput. Chem. 11*, 121–151.

Fine, R.M., Wang, H., Shenkin, P.S., Yarmush, D.L., & Levinthal, C. (1986). Predicting antibody hypervariable loop conformations II: Minimization and molecular dynamics studies of MCPC 603 from many randomly generated loop conformations. *Proteins Struct. Funct. Genet. 1*, 342–362.

Foote, J. & Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol. 224*, 487–499.

Garel, T. & Orland, H. (1990). Guided replication of random chain: A new Monte Carlo method. *J. Phys. A 23*, L621–L626.

Garnier, J. & Levin, J.M. (1991). The protein structure code. What is its present status? *Cabios 7*, 133–142.

Gibrat, J.F., Higo, J., Collura, V., & Garnier, J. (1992). A simulated annealing method for modeling the antigen combining site of antibodies. *Immunomethods 1*, 107–125.

Goldman, S. (1983). A simple new way to help speed up Monte Carlo convergence rates: Energy-scaled displacement Monte Carlo. *J. Chem. Phys. 79*, 3938–3947.

Higo, J., Collura, V., & Garnier, J. (1992). Development of an extended simulated annealing method: Application to the modeling of complementary determining regions of immunoglobulins. *Biopolymers 32*, 33–43.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.

Kirkpatrick, S., Gelatt, C.D., Jr., & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science 220*, 671–680.

Kotelyanskii, M.J. & Suter, U.W. (1992). A dynamic Monte Carlo method suitable for molecular simulations. *J. Chem. Phys. 96*, 5383–5388.

Levin, J.M. & Garnier, J. (1988). Improvements in a secondary structure prediction method based on a search for local sequence homologies and its use as a model building tool. *Biochim. Biophys. Acta 955*, 283–295.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983). The geometry of the reaction site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. B39*, 480.

Martin, A.C.R., Cheetham, J.C., & Rees, A.R. (1989). Modeling antibody hypervariable loops, a combined algorithm. *Proc. Natl. Acad. Sci. USA 86*, 9268–9272.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys. 21*, 1087-1092.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G., & Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins Struct. Funct. Genet. 12*, 345-364.

Noguti, T. & Gō, N. (1985). Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. *Biopolymers 24*, 527-546.

Northrup, S.H. & McCammon, J.A. (1980). Simulation methods for protein structure fluctuations. *Biopolymers 19*, 1001-1016.

Pangali, C., Rao, M., & Berne, B.J. (1978). On a novel Monte Carlo scheme for simulating water and aqueous solutions. *Chem. Phys. Lett. 55*, 413-417.

Robson, B. & Platt, E. (1986). Refined models for computer calculations in protein engineering. Calibration and testing of atomic potential functions compatible with more efficient calculations. *J. Mol. Biol. 188*, 259-281.

Rossky, P.J., Doll, J.D., & Friedman, H.L. (1978). Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys. 69*, 4628-4633.

Satow, Y., Cohen, G.H., Padlan, E.A., & Davies, D.R. (1987). Phosphocholine binding immunoglobulin fab MC/PC603. An X-ray diffraction study at 2.7 Ångstroms. *J. Mol. Biol. 190*, 593-604.

Shenkin, P.S., Yarmush, D.L., Fine, R.M., Wang, H., & Levinthal, C. (1987). Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ringlike structures. *Biopolymers 26*, 2053-2085.

Shin, J.K. & Jhon, M.S. (1991). High directional Monte Carlo procedure coupled with the temperature heating and annealing as a method to obtain the global energy minimum structures of polypeptides and proteins. *Biopolymers 31*, 177-185.

Srinivasan, S., March, C.J., & Sudarsanam, S. (1993). An automated method for modeling proteins on known templates using distance geometry. *Protein Sci. 2*, 277-289.

Sutcliffe, M.J., Haneef, I., Carney, D., & Blundell, T.L. (1987). Knowledge based modelling of homologous proteins. Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng. 1*, 377-384.

Walter, J., Steigemann, W., Singh, T.P., Bartunik, H., Bode, W., & Huber R. (1982). On the disordered activation domain in trypsinogen. Chemical labeling and low-temperature crystallography. *Acta Crystallogr. B38*, 1462.