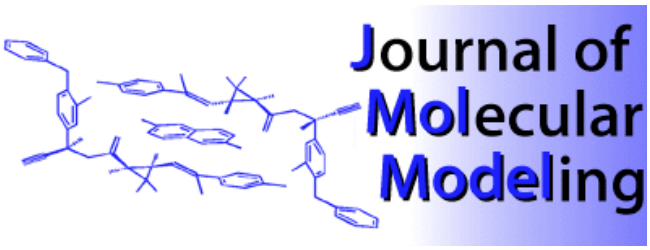
J. Mol. Model. **2000**, *6*, 1 – 8



FULL PAPER

Calculation of Protein Surface Loops Using Monte-Carlo Simulated Annealing Simulation

Zhijie Liu, Fenglou Mao, Weizhong Li, Yuzhen Han, and Luhua Lai

Institute of Physical Chemistry, Peking University, Beijing 100871, P. R. China. Tel: +86-10-62756833; Fax: +86-10-62751725; E-mail: lai@linux2.ipc.pku.edu.cn

Received: 25 January 1999/ Accepted: 11 November 1999/ Published: 14 January 2000

Abstract A modified program for protein loop modelling is presented with significant improvements on our previous study (Zhang et al. 1997, Biopolymers, Vol. 41, pp. 61-72), which is capable of sampling the entire conformational space and identifying the low-energy candidates by Monte-Carlo simulated annealing simulation and a cluster analysis method. Twenty flexible surface loops connecting different secondary structures were selected to test the efficiency of this program. The averaged deviations of backbone heavy atoms for four to eight-residue-loops are 0.19, 0.27, 0.46, 0.41 and 0.87Å respectively. High speed of calculation is achieved with a simplified energy function and a grid-mapping method. As a comparison of single simulation, it takes only four seconds for the simplest four-residue loop and forty-three seconds for the most complex eight-residue loop on a PII-350-Linux platform.

Keywords Protein loop modelling, Monte-Carlo simulated annealing, Cluster analysis, Simplified energy function, Grid-mapping method

Introduction

Calculation of peptide segments and protein local structures is an essential problem in protein structure prediction and protein modelling. For a peptide segment that is very short or is strictly limited by its surrounding protein environment, normal energy minimisation is sufficient to pick out a reasonable structure. However when dealing with the flexible loops, more powerful computational methods are needed to search the vast conformational space and identify the native-like conformations among the numerous candidates. In recent years, some methods based on Monte-Carlo Simulated Annealing (MCSA)[1-4], the multiple copy sampling method[5] and scaling-relaxation [6] have been developed to overcome this problem.

However, due to the huge amount of computation, all these current methods could not achieve sufficient both on accuracy and speed. In our previous study [7], we devised a very efficient MCSA-based package to screen out the lower energy candidates in the conformational space of a protein loop. This high efficiency is due to our grid mapping algorithm and a simplified energy function. After the conformation sampling, a standard CHARMm [8] run continues to pick out the final native-like conformation by a couple of cycles of energy minimisation, which compensate some inaccuracy resulting from the simplified energy function.

In this paper, we make some significant improvements on the algorithm, test the program on more extensive exam-

Correspondence to: L. Lai

ples, and also solve the problems which remained previously. For the MCSA-based method, one concern is to determine the number of simulations necessary to cover the most conformational space. The current work will address this point rather than using an arbitrary value such as 100 times as before. On the other hand, since the coarse energy function and completely random generation method for loop conformation often introduced some unfavourable backbone torsion angles previously, we modify the simulation algorithm here in order to reduce the probability of unreasonable conformations significantly.

Algorithm

The main framework of the current program is inherited from our previous work, which mainly consists of three parts: simulation system, force field and grid-mapping method. The principles and the new enhancements of these methods are illustrated below.

Features derived from our previous work [7,8]

Simulation The first step is to set up the initial configuration of the simulation system (see Figure 1). All atom coordinates of the protein except those of the modelled loop are taken from the Protein Data Bank (PDB)[9] and fixed during the simulation (see Figure 1a). The target loop is built up in an extended conformation in the beginning, with the standard bond lengths and bond angles adopted in CHARMm. Then the loop is attached by its N-terminus to the protein framework (see Figure 1b), and two dummy atoms N' and CA' are generated at the end of the loop. These two dummy atoms are useful for calculation of the harmonic energy, which ensures the smooth closure between the loop and the protein framework in simulation.

The second step is the standard MCSA-simulation, which is composed of two items, iterative generation of random conformations and candidate selection at given temperatures decreasing from the molten state to absolute zero degree with a certain step. During the loop simulation, all the bond lengths and bond angles of the loop are fixed at the initial values, ω dihedrals of the backbone are kept to be 180°, and ϕ dihedrals of prolines are fixed at –75°. The new conformations are produced by changing one of the ϕ, ψ dihedrals of backbone or one of the four χ dihedrals of the side chain at a

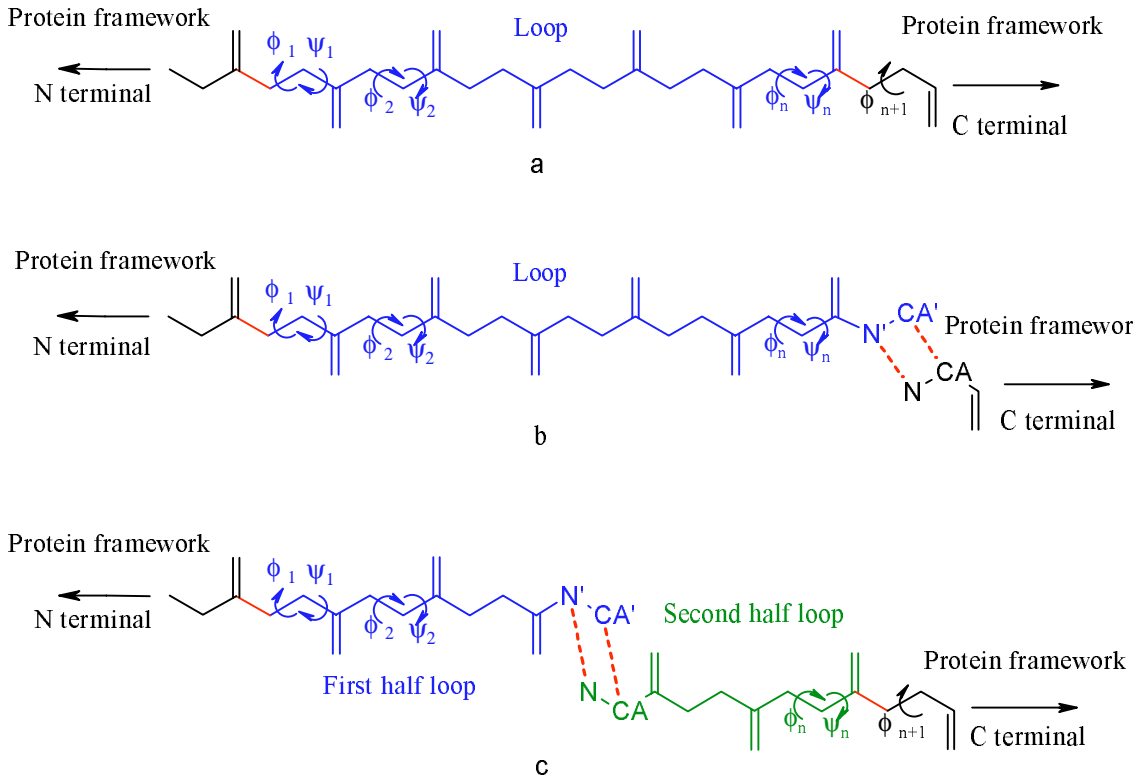


Figure 1 ***Generation of the initial loop conformation*** **(a)** *The situation of the calculated loop;* **(b)** *The method used in previous work, which does not split the loop and only attaches it to the protein framework with pseudo atoms directly.* **(c)** *The method adopted in this work with a split in the middle and connection by pseudo atoms.*

random degree every time. The acceptance of the new conformation is dependent upon conformity with the classical Metropolis criterion.

The final step is to rank the conformation by the classical CHARMm force field. Since the energy function in the simulation is partially coarse (see next part), it will inevitably introduce some unsuitable interactions and unfavourable conformations into the results and the energy of candidates also will not be precise enough. Therefore, all the candidates are minimised by 200 steps of steepest descent (SD), conjugate gradient (CONJ) and adopted basis Newton-Raphson (ABNR) method successively in CHARMm to evaluate the lowest energy conformation.

Energy function In order to screen the conformational space with high speed and relatively high accuracy, a simplified energy function is exploited. Since the bond lengths and bond angles are constant, the total energy E only consists of two parts, the soft-sphere van der Waals non-bonded energy E_s and the harmonic constrained energy E_h. The E_s E_h are given below (see Equation 1 and 2):

$$Es = \begin{cases} k_s\left(d_0^2 - d^2\right), & d_0 \geq d \\ 0 & d_0 < d \end{cases} \quad (1)$$

$$Eh = k_h\left\{\left[\mathbf{r}_{(N)} - \mathbf{r}_{(N')}\right]^2 + \left[\mathbf{r}_{(CA)} - \mathbf{r}_{(CA')}\right]^2\right\} \quad (2)$$

Here, k_s and k_h are force constants for soft-sphere van der Waals non-bonded energy and harmonic constrained energy. d_0 and d represent the sum of van der Waals radii of a pair of atoms and the distance between these atoms respectively. The $\boldsymbol{r}_{(X)}$s are the position vectors of the dummy atoms and their reference atoms are illustrated in Figure 1b.

Grid-mapping In order to speed up computing further, a grid-mapping method is developed to accelerate the calculation of the environmental interactions. For a loop atom located in a defined space, since all the environmental atoms are fixed during the simulation, its nearby non-loop part within the cutoff distance of non-bonded energy is also constant, therefore this non-loop part can be calculated before the simulation. As adopted in this method, a cube that contains the van der Waals volume of the whole fixed protein part is generated initially, then it is divided into more small cubes with a certain grid size such as 1Å. The non-loop atoms are mapped onto each small cube next, if they will possibly clash with a certain cube, they are recorded in a special array. Therefore, each cube is related with some environmental atoms. When calculating the energy between the loop atom and the protein framework, we can determine the non-bonded atom pairs quickly by ascertaining which cube the loop atom is located in instead of calculating numerous atom-distances. Because the calculation of non-bonded interactions between modelled loop atoms and the protein environmental atoms makes up a remarkable portion of the total energy calculation, the grid-mapping algorithm reduces the computation dramatically. For example, the protein BPTI has 454 atoms, when it is divided into 1.0 Å grids, each grid will only interact with 24 atoms [7]. Thus it will reduce almost 95% of the calculation of atom-distances and raise the computational speed about one order of magnitude.

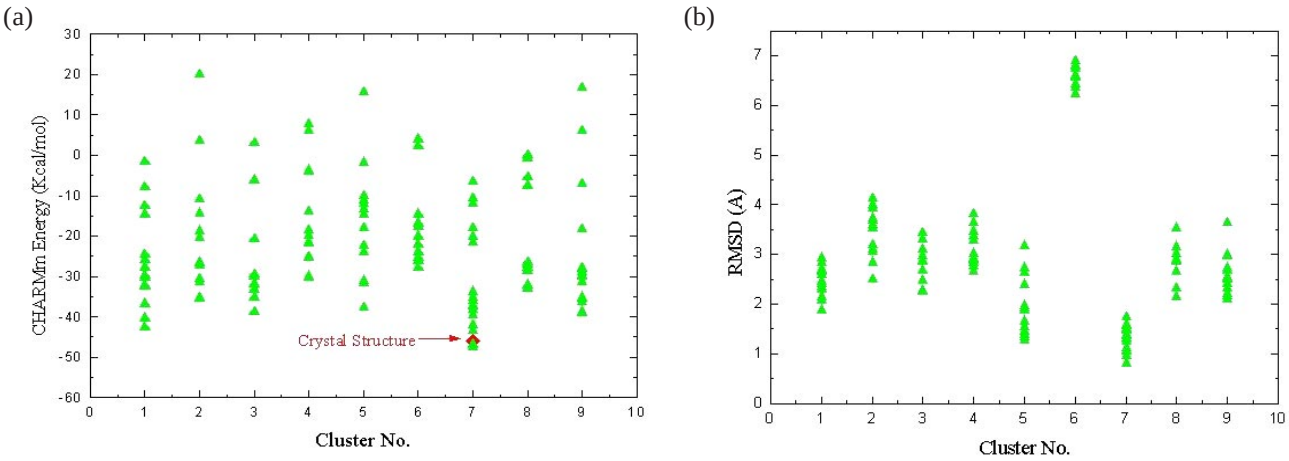


Figure 2 ***The distribution of CHARMm energies and RMSDs of clusters of the loop 5icb_15_8.*** *Only the clusters with more than nine candidates are shown in the figure. Each candidate is plotted as a point;* (**a**) *The energy is calculated by 400 steps of SD minimisation after MCSA simulation;* (**b**) *The RMSD is of the loop backbone heavy atoms between the calculated structure and the crystal structure*

Significant improvements on previous work

In this work, we keep the framework of the previous program and most of the settings of parameters. In addition, we also pay great attention to the efficiency and the rationality of the method to avoid some unnecessary and unreasonable factors. Especially in the simulation part some significant enhancements have been made.

Split-connect procedure for smooth closure of the modelled loop In the first step of simulation, when the modelled loop is built up in an extended conformation, with the standard bond lengths and bond angles adopted in CHARMm, it is split into two parts down the middle. Then two dummy atoms N' and CA', which are the copies of the N and CA atoms of the first residue in the second half of the loop, are attached to the C-terminus of the first half loop. These two dummy atoms will be used for calculation of the harmonic energy to ensure the smooth loop closure in simulation. After that, these two loop segments are connected to the corresponding positions of the protein framework respectively (see Figure 1c). This procedure is a great enhancement on the original one. In this simulation, when the loop is long, it is too free to draw back to the protein framework, even if it is forced to connect to the protein bulk, the dihedral angles especially the ϕ_{n+1} dihedral of the last residue of loop C-terminus are often distorted. The new method, by cutting the long loop into two short ones, can easily control the closure of loop. This will speed up the computational convergence and make the conformation more reasonable.

Reasonable generation of new candidates In the current work, the new conformations of the modelled loop are still produced by changing one of the ϕ , ψ dihedrals of the backbone or one of the four χ dihedrals of the side chain at a random degree. Also all of the bond lengths and the bond angles are fixed at the initial values and ω dihedrals of backbone are kept at 180°. However, the rotating degrees of backbone dihedrals are not set completely randomly as before. For the non-glycine and non-proline residues, when the ψ dihedral is chosen to be rotated, an arbitrary degree is used. While the ϕ dihedral is involved, the program will select a reasonable value from, at most 10, random ones so that the pair of ϕ and ψ values would be located in favourable regions in the Ramachandran map with higher probability. The ϕ dihedrals of prolines are tolerated within ±20° deviation around –75°. This adjustment will ensure the conformations' rationality.

Average Linkage Cluster Analysis (ALCA) As a stochastic method, MCSA has a potential defect that is it cannot find out how much computation will be sufficient to cover most of the conformational space and obtain a favourable result. In order to overcome this problem, the average linkage cluster analysis (ALCA) is introduced in this work instead of an arbitrary number such as 100 times used previously by our method. When the MCSA begins to work, n (here n=20) conformational candidates of the modelled loop are generated with different random seeds. Then a matrix composed of the root mean square deviations (RMSDs) between each pair of these conformations is calculated, and an Average Linkage Cluster Analysis (ALCA) is performed to divide the candidates into different structural classes. When another new iteration of MCSA works, *n* new candidates are produced and classified by ALCA again. The computation will not be stopped until new structural classes are no longer developed. Otherwise, another *n* simulation and ALCA will be continued until the structural classes of the last two clustering results are matched or the maximum simulation number (in this paper it is 1000) is reached. The ALCA procedure mainly provides a suitable sampling judgement in conformational space and will guarantee the sampled candidates to cover the most accessible region.

The ALCA mainly involves some typical average conformations to cluster the candidates. At first, it treats each candidate as a separate conformational class. Then the RMSDs of the average conformations between each pair of typical classes are calculated (see Equation 3). Here RMSDavg is the RMSD of the typical average conformation of each class. M and N are conformational classes, i and j are candidates of M and N classes respectively, C_M and C_N are numbers of candidates which belong to M and N classes respectively. After the average RMSD is compared with a given threshold, all of the conformational candidates are reformed into new classes, the classes whose average conformational RMSDs are less than the threshold are grouped into the same new class. This procedure will repeat until the classes are converged and unchanged. When new candidates are produced, they are also treated as new classes separately. Then they are clustered together with the previous calculated classes. The threshold for ALCA is set to be 1.0 Å for a four or five-residue loop, 1.5Å for a six or seven-residue loop and 2.0Å for all longer loops.

$$RMSD_{avg} = \sum_{i,j} RMSD_{M_i N_j} \Big/ C_M \bullet C_N \qquad (3)$$

Some improvements have also been made on other fields, especially on the Grid-mapping method. Here, we map the multiple-dimension assay of grid information onto a one-dimension assay. This management greatly reduces the occupation of the memory, so that the program can deal with a much larger protein and a more complex system. It lowers the demands for computer hardware and raises the computational speed. The final step-minimisation includes 400 SD steps in order to eliminate the acute non-bonded interaction caused by the coarse energy function. Then the conformation with lowest energy is further optimised thoroughly to the reasonable result.

Table 1 *Structural information of the selected target loops in this work*

PDBid	Loop starting number	Loop Length	Type [a]	Resolution (Å)	Area [b] (Å^2)	Sequence
1utg	46	4	αα	1.34	291.0	DSLP
1utg	27	5	αα	1.34	510.0	EFEPD
1poa	101	6	αα	1.50	350.0	AGAPYN
2abk	149	7	αα	1.85	366.0	QFAPGKN
2abk	100	8	αα	1.85	551.0	HNGEVPED
2tgi	29	4	αβ	1.80	457.0	GWKW
2phy	24	5	αβ	1.40	178.0	DGLAF
2rn2	58	6	αβ	1.48	438.0	ALKEHC
1xnb	156	7	αβ	1.49	357.0	HGMNLGS
5icb	15	8	αβ	1.50	665.0	AKEGDPNQ
1frd	22	4	βα	1.70	257.0	EETT
1poa	79	5	βα	1.50	386.0	KGGNN
1mrj	105	6	βα	1.60	218.0	LPYSGN
2rn2	121	7	βα	1.48	538.0	VKGHAGH
1dpe	8	8	βα	2.00	250.0	EGSPEGFN
7rsa	75	4	ββ	1.26	309.0	SYST
1iro	7	5	ββ	1.10	292.0	TVCGY
2phy	97	6	ββ	1.40	589.0	DYQMTP
1xnb	43	7	ββ	1.49	443.0	TTGSPFR
1hyt	105	8	ββ	1.70	474.0	HYSQGYNN

[a] α and β describe the types of anchoring secondary structures of the loop.
[b] The area is the sum of solvent accessible area of each loop residue calculated by DSSP [12].

Results and discussion

Loop selection

Most structure modelling programs perform better in calculating internal segments or regular secondary structures than in calculating surface loops. In order to test our program on more complicated cases, twenty target surface loops that connect different types of secondary structures and have different lengths were selected (see Table 1). All of these loops were chosen from 364 representative protein structures provided by the FSSP database [10,11] according to the following criteria: have crystal structures solved with resolution better than 2.0Å, have a solvent accessible area more than 30Å^2 per residue, and have a single unbroken peptide chain. The last condition is convenient for CHARMm operation.

Accuracy and efficiency

The computational results of the 20 loops are listed in detail (see Table 2).

According to the algorithm, the number of necessary simulations depends on the flexibility of the modelled loops, which can be deduced from the solvent accessible area of the loop. Therefore, the CPU times in calculating loops of the same length may vary dramatically. The results of ALCA calculations can illustrate the conformational variability of each loop clearly. For example, the solvent accessible area of loop 5icb_15_8 is 665.0 Å^2, it takes 420 simulations to converge at 109 structural clusters. Among the 109 clusters, there are nine clusters each of which contain at least 10 conformations and 24 clusters each of which comprise at least five conformations. The other clusters are very small, and 39 clusters only have one single conformation. Here the energies and RMSDs of the conformations in the most populated nine clusters of the loop 5icb_8_8 are plotted (see Figure 2). The figure shows that cluster 7, in which the crystal structure is located, consists of many low-energy conformations. The lowest one is picked out for further minimisation and leads to the final result (RMSD = 0.34Å, see Figure 3). While in cluster 6, which has a large RMSD, there are few low-energy conformations. Contrary to loop 5icb_15_8, loop 1dpe_8_8 also has eight residues but a solvent accessible area of 250.0 Å^2, it takes only 80 simulations and nine clusters to cover the conformational space.

A key step is to pick out the near native conformation from hundreds of candidates. After MCSA-simulation, all the candidates are minimised by 400 cycles of the SD routine and sorted according to their CHARMm energy. In fact the native-like structures can not be distinguished in most cases (comparing column 5 and column 6 in Table 2) since they do

Table 2 *Computational results of the 20 loops*

Loop id [a]	Simulation [b]	Threshold (Å)[c]	Classes[d]	rmsd (Å)[e]	rmsd (Å)[f]	rmsd (Å)[g]	CPU (s)[h]
1utg_46_4	40	1.0	5	0.34	0.58	0.12	4
1utg_27_5	140	1.0	20	0.47	0.93	0.45	13
1poa_101_6	360	1.5	61	0.51	0.92	0.59	12
2abk_149_7	460	1.5	78	0.61	0.65	0.47	22
2abk_100_8	1000	2.0	183	1.40	1.79	1.33	34
2tgi_29_4	300	1.0	60	0.55	0.55	0.39	11
2phy_24_5	360	1.0	87	0.35	0.70	0.39	8
2rn2_58_6	460	1.5	85	0.79	1.04	0.52	15
1xnb_156_7	380	1.5	75	1.03	1.15	0.39	15
5icb_15_8	420	2.0	109	0.92	0.92	0.34	33
1frd_22_4	100	1.0	5	0.26	0.49	0.09	6
1poa_79_5	380	1.0	141	0.55	0.55	0.12	7
1mrj_105_6	380	1.5	19	0.48	0.82	0.36	14
2rn2_121_7	420	1.5	144	0.52	1.22	0.58	18
1dpe_8_8	80	2.0	9	0.64	0.94	0.62	27
7rsa_75_4	40	1.0	6	0.14	0.31	0.15	5
1iro_7_5	160	1.0	50	0.40	0.54	0.13	6
2phy_97_6	300	1.5	58	0.93	0.93	0.38	20
1xnb_43_7	300	1.5	61	0.63	0.93	0.18	21
1hyt_105_8	320	2.0	56	1.42	1.42	1.20	43

[a] The loop id is composed of PDB id, starting residue number and length of loop.
[b] The number of MCSA simulations.
[c] The threshold used for ACLA computation in dividing the simulated loops into structural classes.
[d] The number of final structural classes after the ACLA procedure.
[e] The lowest RMSD of loop backbone heavy atoms between the calculated candidates and the crystal structure.
[f] The RMSD of loop backbone heavy atoms between the result of the lowest CHARMm energy after MCSA simulation and the crystal structure.
[g] The RMSD of loop backbone heavy atoms between the result of the lowest CHARMm energy after minimisation and the crystal structure.
[h] The average CPU time for one single MSCA running on PII-350-LINUX system.

not have the lowest energy. However, candidates with the lowest energy are often located near the native structure. On the another hand, since the loop often lies on the surface of protein, its side chains has great flexibility to extend into the solvent. Therefore the conformation of the side chain is so difficult to be determined that it often leads to relatively lower precision, as is the case in our work. When the loop acts as the active region and associates with other molecules, its side chain conformation may be fixed in a certain degree. Here only the RMSDs of the heavy backbone atoms between the completely minimised conformation with the lowest energy and the crystal structure are provided (see the last column of RMSD value in Table 2). The averaged deviations of backbone heavy atoms for four to eight-residue-loops are 0.19, 0.27, 0.46, 0.41 and 0.87Å, respectively. The minimised results show that this method has high reliability in obtaining satisfactory predicted loop structures. Moreover, we have compared our results with other methods [1,3,5,6]. Because the calculated cases are different, we can not compare the results precisely and directly. Here we adopt the average RMSD value of backbone heavy atoms as a basic judgement. Contrasting to the value with 0.6 Å in the Multiple Copy Sampling method [5], 0.7 Å in the Scaling-Relaxation method [6], and 1.0 Å in another Monte Carlo Simulated Annealing method [3], our method has high performance with 0.44 Å.

This program has been compiled and tested on SGI-indy/indigo2/o2-irix5.2/6.3 and PII-Linux platforms successfully. All the above calculations are run on the PII-350-Linux system, and the averaged CPU time of one single MCSA for the four, five, six, seven and eight-residue loops are about 7, 9, 15, 18 and 36 seconds, respectively. Therefore, 45 minutes are enough for a five-residue loop with 300 expected MCSA cycles, and five hours are sufficient for an eight-residue loop with 500 simulations. For most cases, the above computations are able to cover the overall conformational space with confidence.

Algorithm analysis

In order to analyse the factors affecting the final results, we made some adjustments to the parameters and settings of the

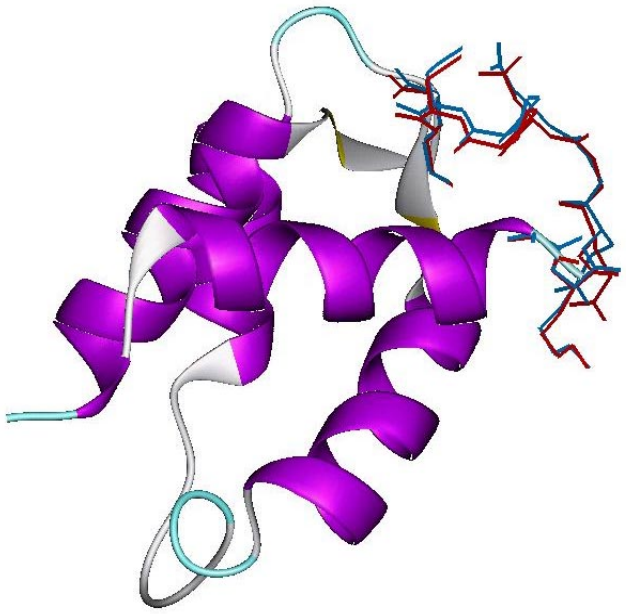


Figure 3 ***Comparion between the minimised result of the lowest CHARMm energy and the crystal structure of the loop 5icb_15_8.*** *The cartoon is the protein framework of 5icb, the red structure is the final calculated loop conformation, the blue one is the crystal structure*

simulation system. We especially concentrated on which item may be potentially improved, the MCSA algorithm or the energy function.

In this work, only the heavy atoms are calculated in MCSA procedure. In order to be consistent with the standard CHARMm topology file, the polar hydrogen atoms are considered. The adding of H atoms brings more non-bonded interactions, more rotary dihedral angles and consequently expends more CPU time than the original calculation. However it does not improve the final results

Another attempt is to replace the energy function with a more ideal one, the RMSD between the calculated conformation of the loop and the crystal structure in PDB. Four loops, 2abk_100_8, 5icb_15_8, 1dpe_8_8 and 1hyt_105_8 have been tested with the pseudo energy function. The results show that for each of the four loops, the computations converged in two cycles of simulation, all the conformational candidates were located in only one cluster after the ALCA procedure, and the final RMSDs of the backbone atoms were 0.12 0.26 0.41 and 0.18Å respectively. Although the RMSD pseudo energy function cannot represent the real situation because it only has one global energy minimum while there are many local minima in the real conformational space, it still indicates that our algorithm performs well with a perfect energy function.

One advantage of the method is the implementation of the simplified energy function, which can greatly speed up the calculation and get reasonable results by an additional minimisation procedure. Considering the classic complicated CHARMm or AMBER force fields often adopted to model protein structures accurately, the energy function including the standard 6-12 Lennard-Jones potential and electrostatic energy has been incorporated into the MCSA-simulation and tested on the above target loops [12]. However, for the 6-12 Lennard-Jones potential is partially sharp, the simulation will suffer from many insurmountable high energy points resulting from arbitrary rotations of dihedrals, therefore it does not make obvious improvements on the results and has less power to explore the loop conformation. In fact, the performance of the simplified energy function is always comparable to or better than that of the complicated force field both in the speed and accuracy.

In general, optimising the energy function and parameters is not a good and feasible means to improve the method. A more practical way is to exploit the knowledge derived from PDB further. In recent years, more extensive studies on protein loops have been published [13-17], we also carried out a detailed loop conformational analysis based on all structures in PDB[18]. The integration of the statistical results into this program is in progress.

Acknowledgements This work was supported in part by Chinese National High Technology Development Program (863-103-13-03-05) and the National Science Foundation of China, NSFC (29703001).

Supplementary material available statement The source codes of the program (in C++) are available from the authors upon request.

References

1. Collura, V.; Higo, J.; Garnier, J. *Protein Sci.* **1993**, *2*, 1502-1510.
2. Higo, J.; Collura, V.; Garnier, J. *Biopolymers* **1992**, *32*, 33-43.
3. Carlacci, L.; Englander, S. W. *Biopolymers* **1993**, *33*, 1271-1286.
4. Vasmatzis, G.; Brower, R.; Delisi, C. *Biopolymers* **1994**, *34*, 1669-1680.
5. Zheng, Q.; Rosenfeld, R.; DeLisi, C.; Kyle, D.J. *Protein Sci.* **1994**, *3*, 493-506.
6. Zheng, Q.; Kyle, D.J. *Proteins* **1996**, *24*, 209-217.
7. Zhang, H.; Lai, L.; Wang, L.; Han, Y.; Tang, Q. *Biopolymers* **1997**, *41*, 61-72.
8. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J.Comp.Chem.* **1983**, *4*, 187-217.
9. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Jr, B.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J.Mol.Biol.* **1977**, *112*, 535-542.
10. Holm, L.; Sander, C. *Nucleic.Acids.Res.* **1996**, *24*, 206-209.
11. Holm, L.; Sander, C. *Nucleic.Acids.Res.* **1997**, *25*, 231-234.

12. Cornell, W. D.; Cieplak, P.; Bayly, C. L..; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J.Am.Chem.Soc.* **1995**, *117*, 5179-5197.
13. Ring, C.S.; Kneller, D.G.; Langridge, R.; Cohen, F. E. *J.Mol.Biol.* **1992**, *224*, 685-699.
14. Van Vlijmen, H. W.; Karplus, M. *J.Mol.Biol.* **1997**, *267*, 975-1001.
15. Kwasigroch, J. M.; Chomilier, J.; Mornon, J. P. *J.Mol.Biol.* **1996**, *259*, 855-872.
16. Oliva, B.; Bates, P. A.; Querol, E.; Aviles, F. X.; Sternberg, M. J. *J.Mol.Biol.* **1997**, *266*, 814-830.
17. Geetha, V.; Munson, P. J. *Protein Sci.* **1997**, *6*, 2538-2547.
18. Li, W.; Wang, R.; Liang, S.; Lai, L.; Han, Y. *Protein Eng.* submitted.