Modeling the Third Loop of Short-chain Snake Venom Neurotoxins: Roles of the Short-range and Long-range Interactions

Zhijie Liu, Weizhong Li, Hongyu Zhang, Yuzhen Han, and Luhua Lai*
Institute of Physical Chemistry & College of Chemistry and Molecular Engineering, Peking University, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, Beijing, P.R. China

ABSTRACT **The influence of long-range interactions on local structures is an important issue in understanding protein folding process and protein structure stability. Using short-chain snake venom neurotoxin as a model system, we have studied the conformational properties of eight different loop III sequences either in the environment of one of the short-chain neurotoxin, erabutoxin b (PDB ID 1nxb), or in free state by Monte Carlo simulated annealing method. The surrounding protein structure was found to be crucial in stabilizing the loop conformation. Although all the eight peptides prefer type V β turn in solution, three of them (KPGI, KPGV, KSGI) turn to type II β turn and the other five (KKGI, KKGV, KNGI, KQGI, and KRGV) are confined to more rigid type V β turn conformation in the protein structure. Using flexible tetra-glycine-peptide to screen the backbone conformational space in the protein environment also validates the results. This study shows that long-range interactions do contribute to the stability and the types of conformation for a surface loop in protein, while short-range interactions may only provide candidate conformations, which then have to be filtered by the long-range interactions further. Proteins 2001;42:6–16.**

Key words: long-range interactions; loop modeling; protein structure stability; Monte Carlo simulated annealing; snake venom neurotoxin; short-range interactions

INTRODUCTION

The interactions in protein can be classified into short-range and long-range interactions according to their effective region.[1] Short-range interaction mainly lies among the neighboring residues in sequence and determines the conformations of single residue and local structures. Long-range interaction exists in the neighboring residues of the structure and is crucial for the protein folding and stability of the advanced protein structure. Though the two kinds of interactions determine the protein structure cooperatively, the long-range interactions have strong influence on the protein structure and function, a conclusion which was validated both by experimental and computational results.[2–4] For example, the lack of structure of the GH loop (G132-S142) of VP1 capsid protein in aqueous solution showed that its conformational determinants sufficient for the structure stabilization were dictated by the interactions with protein environment.[2] It was also found that N-terminal sequence (1-40) in hen lysozyme had little secondary structure in aqueous solution; the stabilization of its structure in native protein was facilitated predominantly by long-range interactions with the C-terminus.[3]

Neurotoxins are the major toxic components of the snake venoms. They bind to the nicotinic acetylcholine receptor proteins on the postsynaptic membranes and cause skeletal muscle paralysis.[5] Two groups of α-neurotoxins, long-chain toxins and short-chain toxins, have been found to take similar three-dimensional structures.[6] In Swiss-Prot (Release 33),[7] there are 43 short-chain neurotoxins, all of which are highly homologous with sequence identity above 48% (alignment by HSSP[8]) (see Table I). One x-ray crystal structure (erabutoxin b)[9] and three NMR structures[10–12] are known for short-chain neurotoxins. Structure comparison of the four structures showed that they are almost identical structures, with main chain root-mean-square deviations (RMSD) less than 1.7 Å.

Loop II and loop III in short-chain neurotoxins are important because they act as toxic residues and are involved in binding with acetylcholine receptor.[6] For the β-bend in loop III (residues 47–50), 19 of these toxins have sequence KPGI as found in erabutoxin b. For the remaining 24 toxins, four of them have sequence of KKGV, 6 KKGI, 6 KPGV, 2KNGI, 1KRGV, 1KQGI, 4KSGI. In all 43 short-chain neurotoxins, Lys and Gly are found in the first and the third positions of the loop.

Roos et al. studied the conformational properties of the blocked tetrapeptides for the third loop by experimental methods and potential energy calculations.[13–15] In the eight tetrapeptides studied they found three peptides (KPGI, KSGI, KPGV) take β-turn conformation, while the other five exist as random coil in water but show preference for helical structure in TFE. Their results indicate

Grant sponsor: Department of Science and Technology of China; Grant sponsor: National Natural Science Foundation of China; Grant number: 29525306; Grant number: 29703001.

*Correspondence to: Luhua Lai, Institute of Physical Chemistry & College of Chemistry and Molecular Engineering, Peking University, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, Beijing 100871, P.R. China. E-mail: lai@mdl.ipc.pku.edu.cn

Received 10 March 2000; Accepted 10 August 2000

TABLE I. Sequences and Alignment (From HSSP[4]) of 43 Short-chain Venom Neurotoxins

No.	Sequence	PDB ID	Name
	12345678901234567890123456789012345678901234567890[a]123456789012		
1	RICFNHQSSQPQTTKTCSPGESSCYNKQWSDFRGTIIERGCGCPTVKPGIKLSCCESEVCNN	1 FRA	nxs1_latse
2	RRCFNQQSSQPQTNKSCPPGENSCYRKQWRDHRGTIIERGCGCPTVKPGIKLRCCQSEDCNN		nsc_latcr
3	RRCFNHPSSQPQTNKSCPPGENSCYNKQWRDHRGTIIERGCGCPTVKPGIKLRCCQSDDCNN		nxsa_latcr
4	RRCYNQQSSQPKTTKSCPPGENSCYNKQWRDHRGSITERGCGCPTVKPGIKLRCCESEDCNN		nxs2_latco
5	RRCFNQQSSQPKTTKSCPPGENSCYNKQWRDHRGSITERGCGCPKVKPGIKLRCCESEDCNN		nxsd_latco
6	RRCFNQQSSQPQTNKSCPPGENSCYRKQWRDHRGTIIERGCGCPTVKPGVKLRCCQSEDCNN		nxsc_latla
7	RRCYNQQSSQPKTTKSCPPGENSCYNKQWRDHRGSITERGCGCPKVKPGIKLRCCESEDCNN		nxsc_latco
8	RRCFNHPSSQPQTNKSCPPGENSCYNKQWRDHRGTIIERGCGCPQVKSGIKLRCCQSDDCNN		nxsb_latcr
9	RRCFNHPSSQPQTNKSCPPGENSCYNKQWRDHRGTITERGCGCPQVKSGIKLTCCQSDDCNN		nxsb_latla
10	CHNQQSSQPPTTKTC.PGETNCYKKVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN	1NEA	nxs1_najpa
11	KICYNQPSSQHPTTKAC.PGEKNCYRKQWSDHRGTIIERGCGCPTVKPGVKLHCCTTEKCNN		nxs1_bouan
12	RICYNHQSTTPATTKSC..GENSCYKKTWSDHRGTIIERGCGCPKVKOGIHLHCCQSDKCNN		nxs1_denja
13	CHNQQSSQPPTTKTCS.GETNCYKKWWSDHRGTIIERGCGCPKVKPGVNLNCCRTDRCNN	1NOR	nxs1_najox
14	CHNQQSSQAPTTKTCS.GETNCYKKWWSDHRGTIIERGCGCPKVKPGVKLNCCRTDRCNN		nxs1_najph
15	CHNQQSSQPPTTKSC.PGDTNCYNKRWRDHRGTIIERGCGCPTVKPGINLKCCTTDRCNN		nxs1_hemha
16	CHNQQSSQPPTTKTC.PGETNCYKKQWSDHRGTIIERGCGCPSVKKGVKINCCTTDRCNN		nxs1_najme
17	ICHNQQSSQPPTIKTCSEGQ..CYKKTWRDHRGTISERGCGCPTVKPGIHISCCASDKCN		nxs1_micni
18	CCNQQSSQPKTTTNCA..ESSCYKKTWSDHRGTRIERGCGCPQVKPGIKLECCHTNECNN		nxs1_enhsc
19	CCNQQSSQPKTTTNCA..ESSCYKKTWRDFRGTRIERGCGCPQVKPGIKLECCHTNECNN		nxs1_hydla
20	RICYNHQSTTPATTKSC..GENSCYKKTWSDHRGTIIERGCGCPKVKRGVHLHCCQSDKCNN		nxs1_denyi
21	CCNQQSSQPKTTTNCA..ESSCYKKTWSDHRGTRIERGCGCPQVKKGIKLECCHTNECNN		nxs1_hydcy
21	CCNQQSSQPKTTTNCA..ESSCYKKTWSDHRGTRIERGCGCPQVKSGIKLECCHTNECNN		nxs2_enhsc
23	CHNQQSSQPPTTTHCSGGETNCYEKRWHDHRGTIIERGCGCPTVKPGVKLNCCTTDKCNN		nxs1_bouch
24	CCNQQSSQPKTTTTCPGGVSSCYKKTWRDHRGTIIERGCGCPRVKPGIRLICCKTDECNN		nxs1_acaan
25	CHNQQSSQTPTTQTC.PGETNCYKKQWSDHRGSRTERGCGCPTVKPGIKLKCCTTDRCNK		nxs2_hemha
26	CCNQQSSQPKTTTNCA..GNSCYKKTWSDHRGTIIERGCGCPQVKSGIKLECCHTNECNN		nxs1_astst
27	CCNQQSSQPKTTTICAGGESSCYKKTWSDHRGSRTERGCGCPHVKPGIKLTCCKTDECNN		nxs1_pseau
28	ICHNQQSSQPPTIKTC.PGETNCYKKQWRDHRGTIIERGCGCPSVKKGVGIYCCKTDKCN		nxs2_najhh
29	CCNQQSSQPKTTTDCA..DDSCYKKTWKDHRGTRIERGCGCPQVKPGIKLECCKTNECNN		nxs4_aipla
30	CCNQQSSQPKTTTDCA..DNSCYKKTWQDHRGTRIERGCGCPQVKPGIKLECCKTNECNN		nxs1_aipla
31	CCNQQSSQPKTTTDCA..DNSCYKKTWKDHRGTRIERGCGCPQVKPGIKLECCKTNECNN		nxs3_aipla
32	CHNQQSSQPPTTKTC.PGETNCYKKRWRDHRGSITERGCGCPSVKKGIEINCCTTDKCNN		nxs1_najha
33	RICYNHQSTTRATTKSCE..ENSCYKKYWRDHRGTIIERGCGCPKVKPGVGIHCCQSDKCN	1NTX	nxs1_denpo
34	ICHNQQSSQPPTIKTC.PGETNCYKKRWRDHRGTIIERGCGCPSVKKGVGIYCCKTNKCN		nxs2_najha
35	CCNQQSSQPKTTTDCA..DNSCYKMTWRDHRGTRIERGCGCPQVKPGIKLECCKTNECNN		nxs2_aipla
36	ICHNQQSSQRPTIKTC.PGETNCYKKRWRDHRGTIIERGCGCPSVKKGVGIYCCKTDKCN		nxs2_najni
37	CHNQQSSEPPTTTRCSGGETNCYKKRWRDHRGYRTERGCGCPTVKKGIELNCCTTDRCNN		nxs1_najmo
38	ICYKQQSLQFPITTVC.PGEKNCYKKQWSGHRGTIIERGCGCPSVKKGIEINCCTTDKCN		nxs3_najha
39	CHNQMSAQPPTTTRCSRWETNCYKKRWRDHRGYKTERGCGCPTVKKGIQLHCCTSDNCNN		nxs3_najmo
40	ICYKQRSLQFPITTVC.PGEKNCYKKQWSGHRGTIIERGCGCPSVKKGIEINCCTTDKCN		nxs4_najha
41	CHNQQSSQTPTTTGCSGGETNCYKKRWRDHRGYRTERGCGCPSVKNGIEINCCTTDRCNN		nxs1_najat
42	CHNQQSIQTPTTTGCSGGETNCYKKRWRDHRGYRTERGCGCPSVKNGIEINCCTTDRCNN		nxs1_najat
43	RICLNQQQSTPEDQPTNGqiKTDCQNKTWNTHRGSRTDRGCGCPKVKPGINLRCCKTDKCNE		nxs1_bunfa

[a]The underlined sequences belong to the third loops of neurotoxins.

that short-range interactions are important in determining the conformation of the first group, but additional long-range interactions are necessary in defining the β-bend conformation of the latter group in protein.

We have developed an efficient Monte Carlo simulated annealing program to study protein loop conformations,[16–18] which was shown to reproduce loop conformation reliably for proteins. In order to study the long-range interactions on the conformations of the loop III in short-chain snake neurotoxins, we used the Monte Carlo simulated annealing program to study eight different loop III sequences in the protein structure and in solution separately. Because short-chain neurotoxins are highly homologous, it is reasonable to assume they all take similar three-dimensional structures. The calculation of one homologous sequence KPGV validates that loop III in the NMR structure of toxin from naja naja oxiana venom can be reproduced well by simulation with the protein environment of erabutoxin b (see Results and Discussion).

Short peptides in proteins can be classified into two groups: Conserved peptides will always take the same conformation independent of protein environment; variable peptides will change conformation under different protein environment.[19–21] Structurally conserved peptides may act as nucleation site in protein folding. In this work, we have carefully studied the conformations of loop III in short-chain neurotoxins, the roles of short-range and

TABLE II. Lowest Energy Conformations for the Eight Loop III Peptides in the Protein Environment of Erabutoxin b

Peptides	Torsion angles				Energy (kcal/mol)	Secondary structure
	ϕ-2	ψ-2	ϕ-3	ψ-3		
KPGI	−39.6	113.8	106.9	49.3	−160.34	β-turn-2
KPGV	−39.4	114.4	106.6	49.5	−158.59	β-turn-2
KSGI	−75.3	82.0	86.9	70.5	−187.52	β-turn-2
KKGI	−71.0	87.6	117.9	−72.0	−188.73	β-turn-5
KKGV	−69.2	109.0	100.3	−83.1	−188.64	β-turn-5
KNGI	−67.4	86.9	106.1	−58.8	−220.71	β-turn-5
KQGI	−71.9	84.6	119.5	−74.5	−213.88	β-turn-5
KRGV	−68.0	79.3	114.8	−65.9	−200.43	β-turn-5

long-range interactions in the loop formation, and whether the short-range interaction can sufficiently force the peptide to form conserved structure.

MATERIALS AND METHODS

Data

The crystal structure of erabutoxin b (PDB code 1nxb, resolution 1.38Å)[9] was used as the protein environment. The PDB codes for the three NMR structures are 1nea,[10] 1nor,[11] and 1ntx.[12] Sequences for short-chain neurotoxins were taken from Swiss-Prot (release 33).[7] Sequence Alignments were taken from HSSP[8] and shown in Table I. Eight different sequences of loop III of short-chain neurotoxins were studied and are listed in Table II.

Conformational Analysis

LPSA,[16,18]a program for modeling protein loops by Monte Carlo simulated annealing method, was employed in this work. In this program the loop is developed from its N-terminal in extended conformation, then two dummy atoms, N′ and CA′, are grown at the end of the loop. Based on a simplified soft-sphere potential which allows atomic interpenetration in certain degree (see Equation 1; k_s is the coefficient of simplified soft-potential, d is the distance between a pair of atoms, and d_0 is the sum of their van der Waals radii) and harmonic constrained potential which guarantees the smooth closure of the loop (see Equation 2; k_h is the coefficient of harmonic constrained potential, $\boldsymbol{r}$(x)s are the position vectors of the dummy atoms and their reference atoms), and combined with grid-mapping accelerating method, the simulated annealing process continues to screen the conformational space entirely. In order to eliminate the coarseness of above simplified potential and evaluate the conformational energies precisely, an additional CHARMm minimization step is performed. The program can take long-range forces into consideration and precisely model the protein loop conformation with high speed. LPSA can be obtained from anonymous ftp server of the author's laboratory: ftp://mdl.ipc.pku.edu.cn.

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

$$Eh=k_h\{[r_{(N)}-r_{(N')}]^2+[r_{(CA)}-r_{(CA')}]^2\} \quad (2)$$

The simulation procedures for the loops under protein environment are the same as in references 16 and 18. In order to remove bad contacts, the crystal structure of erabutoxin (1nxb) was optimized with 200 steps of steepest descent, 200 steps of conjugated gradient, and 800 steps of Adopted Basis Newton-Raphson method successively using CHARMm with polar hydrogens. The dielectric constant was set to be one. For each loop, 100 conformations were generated, which were then submitted to energy minimization by 400 steps of SD, CONJ, and ABNR methods successively using CHARMm. During these procedures, only loop atoms were allowed to move in the optimization, while others were kept fixed. It is incontestable that optimizing the whole protein structure after the enclosure of the loops will be more appropriate. However, because the side-chains of loop residues were mainly extended out they had less influence on the protein environment. On the other hand, the sequences of different loops primarily had differences at the second residue. The similar sequences inspired us that it would introduce less distortion that the protein environments were kept fixed during CHARMm minimization. More important, this treatment would greatly reduce the computational demand. For example, when reproducing the crystal structure of loop KPGI, the minimization of the whole protein took about 21 min, while only 37.5 sec were spent in minimizing loop when protein environment was kept fixed. The computational speed was accelerated about 30 times. The energy coefficients for the soft-sphere potential and the constrained potential are 10 and 100 kcal/(mol.atom. Å^2), respectively. The simulations were carried out on SGI O_2/R10000, Indy/R4400, or Indigo^2 /R4000.

The conformation of free peptides was analyzed by using SA program.[22,23] SA is a program for peptide conformational analysis using Monte Carlo simulated annealing protocol and based on ECEPP/2 force field.[24,25]. The N-terminal of each peptide was blocked by acetyl, and the C-terminal was blocked by N-methylamide groups. One hundred candidate conformations were generated by 100 runs of annealing for each peptide. Main chain ω angles were kept to 180° during the simulation, except for those before proline (which were kept to either 180° or 0°). In order to compare with those obtained under protein environment, all the conformations were then optimized using

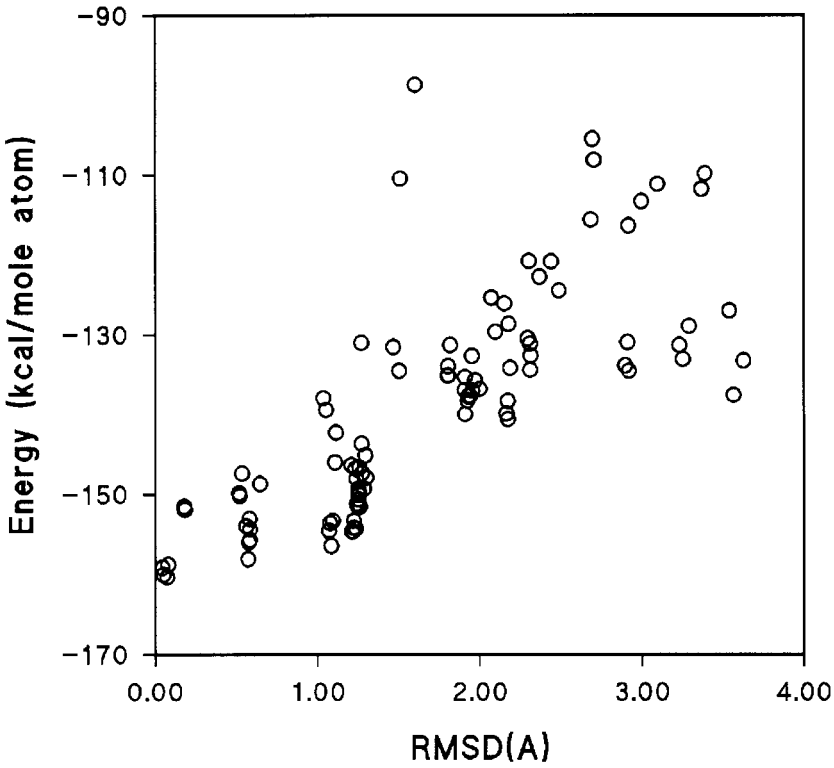

Fig. 1. Reproduce the loop III conformation in erabutoxin b. Energy verse RMSD from crystal structure of erabutoxin b

CHARMm, and the dielectric constant was set to be the value of vacuum, i.e., one unit, for simplicity.

Conformational Clustering Method[18]

Average Linkage Cluster Analysis (ALCA) method was performed to divide the conformations into different structural groups. 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 of which the average conformational RMSDs are less than the threshold are grouped into the same new class. This course will repeat until the classes are converged and unchanged.

$$\mathrm{RMSD}_{\mathrm{avg}}=\frac{\sum_{i,j}\mathrm{RMSD}_{\mathrm{MiNj}}}{C_M \cdot C_N} \quad (3)$$

RESULTS

Conformation of the Native Loop III Sequence in Erabutoxin b Structure

The loop III of erabutoxin (KPGI) was first calculated to see whether the crystal structure could be reproduced by the calculation procedure. From the 100 conformations obtained from the simulation, a good correlation between CHARMm energy and RMSD from the crystal structure was established (see Fig. 1). The lowest energy conformation deviates from the crystal structure only by 0.05 Å, which shows that we can reliably model the loop conformation by LPSA simulation.

Conformation of the Loop III Sequence From Naja Naja Oxiana Venom in Erabutoxin b Structure

In all the three known NMR structures of short-chain neurotoxins, one of the loop III has sequence of KPGI, while the other two have sequence of KPGV. We took neurotoxin from naja naja oxiana venom (1nor) to test whether the KPGV conformation in 1nor can be reproduced when erabutoxin structure is used as protein environment. We first compared the conformations of loop III among the 19 NMR structures of 1nor by superposing the loops with one additional residue at both ends and found that the RMSDs are in the range of 0.19–1.36Å. We then compared the lowest energy conformation for KPGV obtained in the environment of erabutoxin b with the NMR structures. The value of RMSD (between 1.36 and 1.63Å)

shows that KPGV conformation can be reproduced under the environment of erabutoxin. The successful reproduction of KPGV offers us a great support to simulate other loop III sequences under erabutoxin environment.

Conformations of the Other Six Loop III Sequences in Erabutoxin B Structure

All the other loop III sequences were calculated in the same way as KPGI. One hundred conformations were obtained for each loop sequence. The lowest energy conformations were regarded as the expected conformations in protein environment. The secondary structures of these conformations are listed in Table II. Definition of β-bend was based on main chain Cα distances and backbone torsion angles.[26] The results show that all the eight peptides take β-bend structure, but belong to different types. The lowest energy conformations for KPGI, KPGV, KSGI belong to type II β turn, while KKGI, KKGV, KNGI, KQGI, and KRGV belong to type V β turn.

The Accessible Conformational Space of Loop III in Erabutoxin b Strucutre

In order to study the influence of protein environment on local structures, a flexible polytetraglycine peptide $(Gly)_4$ is used to search the accessible conformational space of loop III in erabutoxin b structure. As glycine is the smallest amino acid residue with no side-chain, polyglycines should be able to give the full conformational space for all possible sequences. One hundred conformations were generated by using LPSA which were clustered into nine groups with an RMSD cutoff of 1.6Å.[18] Except for three small groups, which may have resulted from conformational perturbation and fluctuation, six major groups were found (see Fig. 2). Comparing the lowest energy conformations for the eight loop III sequences in the protein environment of erabutoxin b, KPGI and KPGV were found to belong to group I (with a conformational population of 12%) and KKGI, KKGV, KRGV, KQGI, KNGI, and KSGI to group II (with a conformational population of 7%). Polyglycine peptide simulation shows that there are six possible groups of conformation restricted by the erabutoxin b structure. The eight short-chain snake venom neurotoxin loop III sequences only use two of them. Because the peptide is so flexible and has no C_β atom and chirality of C_α atom, groups III, IV, V, and VI (with conformational populations of 45%, 27%, $3\%_A$ and $\%3_{B,}$ respectively) may come from the intrinsic properties of polyglycines or be accessible to other sequences. It may be possible to make mutations in this loop and force it to take these conformations.

Conformations of the Eight Corresponding Isolated Peptides

One hundred conformations for each of the eight blocked peptides were calculated by the SA program and then clustered into groups according to the RMSDs of backbone dihedral angles. Because turn types are determined mainly by the backbone conformation of the two central residues, only the dihedral angles were taken into consideration in calculation. The conformational groups are listed in Table III.

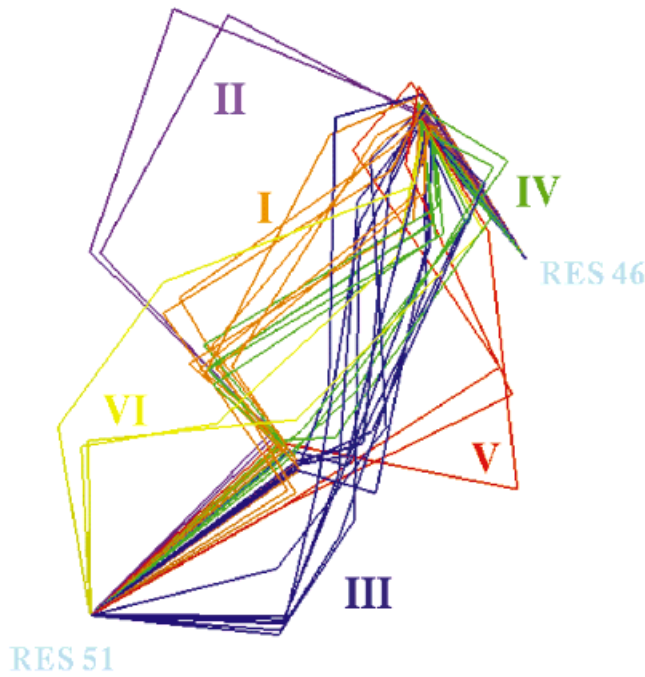

Fig. 2. The conformational groups of polyglycine under erabutoxin environment. Only the traces of Cα atoms of six residues, which include two additional residues in the beginning and the end of the loop, were illustrated. The conformational populations of groups I, II, III, IV, V, and VI are 12%, 7%, 45%, 27%, $3\%_A$, and $3\%_B$, respectively.

DISCUSSION

Effect of CHARMm Minimization on Protein Loops and Isolated Peptides

In order to remove the imprecision from the soft-sphere potential and the biases from different force fields, the CHARMm minimization was performed. The RMSDs of protein loop and isolated peptide between the unminimized and minimized conformations were figured out (see Fig. 3). It was found that, for isolated polypeptide, the RMSDs were scattered between 0.3 Å and 1.8Å (see Fig. 3a). The result implied that for the most conformations with smaller RMSDs of isolated peptides, they were predicted reliably and had good agreements between ECEPP force field and CHARMm force field, they only had been modified a little according to CHARMm parameters. While for those that had large RMSDs, there were some conflicts between the above two force fields, then the conformations underwent great changes in CHARMm optimization to remove the biases from ECEPP models. For the conformations of protein loop, the RMSDs were slightly larger than the isolated peptide, ranging mainly from 1.0Å to 2.0Å, with a few exceeding 3.0Å (see Fig. 3b). The program LPSA produced proper initial conformations, then the procedure of CHARMm optimization considered more interactions in details and made the finial result more reasonable. Thus the final results mainly built from very approximate structures. In general, the CHARMm minimization step

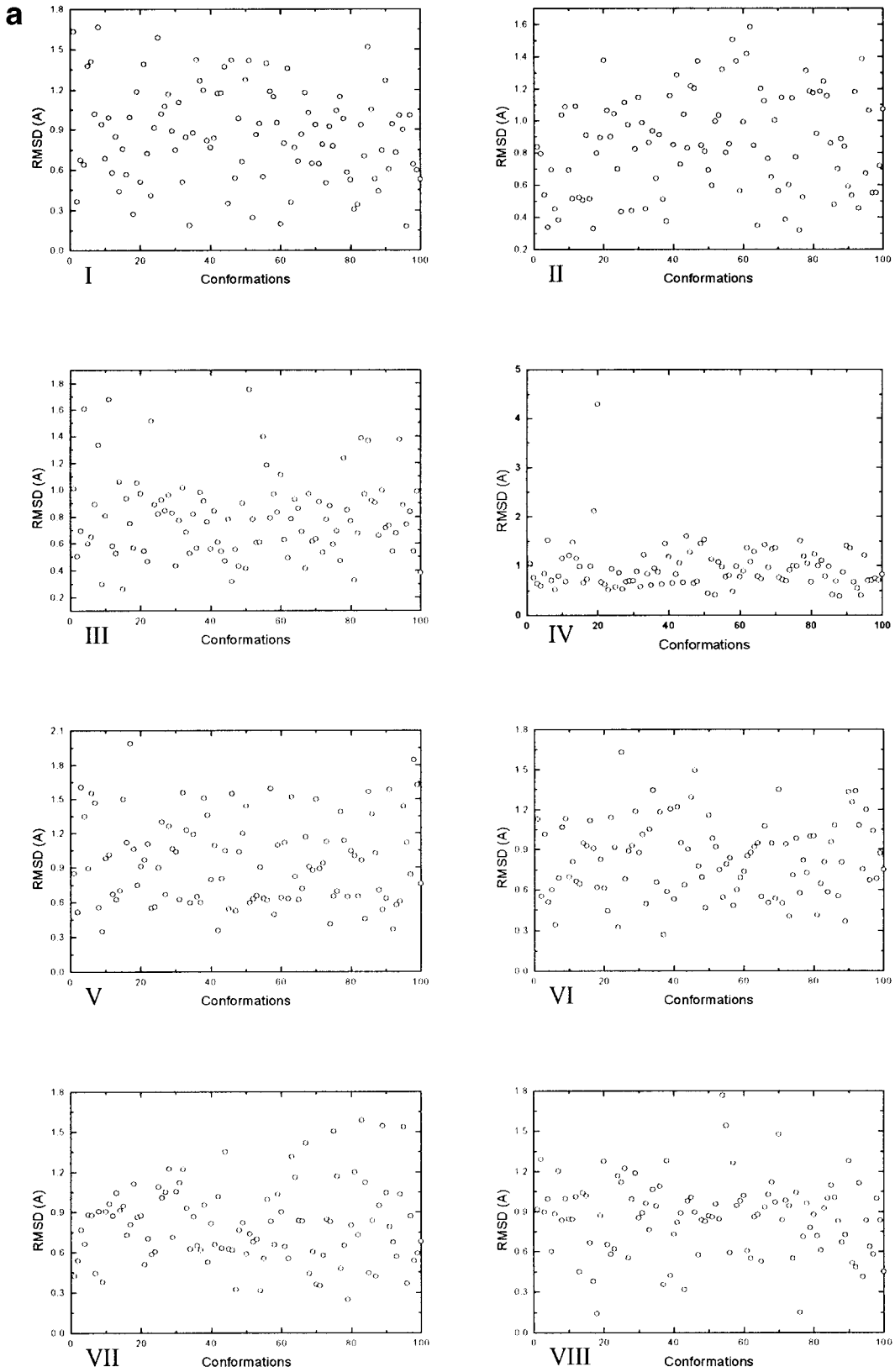

Fig. 3. Comparison of RMSDs between the CHARMm minimized conformations and the corresponding unminimized ones. **a:** The result of isolated polypeptides with SA program. **b:** The results of protein loop with LPSA program. Results of I, II, III, IV, V, VI, VII, and VIII were sequences of KKGI, KKGV, KNGI, KPGI, KPGV, KQGI, KRGV, and KSGI, respectively.

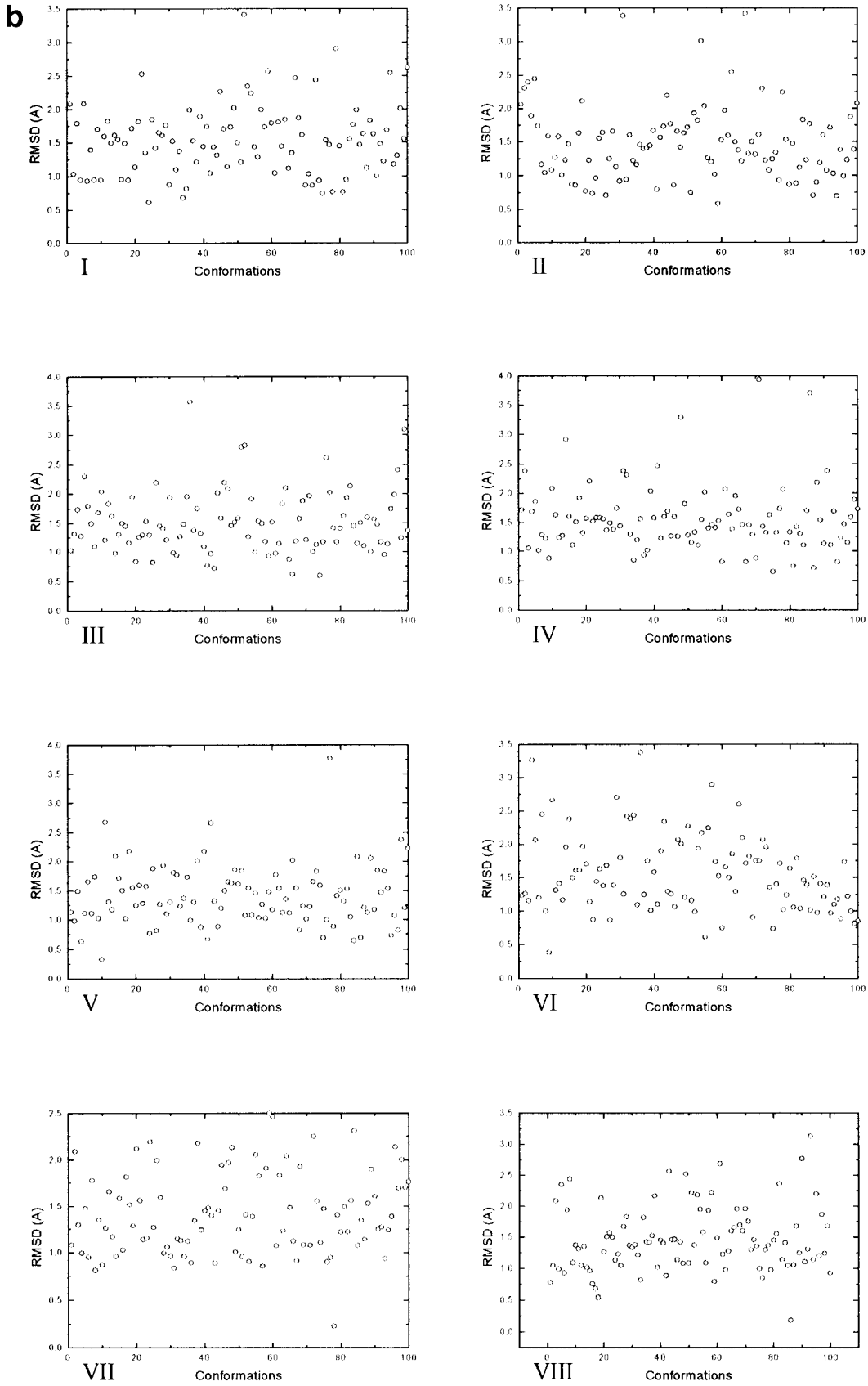

Figure 3. (Continued.)

TABLE III. Conformations of the Eight Loop III Peptides in Isolated State[†]

		Density of secondary structure (%)							
		β-turn-1	β-turn-1′	β-turn-2	β-turn-2′	β-turn-3	β-turn-4	β-turn-5	Random coil
Peptides	KPGI	27	0	10	0	6	7	43	7
	KPGV	0	0	2	0	7	13	61	17
	KSGI	16	7	7	1	0	8	49	12
	KKGI	20	7	17	0	9	0	38	9
	KKGV	8	4	4	0	8	2	64	10
	KNGI	14	5	13	0	11	2	40	15
	KQGI	11	4	3	1	11	2	51	17
	KRGV	5	4	1	0	19	2	53	16

[†]Conformations were separated into clusters according to their RMSDs of the main chain dihedral angles of the central two residues. The cut-off value is 40 degrees.

can not only find the correct conformations, but also provide the common comparable standard between the results from protein loop and isolated peptide.

All Eight Loop III Sequences Take β-Bend Conformation Under Protein Environment

Under the bulk protein environment, all the eight loops form β-bend structure. But they are separated into two distinct groups: group one contains KPGI, KPGV, and KSGI and forms type II turn; group two contains KKGI, KKGV, KNGI, KQGI, and KRGV and forms type V turn. When studying free peptides, Roos et al. also found that the two groups of peptides behave differently in solution.[15] This may come from the difference of amino acids residues at the second position. The first group has proline or serine, while the second group has either positive charged or highly polar residues. In general, the results show that side-chains and their interaction with protein environment have strong influence on the backbone structure of this loop.

Isolated Peptides Take Diverse Conformations

From Table III we can see that for all eight peptides, the most populated conformation is type V turn, with a population between 38% and 64%. At the same time, the peptides are able to take a large variety of other conformations. For KPGI, apart from 43% of type V turn, other conformations include 27% of type I turn, 10% of type II turn, 7% type IV turn, 6 % type III turn, and random coil conformation. Roos et al. also found that KPGI, KPGV, and KSGI appear to be a type V turn from NMR studies and force field calculations, but they did not have evidence of β turns for the other five peptides.[12–14] From the calculation here we can see that all eight peptides dominate in type V turns. Since all of them can reach to other conformations at the same time, less flexible peptides, like those with a proline residue, can have a detectable conformation in solution, while no single conformation can be detected for rapid changing peptides.

Long-range Interactions Play a Key Role in Stabilizing the β-Bend Structure of the Third Loop in Short-Chain Neurotoxins

Although all eight free peptides have a dominate type V β turn structure, under the protein environment of erabutoxin b three of them (KPGI, KPGV, and KSGI) change to type II β turn. The traces of Cα atoms of conformations for eight loops in protein and their corresponding represent conformations of isolated peptides were superposed in Figure 4. The classification of conformations obtained from polyglycine simulation also validated the result. In protein environment, the conformational space of the peptide is highly restricted to a specific region, which may not coincide with its dominant structure when it is alone in solution. Structural analysis of short-chain snake venom neurotoxins indicated that, within 5.0 Å from the third loop in space, there was an electrostatic center in the central β-sheet which formed by polar heteroatoms in side-chains of residues LYS27 (Met in one sequence), TPR29, and GLU38 (ASP in one sequence). When the second residue of the loop was polar residue (i.e., LYS, ASN, GLN, ARG), it had strong interactions with the electrostatic center. While for non-polar residue (i.e., Pro), the first residue LYS interacted with the center strongly, thus making the loop distorted and forcing the second residue far away from the center. For less polar residue at the second position (i.e., Ser), the situation was located between the above two circumstances. All these results clearly showed that long-range interactions, which might include van der Waals steric constraint and hydrogen bond (especially the electrostatic effect), played an important role both in confining and stabilizing the β-bend structure of the third loop in short-chain neurotoxins. The loop peptide has intrinsic tendency to form β turn, but the final structure is tuned by its environment. This should be true even for the nucleation sites in protein folding, which may form some kind of structures due to their intrinsic tendency in the early stage of folding, but will have to modify their structure according to the surroundings in the late stage of folding.

Baker and coworkers had performed cluster analysis of the distance between sequence profiles generated from multiple sequence alignments to identify some local sequence motifs that transcend protein family boundaries.[27] Based on the similar distance analysis, the structural correlates of these patterns were explored. For each typical pattern of sequence segment of 3-15 residues in length from 154 different protein families, only one or a small number of local structures predominated.[28] They explored

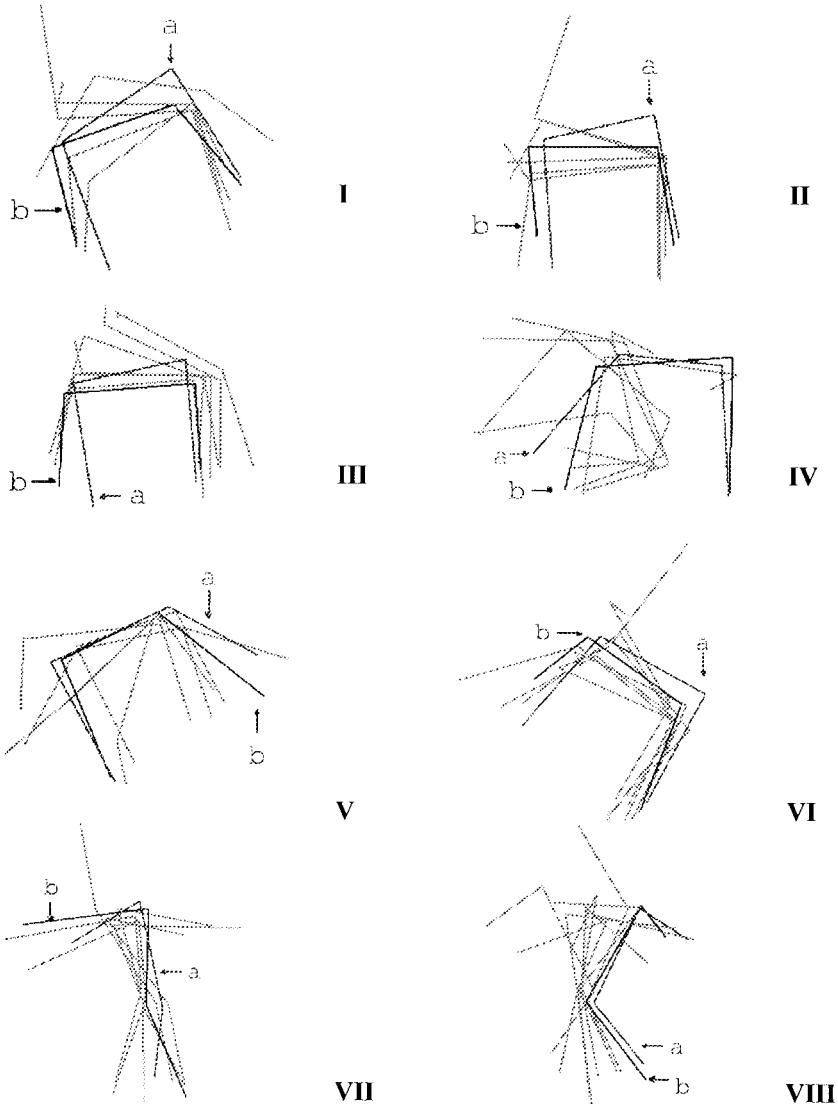

Fig. 4. The comparisons between the lowest energy conformation in protein environment and the represent conformations of each turn of each isolated peptide (only traces of Cα atoms are illustrated; I, II, III, IV, V, VI, VII, and VIII are sequences of KKGI, KKGV, KNGI, KPGI, KPGV, KQGI, KRGV, and KSGI, respectively). **a:** The lowest energy conformation of protein loop. **b:** Conformation of the nearest represent conformation of isolated peptides.

the ability of a simple simulated annealing procedure to assemble native-like structures from fragments of unrelated protein structures with similar local sequences using Bayesian scoring functions. The simulated annealing procedure can rapidly and frequently generate native-like structures for small helical proteins and better than random structures for small β sheet containing proteins. It was also found that protein environment and residue pair had specific contributions to the scoring functions.[29] Similar restriction from protein environment on local sequence was also detected in hypervariable regions of immunoglobulins. Chothia and coworkers had found that, for at least five of six hypervariable regions of antibodies, though the sequences had great diversities, the repertoire of conformations appeared to be limited to a relatively small number of discrete structural classes called "canonical structures." Some key residues of the regions that located in conserved β-sheet framework, through packing, hydrogen bonding, or the ability to assume unusual ϕ, ψ, ω torsiona angles, were responsible for the formation of regular structures. These mainly because the tight packing mode in immunoglobulin environment had great effect on the hypovariable regions, which restricted the key residues to determine the final structures.[30–33] Our conclusion had agreement with the above mentioned results. However, in contrast to the tight-packing, hypervariable regions of immunoglobulin and regular structures in proteins, our computation primarily focused on the surface loops, which were packed relatively looser and more flexibly. Recently the works in cluster analysis of loop conformations database also vali-

dated that protein loops had limited conformations and were restrained by the protein bulk.[34–37]

Van Gunster and coworkers had performed molecular dynamic simulations on conformational research of peptides.[38–40] There was good agreement on the conformations of simulated nonnative peptide fragments from proteins with the experimental results of denatured conformers of these peptides in solution. They also found that although the conformational space potentially accessible to the peptides was extremely large, very few conformers (10^1 to 10^2) were significantly populated at 20K above the melting temperature. Hansmann and coworkers adopted generalized-ensemble Monte Carlo simulations and Multicanonical Monte Carlo simulations in studying conformations of small peptides.[41–43] From the landscape of thermodynamic quantities obtained over a wide range of temperatures, it was found that the folding of some small peptides was a multistage process associated with two characteristic temperatures, the collapse temperature and folding temperature. There were some categories of well-defined local minimum energy. Karplus and coworkers exploited a minimum perturbation conformational search approach in modeling mutant protein structure.[44–47] They mainly focused on single-point mutants and found that when an amino acid was introduced, it might not be able to conform to the more general rules that applied to protein structures of evolutionary origin. While the van der Waals energy largely determined the allowed minima, the relative ranking of the final minima was determined by electrostatic effects. Our results were consistent with theirs in part. For the strategies, our two-step treatment, especially adopted in loop modeling, was also concordant with Karplus's conclusion. However, our simulated annealing methods with minimization could both be applied in protein loop prediction and isolated peptide modeling, and obtained reasonable results, though the adopted potentials were different in the first step. The method mainly focused on conformational prediction and did not study the thermodynamic properties more. It had higher speed than molecular dynamic simulations while keeping enough precision. Contrasted with minimum perturbation conformational search approach, our methods could deal with not only multiple residue positions, but also multiple loop segments.

Because protein loop was often involved in protein active sites, such as binding site and catalysis site, the high predictive accuracy of our methods and restriction of long-range forces of protein environment on loops conformation inspired us that the methods might be introduced into drug design field. Though the isolated peptide may exist in various different conformations, it can only take a few ones as protein loop. Especially for protein ligand, we can use flexible polypeptide such as polyglycine or polyalanine to screen the possible conformational motifs of protein loop backbone, then employ proper side-chain building method and docking method to choose reasonable sequences to design functional protein ligand, which may be used as protein antagonism. For the protease engineering, the similar scheme can be applied to design a new enzyme that will associate with special substrate.

CONCLUSION

From the calculation of eight loop III sequences of short-chain snake venom in a protein environment of erabutoxin b and in an isolated state, and searching conformational space in a protein environment with flexible polyglycine peptide, we clearly demonstrated that the short-range interaction could only provide the possible conformational candidates of the local structures. For the formation and stabilization of the advanced intact structure, long-range interactions are necessary. In general, the long-range interactions play a crucial role in protein folding and protein function.

REFERENCES

1. Popov EM. Quantitative approach to conformations of proteins. Int J Quant Chem 1979;16:707–737.
2. Prat-Gay DG. Conformational preferences of a peptide corresponding to the major antigenic determinant of foot-and-mouth disease virus: implications for peptide-vaccine approaches. Arth Biochem Biophys 1997;341:360–369.
3. Yang JJ, Berg VB, Pitkeathly M, et al. Native-like secondary structure in a peptide from the alpha-domain of hen lysozyme. Folding & Design 1996;1:473–484.
4. Burgess AW, Scheraga HA. Assessment of some problems associated with prediction of the three-dimensional structure of a protein from its amino-acid sequence. Proc Natl Acad Sci USA 1975;72:1221–1225.
5. Karlin A. Molecular properties of nicotinic acetylcholine receptors. In: The cell surface and neuronal function (Cell Surface Reviews, vol. 6). Cotman CW, Paste G, Nicolson GL, editors. New York: Elsevier/North Holland Biomedical Press; 1980. p 191–260.
6. Dufon MJ, Hider RC. Conformational properties of the neurotoxins and cytotoxins isolated from Elapid snake venoms. CRC Crit Rev Biochem 1983;14:113–171.
7. Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acid Res 1996;24: 21–25.
8. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 1991;9:56–68.
9. Tsernoglou D, Petsko GA. The crystal structure of a post-synaptic neurotoxin from sea snake at A resolution. FEBS Lett 1976;68: 1–4.
10. Zinn-Justin S, Roumestand C, Gilquin B, Bontems F, Menez A, Toma F. Three-dimensional solution structure of a curaremimetic toxin from Naja nigricollis venom: a proton NMR and molecular modeling study. Biochemistry 1992;31:11335–11347.
11. Brown LR, Wuthrich K. Nuclear magnetic resonance solution structure of the alpha-neurotoxin from the black mamba (Dendroaspis polylepis polylepis). J Mol Biol 1992;227:1118–1135.
12. Golovanov A, Lomize AL, Arseniev S, Utkin YN, Tsetlin VI. Two-dimensional 1H-NMR study of the spatial structure of neurotoxin II from Naja naja oxiana. Eur J Biochem 1993;213:1213–1223.
13. Roos HM, Van Rooyen PH, Wessels PL. Experimental studies and potential energy calculations of the blocked tetrapeptide Ac-Lys-Pro-Gly-Ile-NMA from the third loop of short-chain snake venom neurotoxins. Int J Peptide Protein Res 1993;42:305–311.
14. Roos HM, Van Rooyen PH, Wessels PL. Experimental studies and potential-energy calculations of the blocked tetrapeptide Ac-Lys-Gln-Gly-Ile-NMA from the third loop of a short-chain snake venom neurotoxin. Int J Peptide Protein Res 1994;43:337–343.
15. Roos HM, Van Rooyen PL. Potential-energy calculations of terminally blocked tetrapeptides from the third loop of short-chain snake venom neurotoxins. Int J Peptide Protein Res 1994;44:562–567.
16. Zhang H, Lai L, Wang L, Han Y, Tang Y. A fast and efficient program for modeling protein loops. Biopolymers 1997;41:61–72.

17. Zhang H, Yang Y, Lai L, Tang Y. Conformational analysis of two glycoproteins: a Monte Carlo simulated annealing approach using a soft-sphere potential. Carbohydr Res 1996;284:25–34.
18. Liu Z, Mao F, Li W, Han Y, Lai L. Calculation of protein surface loops using Monte-Carlo simulated annealing simulation. J Mol Model 2000;6:1–8.
19. Cohen BI, Presnell SR, Cohen FE. Origins of structural diversity within sequentially identical hexapeptides. Protein Sci 1993;2: 2134–2145.
20. Unger R, Sussman JL. The importance of short structural motifs in protein structure analysis. J Comp Aided Mol Des 1993;7:457–472.
21. Han KF, Baker D. Recurring local sequence motifs in proteins. J Mol Biol, 1995; 251: 176-187.
22. Deng QL, Han YZ, Lai LH, Xu XJ. Application of monte-carlo simulated annealing on conformational analysis. Chin Chem Lett 1991;2:809–812.
23. Deng QL, Lai LH, Han YZ, Miao ZW, Ji AX, Xu XJ. Application of simulated annealing: study on relationship between conformation of linear peptides and their cyclization yields. Acta Physico-Chimica Sinica 1994;10:444–448.
24. Nemethy G, Pottle MS, Scheraga HA. Energy parameters in polypeptides .9. Updating of geometrical parameters nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acids. J Phys Chem 1983;87:1883–1887.
25. Vasquez M, Scheraga JA. Use of buildup and energy-minimization procedures to compute low-energy structures of the backbone of enkephalin. Biopolymers 1985;24:1437–1447.
26. Zimmerman SS, Scheraga HA. Influence of local interactions on protein structure III. Conformational energy studies of N-Acetyl-N′-Methylamides of Gly-X and X-Gly dipeptides. Biopolymers 1978;17:1871–1884.
27. Han KF, Baker D. Recurring local sequence motifs in proteins. J Mol Biol 1995;251:176–187.
28. Han KF, Baker D. Global properties of the mapping between local amino acid sequence and local structure in proteins. Proc Natl Acad Sci USA 1996;93:5814–5818.
29. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. J Mol Biol 1997;268:209–225.
30. Chothia C, Novotny J, Bruccoleri R, Karplus M. Domain association in immunoglobulin molecules. The packing of variable domains. J Mol Biol 1985;186:651–663.
31. Chothia C, Lesk AM, Tramontano A, et al. Conformations of immunoglobulin hypervariable regions. Nature 1989;342:877–883.
32. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 1987;196:901–917.
33. Chothia C, Gelfand I, Kister A. Structural determinants in the sequences of immunoglobulin variable domain. J Mol Biol 1998; 278:457–479.
34. Li W, Liu Z, Lai L. Protein loops on structurally similar scaffolds: database and conformational analysis. Biopolymers 1999;49:481–495.
35. Donate LE, Rufino SD, Canard LH, Blundell TL. Conformational analysis and clustering of short and medium size loops connecting regular secondary structures: a database for modeling and prediction. Protein Sci 1996;5:2600–2616.
36. Rufino SD, Donate LE, Canard LH, Blundell TL. Predicting the conformational class of short and medium size loops connecting regular secondary structures: application to comparative modeling. J Mol Biol 1997;267:352–367.
37. Kwasigroch JM, Chomilier J, Mornon JP. A global taxonomy of loops in globular proteins. J Mol Biol 1996;259:855–872.
38. Smith LJ, Mark AE, Dobson CM, van Gunsteren WF. Molecular dynamics simulations of peptide fragments from hen lysozyme: insight into non-native protein conformations. J Mol Biol 1998;280: 703–719.
39. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulations. J Mol Biol 1998;280:925–932.
40. Nazer AP, Torda AE, Bisang C, Weber C, Robinson JA, van Gunsteren WF. Dynamical studies of peptide motifs in the plasmodium falciparum circumsporozoite surface protein by restrained and unrestrained MD simulations. J Mol Biol 1997;267:1012–1025.
41. Mitsutake A, Hansmann UH, Okamoto Y. Temperature dependence of distributions of conformations of a small peptide. J Mol Graph Model 1998;16:226–238, 262–263.
42. Hansmann UH, Okamoto Y, Onuchic JN. The folding funnel landscape for the peptide Met-enkephalin. Proteins 1999;34:472–483.
43. Hansmann UH, Masuya M, Okamoto Y. Characteristic temperatures of folding of a small peptide. Proc Natl Acad Sci USA 1997;94:10652–10656.
44. Prod'hom B, Karplus M. The nature of the ion binding interactions in EF-hand peptide analogs: free energy simulation of ASP to ASN mutations. Protein Eng 1993;6:585–592.
45. Joseph-McCarthy D, Petsko GA, Karplus M. Use of a minimum perturbation approach to predict TIM mutant structures. Protein Eng 1995;8:1103–1115.
46. Shih HH, Brady J, Karplus M. Structure of proteins with single-site mutations: a minimum perturbation approach. Proc Natl Acad Sci USA 1985:82:1697–1700.
47. Bartels C, Stote RH, Karplus M. Characterization of flexible molecules in solution: the RGDW peptide. J Mol Biol 1998;284: 1641–1660.