CHEMISTRY LETTERS, pp. 1275-1278, 1992.

Prediction of Low-Energy Structures of Met-Enkephalin by Monte Carlo Simulated Annealing

Yuko OKAMOTO,* Takeshi KIKUCHI,† and Hikaru KAWAI††
Department of Physics, Nara Women's University, Nara 630
† IRL, CIBA-GEIGY (Japan) Limited, Takarazuka, Hyogo 665
†† Department of Physics, University of Tokyo, Hongo, Tokyo 113

The efficiency of the Monte Carlo simulated annealing method for prediction of the low-energy structures of Met-enkephalin is examined with the exact ECEPP function. Among the 40 obtained conformations, eleven, including the lowest-energy conformation, exhibit structure similar to the tertiary structure predicted by the Monte Carlo with minimization method. The average energy of these conformations are considerably lower than those of the remaining conformations. These findings prove the effectiveness of the method.

Based on the Monte Carlo simulated annealing (SA) method,[1] we have been working on the *ab initio* prediction of tertiary structure of polypeptides, using only amino acid sequence as input. We predicted a few low-energy conformations of Met-enkephalin.[2] Our simulations successfully reproduced empirically known α-helix structure of the C-peptide of ribonuclease A[3,4] and α-helix propensities of amino acid homopolymers.[5] We also obtained a structure similar to the β-sheet in the fragment 16–36 of bovine pancreatic trypsin inhibitor.[6] In order to facilitate the analysis of the obtained conformations, we have introduced a classification of conformations into groups of similar structure[2] and proposed a specific criterion from this classification of conformations as to whether the simulation has reached a successful prediction.[4]

Another promising method for peptide structure prediction is the Monte Carlo with minimization (MCM) method.[7,8] The two methods, SA and MCM, were recently compared and it was claimed that MCM is superior to SA.[9] The comparison was made on Met-enkephalin, and they argued that runs with MCM converge to a unique minimum while those with SA do not. They also argued that the energy difference between the minima reached by the two methods is typically 5–15 kcal/mol in favor of MCM.

In this communication, we re-examine our previous work of SA with Met-enkephalin[2] in order to decide whether the claim in Ref. 9 is reasonable or not. The peptide has the amino acid sequence Tyr-Gly-Gly-Phe-Met. The energy parameters are adopted from ECEPP/2[10] (same as in Ref. 9), and the computer code KONF90[3,4] has been employed. We remark that the energy parameters and molecular geometry adopted in Ref. 2 are an approximation to ECEPP/2 and that the results shown presented there cannot be directly compared with those of Ref. 9. Each SA run consists of 10000 Monte Carlo steps with the initial temperature of 1000 K and the final temperature of 50 K. The tempreature

is lowered exponentially.[3,4] The peptide-bond dihedral angles ω are fixed to 180° for simplicity. We made 40 SA runs, starting from completely random initial conformations. Each run takes about one minute on a HITAC S-820/80 computer.

We label the 40 obtained conformations by numbers in increasing order of energy. The ECEPP energies of conformations #1 (the lowest-energy conformation of the present work), #10, #20, #30, and #40 are −11.9 kcal/mol, −9.5 kcal/mol, −7.6 kcal/mol, −6.8 kcal/mol, and −4.0 kcal/mol, respectively. Since the conformation with the global-minimum energy obtained by MCM [7] (Conformaton #0 hereafter) has $E = -12.4$ kcal/mol, conformation #1 does not differ much from #0 in energy. We remark that the energy of #0 (−12.4 kcal/mol) was calculated with KONF90[3,4] from the dihedral angles given in Ref. 9, while the energy value given in Ref. 9 for the same set of dihedral angles is −12.9 kcal/mol. Hence, one should note that the energy functions used in the present work and Ref. 9 have uncertainties of $\approx$ 0.5 kcal/mol.

The lowest-energy conformations of SA (#1) and MCM (#0) not only have similar energy values but also have similar structures. The structures of the 40 obtained conformations as well as #0 are compared in terms of the root-mean-square (rms) distances. Among the 40 conformations, #1 has the second smallest rms deviation from #0 (0.8 Å for the backbone structure and 2.3 Å for the structure of the entire molecule).

Table 1. Dihedral angles and ECEPP energy of the minimum energy conformations for each cluster [a]

Conformation	*A*0	*A*1	*B*	*D*	*C*
E (kcal/mol)	−12.4	−11.9	−10.0	−9.0	−7.8
ϕ_1	−86	98	111	109	62
ψ_1	156	154	157	151	144
ω_1	−177	180	180	180	180
ϕ_2	−154	−161	−71	−164	−67
ψ_2	84	69	78	79	104
ω_2	169	180	180	180	180
ϕ_3	84	65	159	81	130
ψ_3	−74	−93	−37	−68	−27
ω_3	−170	180	180	180	180
ϕ_4	−137	−85	−154	−87	−111
ψ_4	19	−27	151	150	−46
ω_4	−174	180	180	180	180
ϕ_5	−164	−83	−140	−87	−162
ψ_5	160	142	−29	130	123
ω_5	180	180	180	180	180
χ_1^1	−173	−179	−179	172	170
χ_1^2	79	−112	−95	74	−102
χ_1^3	−166	149	169	−24	170
χ_4^1	59	180	59	−174	175
χ_4^2	−86	73	87	−114	71
χ_5^1	53	−65	−68	−67	−173
χ_5^2	175	180	177	−178	180
χ_5^3	180	179	−179	175	179
χ_5^4	−59	−55	60	−63	177

a) *A*0, *A*1, *B*, *D*, and *C* respectively stand for conformations #0, #1, #8, #14, and #19.

In order to analyse the results in more detail, we have carried out classification of the obtained conformations into clusters of similar structure.[2] A set of conformations belong to the same cluster of conformations if their rms distances are less than a cutoff c (see Ref. 4 for the exact definition of the cluster used in the present work). In Table 1 we list the dihedral angles and ECEPP energy of the minimum-energy conformations (#1 for *A*, #8 for *B*,

Table 2. Classification of conformations into clusters of similar structure[a)]

	Backbone				Entire molecule		
c	Cluster	n	$\bar{E}$	c	Cluster	n	$\bar{E}$
0.1	$A_1 = \{1,2,3,4,5,6\}$	6	−11.5	0.4	$\tilde{A}_1 = \{1,2\}$	2	−11.8
0.2	$A_2 = A_1 \cup \{7,9\}$	8	−11.1		$\tilde{A}'_1 = \{3,4,6\}$	3	−11.3
	$B_1 = \{8,11\}$	2	−9.7	0.6	$\tilde{A}_1$	2	−11.8
	$C_1 = \{19,23\}$	2	−7.6		$\tilde{A}'_1$	3	−11.3
0.6	$A_3 = A_2 \cup \{0,15\}$	10	−11.0		$\tilde{A}''_1 = \{5,9\}$	2	−10.5
	$B_2 = B_1 \cup \{12\}$	3	−9.6	1.2	$\tilde{A}_2 = \tilde{A}_1 \cup \tilde{A}'_1 \cup \tilde{A}''_1$	7	−11.2
	$C_2 = C_1 \cup \{21,31\}$	4	−7.4		$\tilde{B}_1 = \{10,36\}$	2	−7.9
	$D_1 = \{14,16\}$	2	−8.8		$\tilde{C}_1 = \{19,23\}$	2	−7.6
	$B'_2 = \{28,29\}$	2	−6.8	2.0	$\tilde{A}_3 = \tilde{A}_2 \cup \{0,15\}$	9	−11.1
	$E_1 = \{30,32\}$	2	−6.8		$\tilde{B}_2 = \tilde{B}_1 \cup \{13\}$	3	−8.3
	$E'_1 = \{33,37\}$	2	−6.3		$\tilde{B}'_2 = \{8,11,12,20,27\}$	5	−8.6
1.0	A_3	10	−11.0		$\tilde{C}_2 = \tilde{C}_1 \cup \{21,31\}$	4	−7.4
	$B_3 = B_2 \cup B'_2 \cup \{10,18,26,36,39\}$	10	−7.8		$\tilde{D}_1 = \{14,16,22\}$	3	−8.3
	C_2	4	−7.4		$\tilde{D}'_1 = \{18,28\}$	2	−7.5
	$D_2 = D_1 \cup \{22,24,25,34\}$	6	−7.7		$\tilde{D}''_1 = \{24,25,34\}$	3	−7.0
	$E_2 = E_1 \cup E'_1$	4	−6.5		$\tilde{D}'''_1 = \{30,32,33\}$	3	−6.7
	$F_1 = \{13,20,27\}$	3	−7.9	2.4	$\tilde{A}_4 = \tilde{A}_3 \cup \{7,38\}$	11	−10.5
1.4	$A_4 = A_3 \cup \{17,38\}$	12	−10.4		$\tilde{B}_3 = \tilde{B}_2 \cup \tilde{B}'_2 \cup \{26,39\}$	10	−7.9
	$B_4 = B_3 \cup D_2 \cup E_2 \cup F_1$	23	−7.5		$\tilde{C}_2$	4	−7.4
	C_2	4	−7.4		$\tilde{D}_2 = \tilde{D}_1 \cup \tilde{D}'_1 \cup \tilde{D}''_1 \cup \tilde{D}'''_1 \cup \{29\}$	12	−7.3

a) The cutoff c is in Å . n and $\bar{E}$ (kcal/mol) are respectively the number of conformations and the average energy of each cluster.

#19 for C, and #14 for D) of the four clusters we identified and those of #0 (labeled as $A0$ in the table). The details of the cluster analysis are summarized in Table 2. The number of conformations (n) and the average energy ($\bar{E}$) for each cluster are also displayed. One of the most remarkable features in these results is the fact that the six lowest-energy conformations, #1—#6, belong to the same cluster with the cutoff c as small as 0.1 Å, as far as the backbone structure is concerned (see A_1 in Table 2). The average energy of cluster A_1 is −11.5 kcal/mol, and this value is close to the energy of conformation #0, −12.4 kcal/mol, which disagrees strongly with the claim in Ref.9 that the minimum energy reached by MCM is about 5 kcal/mol lower than that of SA. Though the side-chain structure is not as rigidly determined as that of the backbone, the above low-energy conformations do have similar side-chain structure. As is shown in Table 2, three clusters, $\tilde{A}_1$, $\tilde{A}'_1$, and $\tilde{A}''_1$, of the lowest-energy conformations (#1—#6 and #9) which are defined for $c = 0.4$ Å and 0.6 Å merge into one big cluster $\tilde{A}_2$ at $c = 1.2$ Å.

Conformation #0 belongs to cluster A of Table 2. When only the backbone structure is taken

into account, it joins cluster A at $c = 0.6$ Å (see A_3 in Table 2). We believe that this slight structural deviation of #0 from #1—#7 and #9 is due to our simplification that we fix the peptide-bond dihedral angles ω to 180°, while #0 obtained in Ref. 7 has generally $\omega \neq 180°$. Since the average energy of cluster A is considerably lower than those of other clusters(for example, see $\bar{E}$ for A_4, B_4, and C_2 in Table 2), we have essentially the same prediction of Met-enkephalin structure by SA as that by MCM as the energy global minimum.

Moreover, close examination of Table 2 indicates that there are roughly three characteristic backbone structures of Met-enkephalin (see clusters A_4, B_4, and C_2). When we consider the structure of the entire molecule, however, we find that cluster B is divided into two (clusters $\tilde{B}_3$ and $\tilde{D}_2$); B has two possible side-chain orientations.

To conclude, we have explored the low-energy structures of Met-enkephalin by SA and classified the obtained conformations into clusters of similar structures.[2,4] We have shown that one cluster has average energy significantly lower than any other cluster and that this cluster indeed includes the global -minimum energy conformation obtained by MCM. These findings prove the effectiveness of the SA method. Furthermore, MCM requires much more computation time than SA (the ratio of four hours to one minute according to Ref. 9). This means that for a prediction of tertiary structure of large peptide and protein, SA will be a more feasible method. We remark that we have predicted a tertiary structure of PTH(1 - 34) which agrees with the implications of NMR data.[11]

We are grateful to the Computer Center, Institute for Molecular Science, Okazaki for computation time on HITAC S-820/80.

References

1) For example, S. Kirkpatrick, C.D. Jr. Gelatt, and M.P. Vecchi, *Science*, **220**, 671 (1983).
2) H. Kawai, T. Kikuchi, and Y. Okamoto, *Protein Engineering*, **3**, 85 (1989).
3) H. Kawai, Y. Okamoto, M. Fukugita, T. Nakazawa, and T. Kikuchi, *Chem. Lett.*, **1991**, 213.
4) Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai, *Protein Engineering*, **4**, 639 (1991).
5) M. Fukugita, T. Nakazawa, H. Kawai, and Y. Okamoto, *Chem. Lett.*, **1991**, 1279.
6) T. Nakazawa, H. Kawai, Y. Okamoto, and M. Fukugita, submitted to *Protein Engineering*.
7) Z. Li and H.A. Scheraga, *J. Mol. Struct. (Theochem)*, **179**, 333 (1988).
8) For another method, see B. von Freyberg and W. Braun, *J. Comp. Chem.*, **12**, 1065 (1991).
9) A. Nayeem, J. Vila, and H.A. Scheraga, *J. Comp. Chem.*, **12**, 594 (1991).
10) M.J. Sippl, G. Némethy, and H.A. Scheraga, *J. Phys. Chem.*, **88**, 6231 (1984), and references therein.
11) Y. Okamoto, T. Kikuchi, T. Nakazawa, and H. Kawai, to be published.

(Received March 12, 1992)