Supramolecular Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713649759

Prediction of the global energy minimum conformation of polypeptides by the High Directional Monte Carlo procedure

Mu Shik Jhon[a]; Jae Kwang Shin[a]
[a] Center for Molecular Science and Department of Chemistry, Korea Advanced Institute of Science and Technology, Taejon, Korea

To cite this Article Jhon, Mu Shik and Shin, Jae Kwang(1993) 'Prediction of the global energy minimum conformation of polypeptides by the High Directional Monte Carlo procedure', Supramolecular Chemistry, 1: 2, 139 — 143
To link to this Article: DOI: 10.1080/10610279308040659
URL: http://dx.doi.org/10.1080/10610279308040659

SUPRAMOLECULAR CHEMISTRY, Vol. 1, pp. 139–143
Reprints available directly from the publisher
Photocopying permitted by license only

Prediction of the global energy minimum conformation of polypeptides by the High Directional Monte Carlo procedure

MU SHIK JHON* and JAE KWANG SHIN

Center for Molecular Science and Department of Chemistry, Korea Advanced Institute of Science and Technology, 373-1 Kusung Dong, Yusung Gu, Taejon, 305-701 Korea

A new Monte Carlo sampling scheme, namely High Directional Monte Carlo procedure, is used to obtain the global energy minimum conformations of polypeptides such as enkephalin and melittin. The resultant structures of enkephalin and melittin agree well with previous results of theoretical and experimental studies. Particularly, it is shown that some important parts in the conformation, such as the hinge region that principally determines the tertiary structure of proteins, are correctly described by the new method. The resultant structures are compared with those of other works and their stereoscopic views are shown.

INTRODUCTION

Since the folding of biomolecules occurs in a spontaneous pathway,[1] it is natural to think that their native structures are simply energy minimum ones reached by thermal structural fluctuations. The theoretical prediction of their native structures, therefore, presents the same problems as the optimization of the conformational energy functions of them. Usually, the solution of this type of optimization problem can be achieved by Monte Carlo simulations. However, since the number of local energy minimum structures increases as $\sim 10^n$, where n is the number of amino acid residues, the global minimization of large bimolecular structures by the conventional Monte Carlo algorithm quickly becomes impossible as the number of residues increases. One of the principal difficulties of the original Metropolis procedure in the simulation of large molecules such as proteins and polypeptides seems to be inefficient random sampling. Hence, it is important to explore the possibility of devising a more efficient sampling scheme.[2-5]

Recently, we have developed the High Directional Monte Carlo procedure (HDMC) that predicts the shape of an energy hyper surface to generate variable steps in Monte Carlo simulation. Results show that the procedure is very much improved compared with other methods in several respects.[2,3] First, the search of the statistically important regions of a conformational space is much faster than with the conventional procedures. Second, the radial distribution functions (rdfs) and energy distribution functions show that there is no bias in the equilibrium distribution of ensembles. In addition, this method can be applied to most of the systems that are only partially accessible by other methods.

In the HDMC procedure, the co-variance tensor is calculated to control the individual trial step distribution of the next conformational region from each second of the actual sampling segment in the previous set which is composed of an appropriately chosen number of steps.

Thermal heating processes compensate for the possible lowering of transition probabilities between the local minima which result from the rigid geometry constraints[2,6] and the equilibrium description of the non-equilibrium phenomena. In other words, there may be some isolated classes in Markov chains that inhibit the transitions between them. If this occurred the total Markov chains would be reducible and not ergodic.[7]* Therefore, some additional thermal heating is necessary to overcome this possibility.

In this paper we describe the use of the HDMC procedure to obtain the global energy minimum conformations of enkephalin, and some preliminary results from the simulation of melittin are reported. Enkephalin is a peptide hormone with morphine-like activity in mammalian brains and melittin is a toxic

* To whom correspondence should be addressed.

* A state x_i is called ergodic if it is aperiodic and persistent with a finite mean recurrence time.

protein found in the venom of honey bees and has 26 amino acid residues. It is found that the resulting structures of HDMC simulations explain well the general structural characteristics of both systems within a reasonable range of computing time.

METHOD

In Monte Carlo simulation, a trial conformation $\mathbf{r}'$ is generated by changing the current conformation $\mathbf{r}$ randomly and it is accepted with the probability of p or rejected with $1-p$ where[9,10]

$$p = \min\left[1, \frac{T(\mathbf{r}|\mathbf{r}')}{T(\mathbf{r}'|\mathbf{r})}\exp[-\beta(\Delta E)]\right] \tag{1}$$

and $T(\mathbf{r}'|\mathbf{r})$ is the joint probability distribution function which is usually called a sampling function. The choice of this sampling function $T(\mathbf{r}'|\mathbf{r})$ is of particular importance and distinguishes the different methods. Next, we briefly discuss the important points included in HDMC sampling function $T_{\text{HDMC}}(\mathbf{r}'|\mathbf{r})$.

The principal concern of HDMC sampling scheme is to sample the random conformations whose distribution resembles the real phase space distribution of conformations. Since the shape of any distribution function can be approximately (neglecting higher order correlation terms) measured by its co-variance, we tried to develop a new method which allows random sampling configurations with a co-variance approaching that of the real system distribution.[2,3]

If a system resides at a certain local equilibrium state during Monte Carlo simulations, then the configurational variables $\mathbf{r}$ can be assumed to be normally distributed according to the probability distribution function $f_N(\mathbf{r})$

$$f_N(\mathbf{r}) = [(2\pi)^{N/2}\det\Sigma]^{-1}\exp(-\tfrac{1}{2}\det[(\mathbf{r}-\bar{\mathbf{r}})^T\Sigma^{-1}(\mathbf{r}-\bar{\mathbf{r}})]) \tag{2}$$

where Σ is an $n \times n$ co-variance matrix such that $\Sigma = \|\sigma_{ij}\|$ and each element σ_{ij} can be expressed as $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, where ρ_{ij} is the correlation coefficient.[3]

From eqn. (2), we can derive the HDMC sampling function $T_{\text{HDMC}}(\mathbf{r}'|\mathbf{r})$ which is written as[3]

$$T_{\text{HDMC}}(\mathbf{r}'|\mathbf{r}) = D(\Sigma')\exp(-\tfrac{1}{2}\det[(\mathbf{r}'-\mathbf{r})^T\Sigma'^{-1}(\mathbf{r}'-\mathbf{r})]) \tag{3}$$

where

$$D(\Sigma') = \frac{1}{(2\pi)^{N/2}\det\Sigma'}$$

$$\Sigma' = \|\sigma'_{ij}\|$$

$$\sigma'_{ij} = \langle(r'_i - r_i)(r'_j - r_j)\rangle \tag{4}$$

The co-variance matrix Σ as an estimation of the shape of the next conformational region can be calculated from the past history of sample distributions of some nearby conformational space.

Next, we describe the implementation procedure in detail.[2,3]

Step 1

The total steps M_{tot} were divided by M, the number of steps per set, to give M_{tot}/M sets. M should be chosen so that the information on the different local energy minima do not mix together and lose statistical significance. M values can be obtained by monitoring the convergence of conformational fluctuations and it is found that $M = 20$–50 MCS (Monte Carlo Steps per particle or torsion) is adequate.

Step 2

The information on the shape of the energy hyper surface was gathered by calculating the co-variance σ_{ij} during Monte Carlo simulations of M steps, and from the co-variance, the second moment of increments σ'_{ij} was calculated by using[3]

$$\sigma'^{k}_{ij} = 2\sigma_{ij} = \frac{2}{M}\left[\sum_{v=k-M}^{k-1}[r_i^v - \bar{r}_i][r_j^v - \bar{r}_j]\right] \tag{5}$$

where k denotes the number of sets. Eqn. (5) was successfully applied to the 12-6 Lennard-Jones and TIP4P water system.[3] However, since the conformational energy hyper surface of large biomolecules changes its geometry rather quickly during the simulations, the direction of the co-variance hyper ellipsoid along which the largest element of the matrix lies, should be adjusted to get the correct direction of movement. It can be achieved by a rather simple generalization of eqn. (5) by introducing the linear conjugation of co-variances between the subsequent sets by using

$$\sigma'^{k+1}_{ij} = 2[\sigma^k_{ij} + (1-\lambda)s^k_{ij}]$$
$$s^k_{ij} = \sigma^k_{ij} - \sigma^{k-1}_{ij} \tag{6}$$

where the parameter λ was chosen to be between 0 and 1. In our applications, we used two extremes of $\lambda = 1$ (enkephalin and melittin)[2] and $\lambda = 0$ (Lennard-Jones and TIP water system)[3].

Step 3

The co-variance matrix was divided into the lower and upper triangular matrices of

$$\Sigma' = C^{\mathrm{T}} \cdot C \tag{7}$$

where the elements c_{ij} of the matrix C can be calculated by the recursive formula of the square root method such that

$$c_{ij} = \frac{\sigma'_{ij} - \sum_{k=1}^{j-1} c_{ik}c_{jk}}{(\sigma'_{ij} - \sum_{k=1}^{j-1} c_{jk}^2)^{1/2}} \quad (8)$$

where

$$\sum_{k=1}^{0} c_{ik}c_{jk} = 0, \quad 1 \leqslant j \leqslant i \leqslant n \quad (9)$$

Since the co-variance matrices are symmetric and positive definite, they are uniquely split into those two parts in eqn. (7).

Step 4

The Monte Carlo simulation for the next set was carried out by generating the new configurations

$$r'_i = r_i + \sum_j c_{ij}n_j \quad (10)$$

where $\{n_1, n_2, \ldots, n_j, \ldots\}$ is the random normal vector with mean zero and unit standard deviation.

Step 5

During the simulations the co-variance was calculated using eqn. (6).

Step 6

If the energy fluctuation η_k at the k^{th} set

$$\eta^k = \frac{\langle E \rangle^k - E^k_{\min}}{\langle E \rangle^k} \quad (11)$$

was greater than the predefined value (10^{-3} kcal/mol for melittin and enkephalin), we went back to Step 3. Otherwise, the simulation was halted.

In our work, in order to compensate for the possible lowering of the transition probabilities between the local minima and to make the molecule more flexible so that it moved in the conformational space more efficiently, we applied the temperature annealing from 500 to 1000 K to room temperature using 2–3% annealing rates. Then the temperature of the $(k+1)^{\text{th}}$ set is given by[2]

$$T^{k+1} = T^k - {}_{\chi T}T^k \quad (12)$$

where ${}_{\chi T}$ is the temperature annealing factor. Finally we used Consistent Valance Force Field (CVFF) and AMBER potential functions and geometries to obtain the conformational energies of the molecules.[11,12]

RESULTS AND DISCUSSIONS

Met-enkephalin

Met-enkephalin is a pentapeptide hormone in the mammalian brain with morphine-like activities.[13]

H—Tyr^1—Gly^2—Gly^3—Phe^4—Met^5—OH

Among the five residues, the principal torsional angles (ϕ, ψ) of Gly^3 and Phe^4 are known to be principally responsible for the overall structure of Met-enkephalin. So the emphasis is placed on the torsional angles of these two residues. Since there are many suggestions of conformational models of Met-enkephalin based on experimental and theoretical studies,[13,14] the reliability of the method can be tested by comparing the results with those of other works.

The stereo views of the resultant structures are shown in Figure 1 and the torsion angles are given in Table 1, where the results of HDMC simulation are compared with those of other theoretical studies. As can be seen from Table 1 and Figure 1, the HDMC structure is consistent with those of other reports which agree that the global energy minimum structure of Met-enkephalin is a II′β-bend type of structure.[13,14] It is clear that the final conformation is stabilized by two hydrogen bonds between the hydrogen on the phenyl ring in Tyr^1 and the oxygens on Gly^3 and Phe^4.

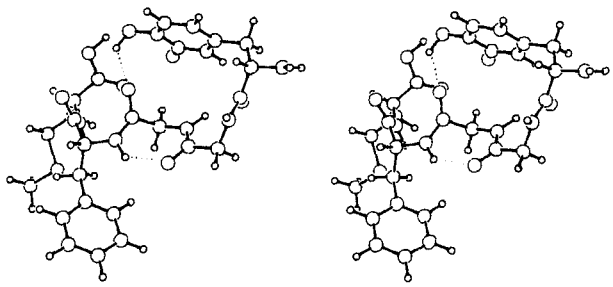

Figure 1 Stereo view of the global energy minimum structure of Met-enkephalin obtained by HDMC simulation.

Table 1 Global energy minimum structure of Met-enkephalin

	Torsional angles (degrees)					
	HDMC			*Previous work*[14]		
Residue	ϕ	ψ	ω	ϕ	ψ	ω
Tyr^1	−61	178	−178	−86	156	−177
Gly^2	−143	84	174	−154	83	169
Gly^3	53	−88	−174	84	−74	−170
Phe^4	−12	86	−176	−13	19	−174
Met^5	−167	−52	−177	−164	160	−180

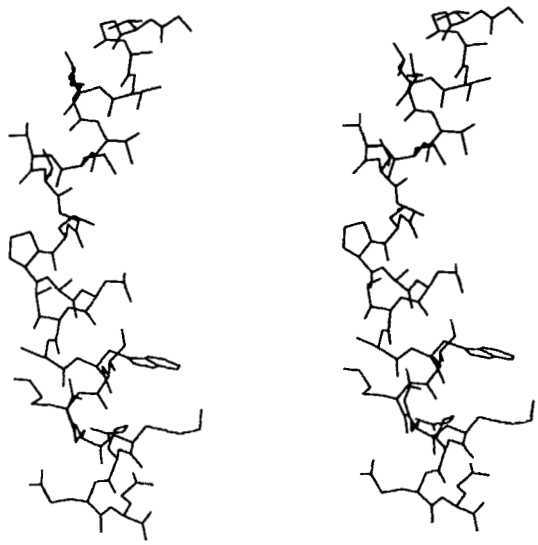

Figure 2 Stereo view of the crystal melittin found by X-ray crystal diffraction; this shows just one asymmetric unit of the tetramer.

In those simulations, it takes approximately 10^5 to 5×10^5 steps to reach the equilibrium and additional time to calculate and divide the co-variance is less than 5% of total CPU time. HDMC is, therefore, comparable to the MMC (Metropolis Monte Carlo) in computing time. Also the average conformational change per step $[\Sigma_i(\Delta\theta_i^2)]^{1/2}$, as an indicator of the sampling efficiency, showed that HDMC was more efficient by at least 20 times than MMC.[2] Other details of HDMC simulations can be found in the reference 2.

Melittin

Melittin is the principal component of the venom from honey bees with 26 amino acid residues and six positive charges:[15]

$$H_3N^+GIGAVL^+KVLTTGLPALISWIK^+R^+K^+R^+QQCONH_2$$

where each of the characters represents the one-letter abbreviations of amino acids.[16]

The molecule is known to be unordered or have traces of helix in water without salt at neutral pH and at low concentration, but at high protein concentration and/or high ionic strength, it adopts a largely helical conformation and aggregates as a tetramer.[15,17,18] In our study, the Melittin conformation at high ionic strength was simulated in its deprotonated form.[17,18]

The stereo view of the tertiary structure of Melittin found by X-ray diffraction studies is shown in Figure 2.[19] Clearly, the structure is composed of two helices jointed by a 'hinge'-like structure between the 11th Tyr and the 12th Gly residues.[20] The side chains are distributed such that the polar side chains point to the convex part of the molecule while the non-polar side chains point to the concave part.

In all simulations, the acceptance ratios are maintained at 0.4–0.6. The structure in Figure 3 is the result of HDMC simulations starting from the conformations initially set by the conventional α-helical form with the principal torsion angles (ϕ, ψ) of $(-60°, -60°)$.

There are clearly two α-helical structures joined by a hinge at residues 11 and 12 (except for some deviations from X-ray structure). The principal torsional angles of three regions (two α-helical regions and one hinge structure) in total amino acids (residues 2–4, 10–12, and 23–25) are shown in Table 2. We also carried out simulations starting from the different conformations with (ϕ, ψ) pairs of $(-40°, -40°)$ and $(-50°, -50°)$, and found that the resultant structures are very similar to the one shown in Figure 3.

The angle at the hinge of the 11th and 12th residues, between two α-helices, is somewhat smaller than that found from the X-ray structure. It seems that these differences result from the absence of intermolecular interactions. Actually, the angles must be relaxed to some extent to form a tetramer in the crystals being X-rayed.[19,20]

We can easily find the polar side chains such as Lys-7, Thr-11 and Ser-18 and non-polar side chains such as Leu-9, Leu-13, and Ile-20, that lie at both sides of the curvature of the molecules. Since there are no external or intermolecular interactions, the HDMC structures should be less compact than those of the tetramer in the crystal.

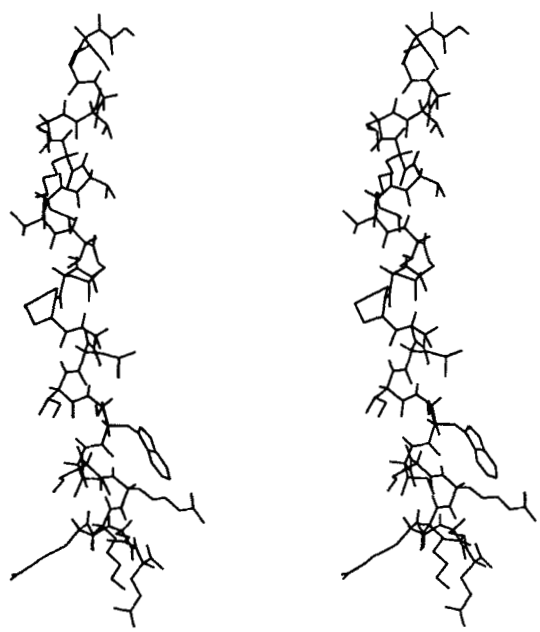

Figure 3 HDMC structure of melittin without charges starting from a typical α-helical structure.

Table 2 Principal torsional angles of nine residues in three regions from the HDMC simulations and crystal structure of melittin*

Residues	*Ile*^{2}	*Gly*^{3}	*Ala*^{4}	*Thr*^{10}	*Thr*^{11}	*Gly*^{12}	*Lys*^{23}	*Arg*^{24}	*Gln*^{25}
ϕ_{HDMC}	−46.24	−59.53	−69.48	−50.38	−99.19	106.27	−53.13	−75.85	−68.25
$\phi_{X\text{-ray}}$	−63.14	−67.07	−69.48	−46.87	−94.63	109.39	−58.35	−70.12	−67.78
ϕ_{HDMC}	−33.58	−23.56	−58.72	−56.21	142.58	−61.96	−29.80	−41.69	−51.90
$\phi_{X\text{-ray}}$	−42.10	−30.04	−51.51	−47.69	139.43	−61.23	−35.42	−38.36	−43.14

* All angles are in degrees.

Significantly, α-helical structures shown in Figure 3 are maintained during the simulation. This means that the strain due to the packing of the positive charges in positively charged melittin is reduced in the deprotonated form.

We expect the monomeric melittin to show helical conformations at high ionic strength. Many experimental studies have reported the existence of helical structures of melittin at high pH (above 10), independently of their concentration. It seems that some difference between the conformations in the experimental studies and those of the HDMC simulations result from ignoring the environments.

CONCLUSION

In this paper, we used the HDMC procedure to obtain the global energy minimum structure of Met-enkephalin and melittin. Except for some deviations in the detailed conformations between the final structures produced by the HDMC simulations and those produced from the X-rays or those of previous reports,[15,17-20] the overall results are found to be satisfactory in explaining the structural characteristics of the two molecules. However, if we are to simulate the more complex folding mechanisms of the larger proteins rather than the somewhat simple cases chosen, it seems that the HDMC simulation algorithm must be incorporated with simplification of the potential energy functions.[2] In addition, there should be more importance placed on the solvent environment.

The improved features of the HDMC procedure can be summarized as follows: First, the method allows the molecule to move efficiently in the conformational space. Second, the time taken to gather information is much smaller than for the other methods which calculate the forces at every step. Third, HDMC is applicable to all systems, ranging from simple classical particles or quantum ones to large biomolecules. In addition, there is no bias in the equilibrium distribution of HDMC ensembles.

Although there are some inadequacies, such as the parametric relevancy of M, the HDMC procedure is found to be successful in locating the tertiary conformations of small proteins from the extended or secondary structures. Further studies to refine the final structure and to successfully calculate the secondary structures from the primary amino acid sequences are required to confirm the more practical capabilities of the HDMC procedure.

(Received July 28, 1992)

REFERENCES

1 Anfinsen, C.B.; *Science* **1973**, *181*, 223–230.
2 Shin, J.K.; Jhon, M.S.; *Biopolymers* **1991**, *31*, 177–185.
3 Shin, J.K.; Yu, J.Y.; Jhon, M.S.; *J. Chem. Phys.*, **1992**, in press.
4 Rossky, P.J.; Doll, J.D.; Friedman, H.L.; *J. Chem. Phys.* **1978**, *69*, 4628–4633.
5 Cao, J.; Berne, B.J.; *J. Chem. Phys.* **1990**, *92*, 1980–1985.
6 Fixman, M.; *J. Chem. Phys.* **1978**, *69*, 1527–1545.
7 Iosifescu, M.; *Finite Markov Processes and Their Applications*, **1980**, John Wiley & Sons, New York.
8 Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E.; *J. Chem. Phys.* **1953**, *21*, 1087–1092.
9 Binder, K.; *Monte Carlo Methods in Statistical Physics*, 2nd edn., (Binder, K., ed.), Topics in Current Physics, **1986**, Vol. 7, Springer, Berlin/Heidelberg.
10 Heerman, D.W.; *Computer Simulation Methods in Theoretical Physics*, **1986**, Chap. 4, Springer-Verlag, New York.
11 Hagler, A.T.; *J. Am. Chem. Soc.* **1979**, *101*
12 Weiner, S.; Kollma, P.A.; Nguyen, D.T.; Case, D.; *J. Comp. Chem.* **1986**, *7*, 230–252.
13 Premilat, S.; Maigret, B.; *J. Phys. Chem.* **1980**, *84*, 293–299.
14 Li, Z.; Scheraga, H.A.; *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6611–6615.
15 Kubota, S.; Yang, J.T.; *Biopolymers* **1986**, *25*, 1493–1504.
16 Stryer, L.; *Biochemistry*, **1981**, 2nd edn., W.H. Freeman, San Francisco.
17 Levin, I.R.; Lavialle, F.; Mollay, C.; *Biophys. J.* **1982**, *37*, 339–349.
18 Degrado, W.F.; Musso, G.F.; Lieber, M.; Kaiser, E.T.; Kezdy, F.J.; *Biophys. J.*, **1982**, *37*, 329–338.
19 Terwiliger, T.C.; Weissman, L.; Eisenberg, D.; *Biophys. J.* **1982**, *37*, 353–361.
20 Bazzo, R.; Tappin, M.J.; Pastore, A.; Harvey, T.S.; Carver, J.A.; Campbell, I.D.; *Eur. J. Biochem.* **1988**, *173*, 139–146.