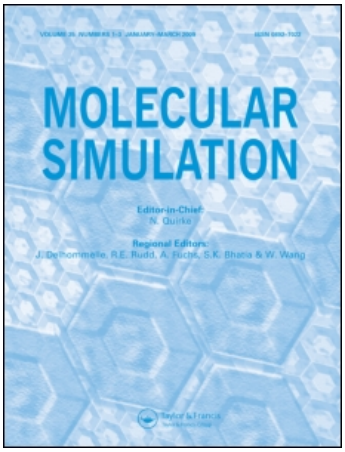


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Valley Restrained Monte Carlo Procedure as a Method to Improve Sampling Efficiency

Seung Hoon Choi[a]; Mu Shik Jhon[a]
[a] Department of Chemistry and Center for Molecular Science, Korea Advanced Institute of Science and Technology, Taejon, Korea

To cite this Article Choi, Seung Hoon and Jhon, Mu Shik(1999) 'Valley Restrained Monte Carlo Procedure as a Method to Improve Sampling Efficiency', Molecular Simulation, 22: 2, 123 — 147
To link to this Article: DOI: 10.1080/08927029908022091
URL: http://dx.doi.org/10.1080/08927029908022091

Molecular Simulation, 1999, Vol. 22, pp. 123–147
Reprints available directly from the publisher
Photocopying permitted by license only

VALLEY RESTRAINED MONTE CARLO PROCEDURE AS A METHOD TO IMPROVE SAMPLING EFFICIENCY

SEUNG HOON CHOI and MU SHIK JHON*

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

(Received March 1998; accepted April 1998)

A straightforward application of Metropolis Monte Carlo method to a protein system has proven to be inefficient owing to the serious anisotropy of the conformational energy surface. We propose the Valley Restrained Monte Carlo procedure, that predicts the topology of the energy hyper-surface using statistical and empirical data, as a method to improve the sampling efficiency. It calculates the Valley Function which goes along the valley between local minima in the energy surface and then reinforces the sampling of the region near the Valley Function in Monte Carlo Procedure. Valley Restrained Monte Carlo procedure samples the minima and the path along the lowest energy barrier between local minima more frequently, it reduces trapping in local minima and increases the convergence rate. This method is successfully applied to a model energy surface, the blocked alanine dipeptide (Ac-Ala-NHMe) and the pentapeptide Met-enkeplain (H-Tyr-Gly-Gly-Phe-Met-OH). The comparison between Valley Restrained Monte Carlo Procedure and the conventional Metropolis Monte Carlo Method shows that the sampling efficiency of our new method is greater than that of conventional Metropolis Monte Carlo. It is expected that this increase in the efficiency will be large when the system is larger.

Keywords: Simulation; proteins; Monte Carlo; sampling; valley

I. INTRODUCTION

In the late 1950s Anfinsen and his colleagues made a remarkable discovery [1]. They were exploring a long-standing puzzle in biology. What causes newly made protein to wind into specifically shaped forms able to perform crucial tasks in the living cell? In the process the team found the answer was

*Corresponding author. e-mail: jwater@sorak.kaist.ac.kr

simpler than anyone had imagined. It was the sequence of amino acids, referred to as the primary structure of the protein, that determines the native conformation, the structure is stable under physiological conditions and has a biological activity.

This successful experiment of refolding the ribonuclease by Anfinsen *et al.*, motivated efforts to predict the tertiary structure of protein theoretically from the sequence of its amino acids. It is assumed that the native conformation of a protein is that which corresponds to the global minimum of a potential energy function. This global optimization of protein and polypeptide structure brings its own formidable problems due to the complexity of energy hyper-surface, *i.e.*, the multiple minima problem. The typical number of local minima in the potential energy surface varies as 10^n, where n is the number of residues of polypeptides. Thus the direct searching of the global energy minimum structure in the whole conformational space quickly becomes impossible as the number of residues increases.

In order to get over this multiple minima problem, many theoretical methods for computing the stable conformations of proteins have been developed [2–8]. These previous works can be classified into two categories; the improvement of methodology of computation and the simplification of the interaction potential function. But there is no one method which can be regarded satisfactory.

Since the Monte Carlo method simulates the thermal Markovian processes, it should be applicable to the problem of obtaining the global energy minimum structure of a protein. However, a straightforward application of conventional Metropolis method to a protein system has proven to be very inefficient [9–11]. Many previous workers tried to improve the efficiency of conventional Metropolis Monte Carlo Method and to over- come the energy barrier between local minima [12–16]. One of the main difficulties is the isotropic sampling of Metropolis Monte Carlo pro- cedure which does not consider the anisotropic potential energy surface of proteins. If we properly take into account this anisotropic energy surface in the progression of the simulation, we may hope that the Monte Carlo method could become a powerful method to simulate the globular structural changes of proteins.

In the series of previous articles, we have proposed the Valley Restrained Monte Carlo Procedure (VRMC) as a method to get over the trapping in local minima and increase the energy convergence rate, using the preferential sampling of the important region in the configurational space. As the chemical reaction is coordinated by the path which cross the lowest activation barrier from the reactant to the product, the conformational change in simulation might occur from one local minimum to other local

minimum *via* the valley between two local minima because that path is the most efficiency way, which avoids high energies. In this present work, we consider the anisotropy of the force field in the process of simulation using the information from *a priori* knowledge of the statistical and/or experimental data.

The Valley Function (VF) which passes through the local minima and goes along the valley between local minima on the energy hyper-surface, is formulated from data acquired in advance. The VF is the most probable way of crossing the barrier in the simulation, because it is the curve which connects two minima by way of the lowest energy barrier.

In Valley Restrained Monte Carlo Procedure, the sampling of the region near the Valley Function is reinforced by an appropriate algorithm the so called Preselection Process, which is similar to the Metropolis test and will be mentioned in the next section. The Preselection Process lets the molecule cross the energy barrier more easily with the help of the preferential sampling of the specific direction that has the lower obstacle. The main aims of this work are as follows. (1) To develop a new Monte Carlo procedure for the molecular systems with connectivity (chemical bond). Since the connectivity creates the high anisotropic potential energy surface, this anisotropy should be taken into account in the simulation. (2) To accelerate the convergence rate and improve the sampling efficiency by the preferential sampling of the more important regions in configurational space. (3) To overcome the energy barrier between local minima on energy hyper-surface by increasing the sampling for a specific direction.

This new method, Valley Restrained Monte Carlo Method (VRMC), is applied to a model energy surface, the alanine dipeptide(Ac-Ala-NHMe) and the pentapeptide Met-enkephalin (H-Tyr-Gly-Gly-Phe-Met-OH). From the comparisons of the results of our VRMC with those of conventional Metropolis Monte Carlo (MMC) Simulation, we find that VRMC is more powerful than MMC in these systems.

II. METHODOLOGY

Our new algorithm is applied to two classes of systems according to the method for the acquisition of the information about the potential energy surface; one is the model energy surface (the known energy surface) and the other is the model peptide (the information about the energy surface from the statistical/empirical data). We show the derivation of Valley Restrained Monte Carlo procedure in three stages for the sake of convenience of

explanation. At first, we illustrate the formulation of the Valley Function (VF) which includes information about the energy hyper-surface. Because each system has a somewhat different formulation of the VF, this part is divided into two subsections (model surface and peptide). Then, we represent the modification of Metropolis Monte Carlo method. Finally, the validity of the VF for each amino acid is investigated before the comparison with Monte Carlo results.

1. Valley Function

1.1. Model Energy Surface

An analytic equation is employed for the model energy surface as represented in Eq. 1 and Figure 1(a),

$$f(x,y) = x^2 + y^2 + A\cos(4\pi x) \quad (x \in [-2,2],\ y \in [-2,2]) \qquad (1)$$

where the value A, the amplitude of cosine part, controls the height of energy barriers between local minima. This surface has the periodic

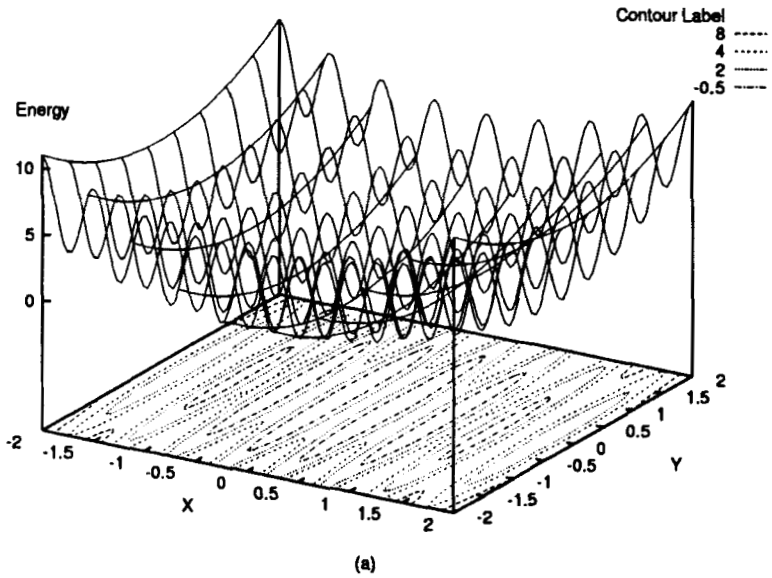


FIGURE 1 Energy map of Eq. 1, where A is equal to 3. (a) 3-dimensional energy surface and contour map and (b) 2-dimensional plot at $y = 0$.

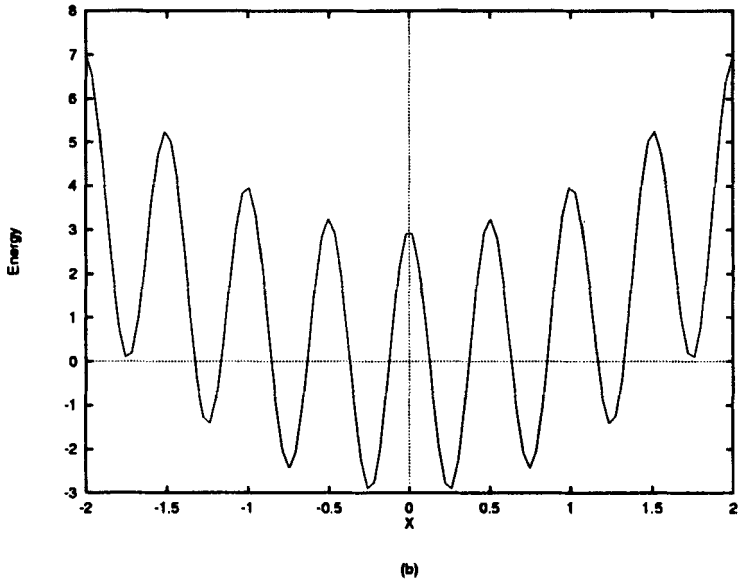


FIGURE 1 (Continued).

boundary condition [17]. As shown in Figure 1(b), the surface has two symmetric global energy minima near $(-1/4, 0)$ and $(1/4, 0)$. Both the local minima and global minima are located on the x-axis. The VF is the curve passing through the local minima and going along the valley between local minima on energy hyper-surface. In this case, we can easily find the VF from the topology of the energy surface. The VF of the model surface coincides with the x-axis ($y = 0$).

1.2. Peptide

In the case of protein, 20 VF's exist because different amino acids have different anisotropy in the energy surface.

In order to get information on this anisotropy, we propose two possible schemes. One is the energy contour map of ϕ and ψ of a single residue (Ramachandran plot) and the other is the X-ray crystallographic data of the Protein Data Bank (PDB). Because the anisotropic energy surface from one residue in protein is somewhat different from that of the isolated single amino acid, the torsional distribution of an isolated single residue is not similar to that of a residue in a protein. This difference stems from the short, medium and long range interactions between amino acids. There is a report

in the literature that indicates that the local structure at the minimum in the energy surface of ϕ and ψ of the isolated single residue has Van der Waals' clashes in a protein owing to the interaction between residues [18]. The aim of the Valley Restrained Monte Carlo Method is not the application to the single residue but that to the protein. Thus, we adopt the second scheme in this work.

However, there is a little doubt in using the known X-ray crystallographic data directly, because of the biased information by the interactions between neighboring residues and the redundancy of PDB data. The first problem is that the residue prefers a certain configurational space according to the residue preceding and following it in the protein. Supposing the data is sufficiently abundant, the bias by the neighboring residue would be negligible because the bias might be eliminated by statistical weighting [19]. In the case of second problem, if we properly remove the redundancy, the problem disappears. The following subsection describes the removal process.

Selection of Representative Set of Proteins Brookhaven Protein Data Bank contains three dimensional structures of over 700 proteins. However, the database is highly redundant, containing several entries for identical or very similar sequences. Reliable structural and statistical analyses of three dimensional protein structures should be based on unbiased data. There have been several reports of work to organize the confused data [20–22]. Thus we have no need to systematize this disordered data over again. We take one of the previous methods which is most adequate for our analysis [21], in which a technique for clustering Brookhaven PDB structures based on the sequences and the contents of α- and β-structures was developed. A representative sample of sequences was then obtained by grouping similar sequences together using a similarity significance threshold and selecting a typical representative from each group. An unbiased sample of 101 high resolution structure, representing a wide variety of proteins, was chosen based on the previous suggestions. We excluded the structure solved by NMR and homology.

Table I shows the selected proteins of which resolution is equal to or higher than 3.0 Å The average length of sequences is 226. In order to remove the bias of the end group effect, we prepare the data of each amino acid without the terminal amino acids in protein. Though there is some difference in the number of data for each amino acid according to the natural abundance, the average number of data for one amino acid is 1136 as shown in Table II. This is sufficient to manipulate statistically.

TABLE I Selected protein set for finding the valley functions

Entry[a]	*Protein*	*# of a.a*[b]	*Entry*	*Protein*	*# of a.a.*
1acx	Actionxanthin	106	21bp	Leucine binding protein	344
1ccr	Cytochrome c	110	21h2	Leghemoglobin	151
1cla	Chloramphenicol acetyltransferase	211	21tn	Lectin	365
			2mhr	Myohemerythrin	98
1cse	Subtilisin Carlsberg-eglin C complex	333	2ovo	Ovomucoid third domain	54
			2pab	Prealbumin	112
1ctf	50S ribosomal protein	66	2paz	Pseudoazurin	121
1ctx	X cobratoxin	69	2pfk	Phosphofructokinase	299
1eca	Erythrocruorin	134	2rnt	Ribonuclease T1	102
1etu	Tu-factor (domain I)	175	2rsp	Protease	113
1fc2	Immunoglobulin Fc-fragment	247	2sga	Proteinase A	179
1gcr	γ-II-crystallin	172	2sns	Nuclease	139
1gd1	Holo-glyceraldehyde-3-phosphate dehydrogenase	332	2sod	Superoxide dismutase	150
			2ssi	Subtilisin inhibitor	105
1gox	Glycolate oxidase	349	2taa	α-amylase	476
1gp1	Glutathione peroxidase	182	2tsl	Tyrosyl tRNA synthetase	315
1hip	High potential iron protein	83	2wrp	Tryptophan repressor	102
1hoe	α-amylase inhibitor	72	3adk	Adenylate kinase	193
1il1b	Interleukin 1B	149	3app	Acid proteinase	321
1ldm	Lactate dehydrogenase	327	3b5c	Cytochrome b5	84
1lrd	λ-repressor	90	3bcl	Bacteriochlorophyll A	342
1lz1	Lysozyme	128	3fxc	Ferredoxin	96
1mbd	Myoglobin	151	3fxn	Flavodoxin	136
1nxb	Neurotoxin B	60	3gap	Catabolite gene activator protein	206
1pcy	Plastocyanin	97			
1phh	p-hydroxybenzoate-hydroxylase complex	392	3grs	Gluthathione reductase	459
			3hmg	Heamagglutinin	499
1prc	Photosynthetic reaction center	1179	3ins	Insulin	44
1pyp	Inorganic pyrphosphatase	281	3lzm	Lysozyme	162
1r69	*N*-terminal domain of 434 repressor	61	3pgk	Phosphoglycerate kinase	414
			3pgm	Phosphoglycerate mutase	228
1hrd	Rhodanese	291	3rns	Ribonuclease A	122
1sgt	Trypsin	221	3tln	Thermolysin	314
1sn3	Neurotoxin	63	451c	Cytochrome c551	80
1tnf	Tumor necrosis factor	150	4bp2	Phospholipase A2	115
1ubq	Ubiquitin	74	4cpv	Ca-binding parvalbumin	107
1utg	Uteroglobin	68	4fd1	Ferredoxin	104
1wsy	Tryptophan synthase	630	4hhb	Hemoglobin	283
256b	Cytochrome b562	104	4pep	Pepsin	324
2aat	Asparatate aminotransferase complex	394	5cha	α-chymotryspin	230
			5cpa	Carboxypeptidase A	305
2aza	Azurin	127	5hvp	HIV-1 proteinase	97
2cab	Carbonic anhydrase B	254	5pti	Trypsin inhibitor	52
2ccy	Cytochrome c'	125	5rxn	Rubredoxin	49
2cdv	Cytochrome c3	105	5tim	Triose phosphate isomerase	247
2ci2	Chymotrypsin inhibitor	63	6acn	Aconitase	749
2cpp	Cytochrome P-450 Cam	403	6xia	D-xylose isomerase	385
2cts	Citrate synthase	435	7api	α1-antitrypsin	373
2cyp	Cytochrome c peroxidase	291	8adh	Alcohol dehydrogenase	372
2fb4	IgG1 Fab	441	8atc	Aspartate carbamyltransferase	452
2fxb	Ferredoxin	79	8cat	Catalase	496
2gbp	D-galactose binding protein	307	8dfr	Dihydrofolate reductase	184
2gn5	Gene-5 binding protein	85	9pap	Papain	210
2hla	HLA-AW68	365	9wga	Agglutinin	168

[a] PDB abbreviation; [b] Number of amino acid in that protein.

TABLE II Number of each amino acid used in least square fitting

amino acid	*frequency*	*amino acid*	*frequency*
alanine	1989	leucine	1876
arginine	933	lysine	1361
asparagine	1008	methionine	411
aspartic acid	1320	phenylalanine	876
cysteine	416	proline	1077
glutamine	846	serine	1522
glutamic acid	1301	threonine	1360
glycine	1916	tryptophan	333
histidine	538	tyrosine	815
isoleucine	1185	valine	1633
total	22716	average	1136

Formulation of Valley Function Figure 2 shows the Ramachandran plots of alanine and glycine from the unbiased 101 protein data set. In order to make a VF which optimally passes through the points in Figure 2, we formulate a model function as a third order polynomial of ϕ and ψ including with their cross terms. Eq. 2 described the Valley Function.

$$\begin{aligned} \mathrm{VF}(\phi,\psi) = a_0 + a_1\phi + a_2\psi + a_3\phi^2 + a_4\phi\psi + a_5\psi^2 \\ + a_6\phi^3 + a_7\phi^2\psi + a_8\phi\psi^2 + a_9\psi^3 \end{aligned} \tag{2}$$

In order to determine the coefficients in Eq. 2 by using least square methods and the data in Table II, we set up the Eqs. 3 and 4.

$$Error(a_0,\ldots,a_9) = \sum_{i=1}^{n}\{\mathrm{VF}(\phi_i,\psi_j)\}^2 \tag{3}$$

$$\left[\frac{\partial\{Error(a_0,\ldots,a_9)\}}{\partial a_j}\right]_{j=0,\ldots,9} = 0 \tag{4}$$

We need a modification of Eq. 2 to Eq. 5 because Eq. 4 has infinite number of solutions.

$$\begin{aligned} \mathrm{VF}(\phi,\psi) = 1 + b_1\phi + b_2\psi + b_3\phi^2 + b_4\phi\psi + b_5\psi^2 \\ + b_6\phi^3 + b_7\phi^2\psi + b_8\phi\psi^2 + b_9\psi^3 \end{aligned} \tag{5}$$

Table III shows the coefficients $(b_1,\ldots,b_9)$ of twenty amino acids obtained by the least square method and Figure 3 displays the Valley Function of alanine and glycine obtained using Eq. 5 and Table III. For an arbitrary torsional set (ϕ_0,ψ_0) of a specific amino acid, the value of

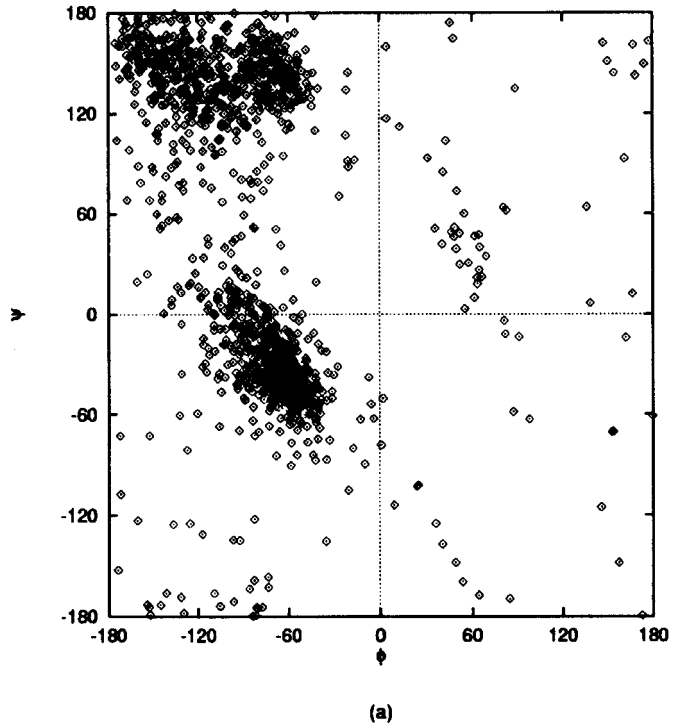


FIGURE 2 Ramachandran plots of torsional data for (a) alanine and (b) glycine. This data is acquired from Protein Data Bank in Table I.

$VF(\phi_0, \psi_0)$ is determined using Eq. 5. The value is zero for the point on VF curve, while the absolute value increases as the point is far from the curve.

2. Modification of Monte Carlo Algorithm

We describe the VRMC algorithm for the case of protein only. The algorithm for the case of model function system is similar.

In order to reinforce the sampling of the configurational space near the VF curve we introduce the Preselection Surface as represented in Eq. 6 and Figure 4.

$$PS(\phi, \psi) = \exp\{-\gamma(1 + b_1\phi + b_2\psi + b_3\phi^2 + b_4\phi\psi + b_5\psi^2 + b_6\phi^3 + b_7\phi^2\psi + b_8\phi\psi^2 + b_9\psi^3)^2\} \quad (6)$$

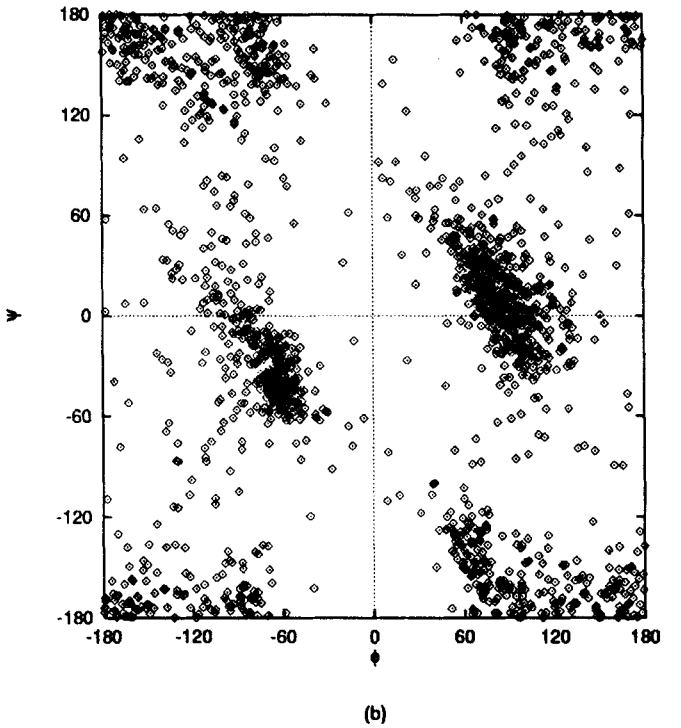


FIGURE 2 (Continued).

If we select an arbitrary torsional set (ϕ_0, ψ_0), a value is determined from Eq. 6. We call it a Preselection Factor. If the chosen torsional set (ϕ_0, ψ_0) is on the VF curve in Figure 3, the Preselection Factor (PS (ϕ_0, ψ_0)) is 1 because VF (ϕ_0, ψ_0) is equal to zero, whereas the Preselection Factor decreases and finally converges to zero, as the torsional set deviates from the VF.

As formulated in Eq. 6, we introduce a parameter, γ, – called a steepness parameter – which controls the smoothness of the Preselection Surface and finally the relative strength of reinforcement for the sampling of specific direction. The steepness parameter (γ) will be described in the last part of this section.

It is well known that Scheme 1 with the exception of Step 3 is the conventional Metropolis Monte Carlo (MMC) Procedure. In the process of MMC, more than 90% of total CPU time is consumed in Step 4 (Metropolis test including the calculation of the interaction potential function) of Scheme 1. We decided to reinforce the sampling of the more acessible region

TABLE III The coefficients in Eq. 5 by least square fitting

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
Ala	1.18×10^{-2}	3.43×10^{-3}	-2.96×10^{-5}	-1.36×10^{-6}	-2.84×10^{-5}	-3.52×10^{-7}	-5.71×10^{-9}	-2.26×10^{-7}	-1.15×10^{-7}
Arg	1.08×10^{-2}	2.36×10^{-3}	-4.30×10^{-5}	-8.74×10^{-6}	-2.55×10^{-5}	-3.96×10^{-7}	1.19×10^{-9}	-1.84×10^{-7}	-9.49×10^{-8}
Asn	6.02×10^{-3}	-7.93×10^{-5}	-9.03×10^{-5}	-2.16×10^{-5}	-2.31×10^{-5}	-4.84×10^{-7}	-2.64×10^{-8}	-1.87×10^{-7}	-6.04×10^{-8}
Asp	9.37×10^{-3}	1.15×10^{-3}	-4.65×10^{-5}	-7.71×10^{-6}	-2.57×10^{-5}	-3.38×10^{-7}	2.81×10^{-8}	-2.13×10^{-7}	-7.51×10^{-8}
Cys	1.08×10^{-2}	6.45×10^{-4}	-4.05×10^{-5}	-3.05×10^{-5}	-2.96×10^{-5}	-4.05×10^{-7}	-1.67×10^{-7}	-2.46×10^{-7}	-5.21×10^{-8}
Gln	1.01×10^{-2}	2.08×10^{-3}	-3.61×10^{-5}	-1.73×10^{-5}	-4.28×10^{-5}	-2.64×10^{-7}	-5.99×10^{-8}	-3.83×10^{-7}	-9.94×10^{-8}
Glu	1.16×10^{-2}	3.17×10^{-3}	-3.48×10^{-5}	-4.72×10^{-7}	-2.85×10^{-5}	-4.20×10^{-7}	2.61×10^{-9}	-1.91×10^{-7}	-8.90×10^{-8}
Gly	-1.31×10^{-3}	3.07×10^{-3}	-5.12×10^{-5}	-2.77×10^{-6}	-1.24×10^{-5}	-2.77×10^{-8}	-1.64×10^{-8}	6.43×10^{-8}	-1.06×10^{-7}
His	7.54×10^{-3}	1.87×10^{-3}	-6.11×10^{-5}	1.23×10^{-5}	-2.58×10^{-5}	-3.23×10^{-7}	2.18×10^{-7}	-1.93×10^{-7}	-1.00×10^{-7}
Ile	1.19×10^{-2}	4.98×10^{-3}	-2.00×10^{-5}	2.79×10^{-5}	-5.06×10^{-5}	-2.99×10^{-7}	5.19×10^{-8}	-4.40×10^{-7}	-8.22×10^{-8}
Leu	1.08×10^{-2}	3.62×10^{-3}	-4.07×10^{-5}	-7.44×10^{-6}	-5.05×10^{-5}	-3.80×10^{-7}	-1.42×10^{-7}	-4.59×10^{-7}	-8.88×10^{-8}
Lys	1.02×10^{-2}	1.80×10^{-3}	-4.27×10^{-5}	-1.56×10^{-5}	-3.02×10^{-5}	-3.69×10^{-7}	-4.94×10^{-8}	-2.34×10^{-7}	-8.06×10^{-8}
Met	7.52×10^{-3}	1.30×10^{-3}	-1.35×10^{-4}	-7.16×10^{-5}	-4.23×10^{-5}	-9.41×10^{-7}	-5.44×10^{-7}	-3.91×10^{-7}	-8.49×10^{-8}
Phe	1.19×10^{-2}	3.45×10^{-3}	-2.39×10^{-5}	9.21×10^{-6}	-3.02×10^{-5}	-3.60×10^{-7}	3.04×10^{-10}	-2.13×10^{-7}	-5.64×10^{-8}
Pro	1.30×10^{-2}	1.62×10^{-3}	-9.84×10^{-5}	3.18×10^{-5}	-3.68×10^{-5}	-9.76×10^{-7}	4.42×10^{-7}	-5.13×10^{-7}	-5.21×10^{-8}
Ser	1.09×10^{-2}	2.70×10^{-3}	-3.26×10^{-5}	4.46×10^{-6}	-2.64×10^{-5}	-3.28×10^{-7}	2.87×10^{-8}	-2.27×10^{-7}	-8.15×10^{-8}
Thr	1.14×10^{-2}	2.85×10^{-3}	-3.08×10^{-5}	-2.24×10^{-6}	-3.59×10^{-5}	-3.68×10^{-7}	-7.03×10^{-8}	-3.12×10^{-7}	-6.87×10^{-8}
Trp	9.86×10^{-3}	1.52×10^{-3}	-5.84×10^{-5}	-6.85×10^{-5}	-5.69×10^{-5}	-4.62×10^{-7}	-4.32×10^{-7}	-4.39×10^{-7}	-7.62×10^{-8}
Tyr	9.33×10^{-3}	1.71×10^{-3}	-7.31×10^{-5}	3.05×10^{-5}	-3.24×10^{-5}	-5.70×10^{-7}	-2.18×10^{-7}	-2.71×10^{-7}	-6.75×10^{-8}
Val	1.25×10^{-2}	3.65×10^{-3}	-2.25×10^{-5}	9.19×10^{-6}	-3.56×10^{-5}	-4.00×10^{-7}	-4.66×10^{-8}	-2.84×10^{-7}	-5.52×10^{-8}

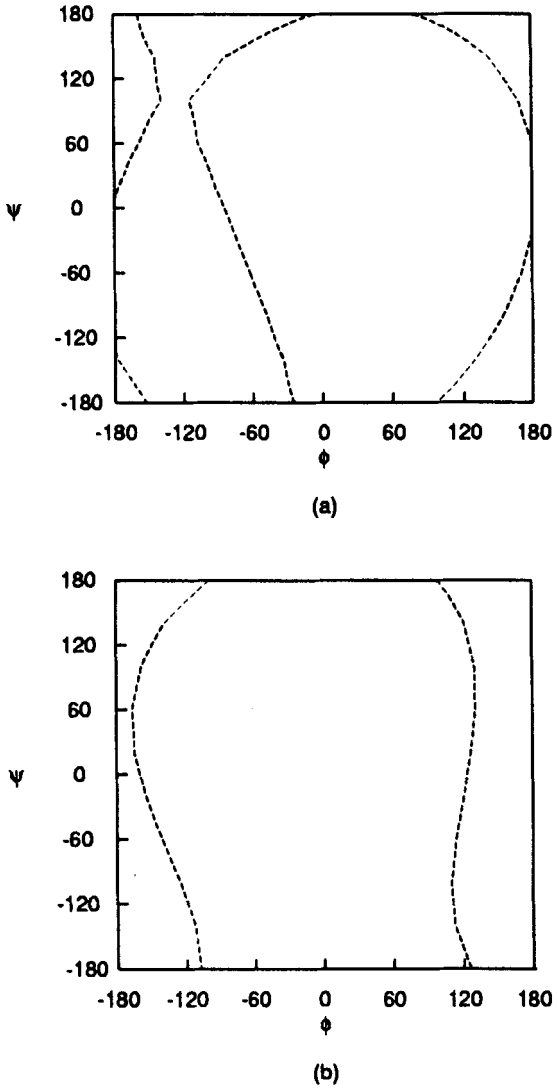


FIGURE 3 Valley Functions for (a) alanine and (b) glycine. The curves are drawn from Eq. 2 and Table III.

and the direction which passes through the lowest energy barrier by *a priori* selection taking into account the anisotropy of potential energy surface before the Metropolis test.

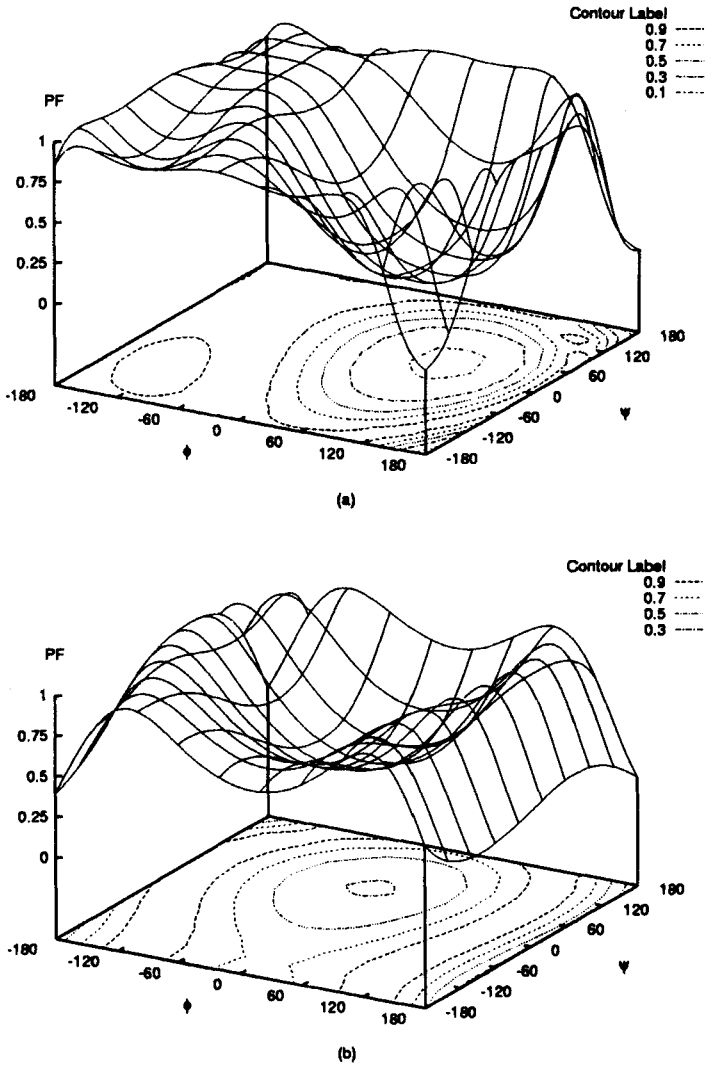
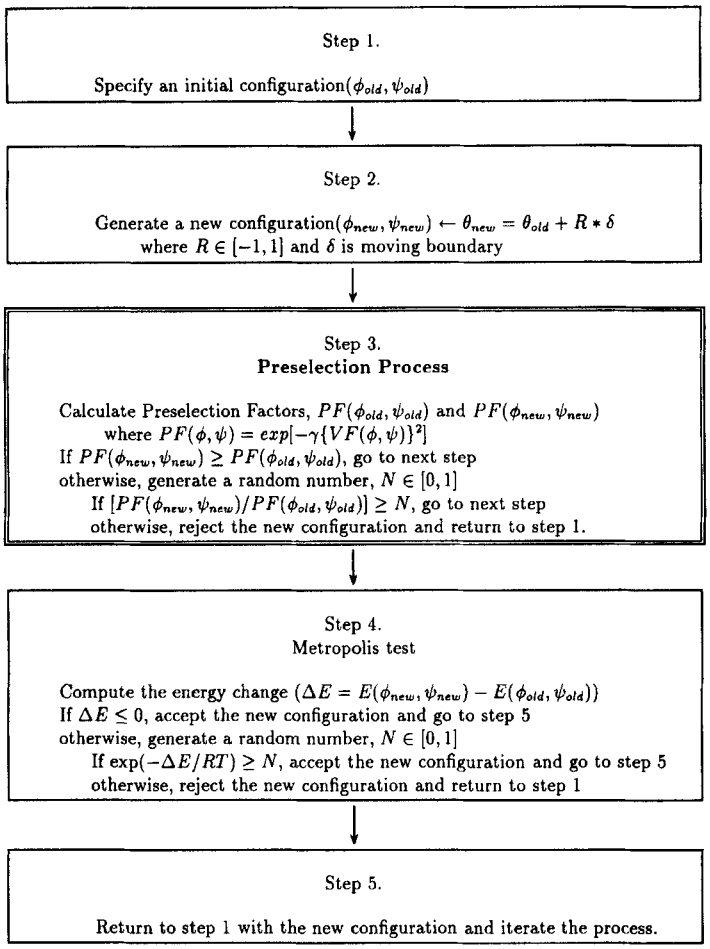


FIGURE 4 Preselection Surfaces for (a) alanine and (b) glycine, where γ is equal to 1. The surfaces are drawn from Eq. 6 and Table III.

Scheme 1 shows the algorithm of the Valley Restrained Monte Carlo Procedure. The difference between VRMC from MMC is the insertion of Step 3 of VRMC – the preselection process.

In the preselection process, we compare the Preselection Factor of an old configuration with that of a new configuration, and then we carry out *a*



SCHEME 1 Algorithm of Valley Restrained Monte Carlo Procedure.

priori selection using this comparison. Explaining in detail, if the Preselection Factor of a new configuration is greater than or equal to that of an old configuration, we sample the new configuration. Otherwise, the new configuration is accepted or rejected by Boltzmann factor determined by two Preselection Factors, which is represented in Eq. 7, likewise Metropolis test.

$$\text{Boltzmann Factor for Preselection} = \frac{\mathrm{PF}(\phi_{\mathrm{new}}, \psi_{\mathrm{new}})}{\mathrm{PF}(\phi_{\mathrm{old}}, \psi_{\mathrm{old}})} \tag{7}$$

If the ratio ($PF(\phi_{new}, \psi_{new})/PF(\phi_{old}, \psi_{old})$) is greater than or equal to random number ($N \in [0, 1]$), Metropolis test is performed. Otherwise the selected configuration is rejected and new sampling is carried out.

In this process, the sampling in the region near energy minima and valley between minima is reinforced. Thus we might expect the high convergence rate and easy transverse of energy barrier between minima by way of efficient path.

As already mentioned, we introduce a steepness parameter (γ) in the formation of the Preselection Surface. It determines the amount of constraint in the direction of sampling for the region near Valley Function by the modulation of the relative strength of Preselection Factor. The higher γ, the bigger difference between $PF(\phi_{new}, \psi_{new})$ and $PF(\phi_{old}, \psi_{old})$, that is to say, the Boltzmann Factor for Preselection is decreased, and the stronger becomes the reinforcement of the sampling the region near VF. We adjust the optimum steepness parameter for the system in the process of simulation.

3. The Validity of Valley Function and Preselection Surface

Here we would like to demonstrate the validity of our Valley Function and Preselection Surface by a comparison between amino acids. It is difficult to investigate the similarity or dissimilarity of Valley Functions between two amino acids directly, owing to the variety of coefficients $(b_1, \ldots, b_9)$ as represented in Table III. Thus we obtain the rank correlation coefficient between mutual amino acids, as formulated in Eq. 8, [23]

$$r_s = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \tag{8}$$

where x and y are the rank orders of the coefficients of Valley Functions of two amino acid, respectively. The rank correlation coefficients in Table IV enable us to evaluate the similarity between VF's qualitatively.

Naturally occurring amino acids can be classified into several systematic groups by their functional abilities and polarities (aliphatic, aromatic, non-polar, polar and charged). However there is no distinct differentiation between VF's according to this usual classification, while the VF's of glycine and as-paragine are clearly different from those of others. This difference is easily explained by the special properties of glycine and asparagine. It is strange that VF of proline is not remarkably different from others.

TABLE IV The rank correlation coefficients between two amino acids

	Ala	*Arg*	*Asn*	*Asp*	*Cys*	*Gln*	*Glu*	*Gly*	*His*	*Ile*	*Leu*	*Lys*	*Met*	*Phe*	*Pro*	*Ser*	*Thr*	*Trp*	*Tyr*	*Val*
Ala	–	1.0000	0.6500	1.0000	0.9667	0.9667	1.0000	0.3167	0.8333	0.8000	0.9500	1.0000	0.9333	0.8167	0.8333	0.8333	0.9500	0.9333	0.9833	0.8362
Arg	1.0000	–	0.6500	1.0000	0.9667	0.9667	1.0000	0.3167	0.8333	0.8000	0.9500	1.0000	0.9333	0.8167	0.8333	0.8333	0.9500	0.9333	0.9833	0.8362
Asn	0.6500	0.6500	–	0.6500	0.6167	0.6000	0.6500	−0.0333	0.4833	0.4333	0.5833	0.6500	0.5833	0.4500	0.4833	0.4833	0.5833	0.5667	0.6333	0.5247
Asp	1.0000	1.0000	0.6500	–	0.9667	0.9667	1.0000	0.3167	0.8333	0.8000	0.9500	1.0000	0.9333	0.8167	0.8333	0.8333	0.9500	0.9333	0.9833	0.8362
Cys	0.9667	0.9667	0.6167	0.9667	–	0.9167	0.9667	0.2667	0.7333	0.6833	0.9333	0.9667	0.9833	0.7000	0.7333	0.7333	0.9333	0.9833	0.9833	0.7542
Gln	0.9667	0.9667	0.6000	0.9667	0.9167	–	0.9667	0.2667	0.8000	0.8333	0.9833	0.9667	0.8667	0.8167	0.8000	0.8000	0.9833	0.9000	0.9500	0.8198
Glu	1.0000	1.0000	0.6500	1.0000	0.9667	0.9667	–	0.3167	0.8333	0.8000	0.9500	1.0000	0.9333	0.8167	0.8333	0.8333	0.9500	0.9333	0.9833	0.8362
Gly	0.3167	0.3167	−0.0333	0.3167	0.2667	0.2667	0.3167	–	0.1500	0.1000	0.2333	0.3167	0.2833	0.1333	0.1500	0.1500	0.2333	0.2333	0.2833	0.1804
His	0.8333	0.8333	0.4833	0.8333	0.7333	0.8000	0.8333	0.1500	–	0.9667	0.7833	0.8333	0.7000	0.9833	1.0000	1.0000	0.7833	0.6333	0.8167	0.9509
Ile	0.8000	0.8000	0.4333	0.8000	0.6833	0.8333	0.8000	0.1000	0.9667	–	0.8167	0.8000	0.6333	0.9833	0.9667	0.9667	0.8167	0.6000	0.7833	0.9346
Leu	0.9500	0.9500	0.5833	0.9500	0.9333	0.9833	0.9500	0.2333	0.7833	0.8167	–	0.9500	0.9000	0.8000	0.7833	0.7833	1.0000	0.9167	0.9667	0.8362
Lys	1.0000	1.0000	0.6500	1.0000	0.9667	0.9667	1.0000	0.3167	0.8333	0.8000	0.9500	–	0.9333	0.8167	0.8333	0.8333	0.9500	0.9333	0.9833	0.8362
Met	0.9333	0.9333	0.5833	0.9333	0.9833	0.8667	0.9333	0.2833	0.7000	0.6333	0.9000	0.9333	–	0.6667	0.7000	0.7000	0.9000	0.9667	0.9667	0.7378
Phe	0.8167	0.8167	0.4500	0.8167	0.7000	0.8167	0.8167	0.1333	0.9833	0.9833	0.8000	0.8167	0.6667	–	0.9833	0.9833	0.8000	0.6167	0.8000	0.9673
Pro	0.8333	0.8333	0.4833	0.8333	0.7333	0.8000	0.8333	0.1500	1.0000	0.9667	0.7833	0.8333	0.7000	0.9833	–	1.0000	0.7833	0.6333	0.8167	0.9509
Ser	0.8333	0.8333	0.4833	0.8333	0.7333	0.8000	0.8333	0.1500	1.0000	0.9667	0.7833	0.8333	0.7000	0.9833	1.0000	–	0.7833	0.6333	0.8167	0.9509
Thr	0.9500	0.9500	0.5833	0.9500	0.9333	0.9833	0.9500	0.2333	0.7833	0.8167	1.0000	0.9500	0.9000	0.8000	0.7833	0.7833	–	0.9167	0.9667	0.8362
Trp	0.9333	0.9333	0.5667	0.9333	0.9833	0.9000	0.9333	0.2333	0.6333	0.6000	0.9167	0.9333	0.9667	0.6167	0.6333	0.6333	0.9167	–	0.9500	0.6722
Tyr	0.9833	0.9833	0.6333	0.9833	0.9833	0.9500	0.9833	0.2833	0.8167	0.7833	0.9667	0.9833	0.9667	0.8000	0.8167	0.8167	0.9667	0.9500	–	0.8526
Val	0.8362	0.8362	0.5247	0.8362	0.7542	0.8198	0.8362	0.1804	0.9509	0.9346	0.8362	0.8362	0.7378	0.9673	0.9509	0.9509	0.8362	0.6722	0.8526	–

The Preselection Factor (PF) modified from the VF is not simply determined by the distance from the VF as shown in Figure 3. The points with the same distance from VF in the Ramachandran plot have different PF according to the kind of amino aicd, as shown in Figures 4(a) and 4(b). The difference stems from the relative statistical weighting. As represented in Figure 2(b), glycine has many torsional data in the region having positive ϕ due to its lack of a side chain, while as represented in Figure 2(a), alanine has a few data points in this region. In the case of the VF, roughly drawn simple curves are located on this region for both case, as shown in Figures 3(a) and 3(b). But the Preselection Surface (PS) of both amino acids are distinctly different. From the comparison Figure 4(a) with Figure 4(b), we can see this clearly. The PS of alanine is steep while that of glycine is very smooth, in spite of the same steepness parameter ($\gamma = 1$).

This difference results from the number of data point in that region. The number of data in specific regions plays an important role in the steepness in that region. In the case of nonlinear least square fitting, because the region having the large number of data gives the possibility of accumulation of error in Eq. 3, in order to diminish the error the interpolating curve becomes zigzag with large amplitude in comparison with the region having the small number of data [24]. From this reason, a region with many statistical data (the region having positive ϕ) in Figure 2(b), has a relative smooth surface and high Preselection Factor (PF), while the same region in Figure 2(a) has a steep surface and a low PF due to there being few statistical data (as shown in Fig. 4).

III. RESULTS AND DISCUSSIONS

To compare the Valley Restrained Monte Carlo (VRMC) simulation with the conventional Metropolis Monte Carlo (MMC) simulation, we carry out MMC and VRMC for two kinds of systems. At first, in order to investigate the validity of VRMC clearly, the global optimization of a simple analytic function is performed. Then we consider an application to real systems, chain molecules having the anisotropic energy hyper-surface due to its own chemical bonds. The chain molecules which are investigated in this work, are the alanine dipeptide (Ac-Ala-NHMe) as the model peptide and the pentapeptide Met-enkephalin (H-Tyr-Gly-Gly-Phe-Met-OH) which is an endogenous peptide from the mammalian brain with morphine like activities [25].

1. Global Optimization of a Model Energy Surface

We employ a model energy surface including periodic boundary condition, as represented in Eq. 1 and Figure 1. In spite of the simpleness of the surface, it is adequate for the investigation of global optimization due to its several minima. In the process of simulation, we investigate the convergence rate, acceptance ratio and the crossing number, where the crossing number is the number of traverses between two symmetric global energy minima (GEM) during the full simulation. Because the high value crossings mean the high possibility of crossing the barrier, it is used as the criterion for the ability to escape from the trapping in local energy minimum.

Table V shows the comparison between MMC and VRMC under various conditions, where 'amp' is the amplitude of cosine part in Eq. 1, 'dx' is the maximum moving in sampling of Monte Carlo process, 'RT' is virtual temperature used in Metropolis test, 'cross #' is crossing number and 'step # at GEM' means the number of Monte Carlo step (MCS) arriving at global energy minimum (it is related to convergence rate). In the case of amp = 2 the total MCS is 10,000 configurations, while 50,000 configurations are processed when amp = 3.

Though the convergence rate is changed according to the variation of steepness parameter (γ), VRMC always has high convergence rate as compared with MMC in all conditions. With respect to the crossing number, VRMC has a higher value than MMC in most cases. Though we do not report it here the acceptance ratio is constantly increased as the steepness parameter (γ) is increased. The increase of acceptance ratio might result from a more frequent sampling of the region near the VF, as we expected. Unfortunately we

TABLE V Comparison of the result between MMC and VRMC for model energy surface

Parameter γ[d]	(*amp*[a] = 2, *dx*[b] = 0.15, *RT*[c] = 0.5)		(*amp* = 2, *dx* = 0.15, *RT* = 1.2)		(*amp* = 3, *dx* = 0.10, *RT* = 1.5)	
	Step # at GEM	*Cross #*	*Step # at GEM*	*Cross #*	*Step # at GEM*	*Cross #*
0[e]	8840	0	3340	26	23890	15
1	2120	3	2190	62	5 10	14
3	2450	2	1200	49	11100	26
5	1550	1	1890	34	12040	20
	4800	1	450	44	115 0	12
10	3 10	2	360	51	1 810	18
20	2680	0	1390	33	6340	13
50	6210	0	1140	58	10590	16

[a] "amp" is the amplitue of cosine part in Eq. 1.
[b] "dx" is the moving boundary in Monte Carlo process.
[c] "RT" means virtual temperature in Metropolis test.
[d] Steepness Parameter controls the extent of restraint around valley region.
[e] $\gamma = 0$, means the conventional Metropolis MC.

cannot find any direct relationship between the value of γ and the efficiency of simulation in this system, but it is found that VRMC is remarkably efficient as compared with MMC in all cases.

2. Application to Chain Molecules

2.1. Alanine Dipeptide

After the full extended structure of alanine dipeptide (Ac-Ala-NHMe) is minimized for 1000 steps using the steepest descents method, we carry out Monte Carlo simulation combined with a simulated annealing procedure (from 1000 K with 3 % annealing rate of each 50 configurations) on the system at 300 K. Simulation proceeds during 10^5 configurations (100 Ksteps) using Consistent Valence Force Field (CVFF) [26] in vacuum condition with acceptance ratio of about ~ 0.5.

In the case of VRMC, we carry out the several short runs of preliminary simulations with the variation of steepness parameter to find the optimal value of steepness parameter for 5 Ksteps in the same condition as MMC with taking averaged conformational change of principal torsion angles (ϕ, ψ), $\langle \Sigma_i (\Delta\theta_i)^2 \rangle^{1/2}$, per every 50 steps. It is used as a indicator of sampling efficiency, which conceptually will be very similar to the rms difference of principal torsion angles between the current structure and the next structure [27]. As shown in Figure 5(a), the optimal value for the steepness parameter (γ) is found to about 5.0×10^2. From this fact, we can guess that the averaged conformational change is decreased by high constraint of Valley Function. Averaged conformational change is increased at medium range of the steepness parameter (γ) owing to the preferential sampling of more feasible region, while it is decreased at high value of γ due to high constraint in spite of the increase of acceptance ratio.

After these preliminary simulations, we carry out the main simulation of the VRMC for 100 Ksteps with the steepness parameter (γ) of 5.0×10^2 which is selected as optimum steepness parameter during the preliminary simulation process. The final 25 Ksteps in VRMC is progressed without the constraint of the VF in order to remove the structural bias by the constraint. The acceptance ratio in preselection process (Step 3 in Scheme 1) is 0.5507 and the consumed CPU time required for this process is only about 1% of the total CPU time.

From the comparison between the final conformation of MMC and that of VRMC, the unique structure is characterized ($\phi = -88^\circ$, $\psi = 86^\circ$). Table VI shows the summary of the comparison between VRMC at

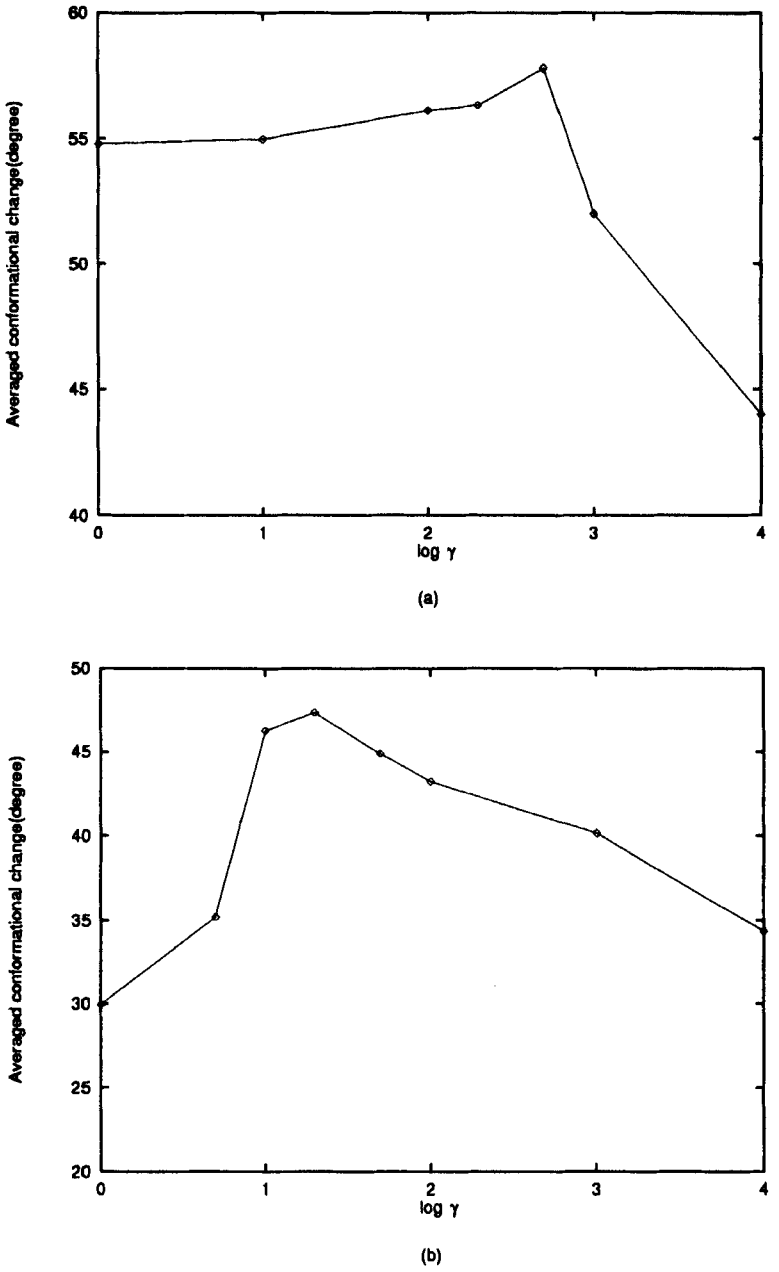


FIGURE 5 Averaged conformational change of principal torsion angles, $\langle \Sigma_i (\Delta\theta_i)^2 \rangle^{1/2}$ according to diverse steepness parameters (γ) during the Process of preliminary simulation for (a) alanine dipeptide for 5 Ksteps and (b) Met-enkephalin for 400 Ksteps.

TABLE VI Comparison of the result for peptide systems

	Alanine dipeptide		*Met-enkephalin*	
	MMC	*VRMC*	*MMC*	*VRMC*
$\langle\Sigma_i(\Delta\theta_i)^2\rangle^{1/2}$ (deg.)	42.34	47.02	19.17	43.23
Acceptance Ratio	0.5369	0.6147	0.4169	0.5539

$\gamma = 5.0 \times 10^2$ and MMC. Here we do not report the energy equilibration processes because we cannot find the remarkable difference of energy fluctuation and convergence rate due to the small size of system. On the other hand, the averaged conformational change of VRMC (first 75 K steps restrained by VF) is increased about 10 % compared to that of MMC and we can find the ascent of acceptance ratio in VRMC about 0.08.

From the results of VRMC and MMC for alanine dipeptide, we gain only a small increase of averaged conformational change and acceptance ratio. This result is magnified in the larger system as follows.

2.2. *Met-enkephalin*

The total simulation steps of MMC and VRMC on the pentapeptide, Met-enkephalin (H-Tyr^1-Gly^2-Gly^3-Phe^4-Met^5-OH), is 8000 Ksteps (6000 Ksteps with the constraint by Valley Function and 2000 Ksteps without the constraint). Preliminary simulations with diverse steepness parameters for determining the parameter (γ) before the main simulation of VRMC, are carried out for 400 Ksteps as in the case of the alanine dipetide. Other conditions are analogous to those of alanine dipeptide.

Figure 5(b) shows the variations of averaged conformational change according to the increase of γ. From this figure, we determine the steepness parameter as 2.0×10^1. This value of Met-enkephalin is smaller than that of alanine dipeptide owing to the size of system. Since there is only a little change in CPU times due to the variation of γ, the preselection process requires only a little time compared to the time for the calculation of the interaction potential function.

In the VRMC with γ of 2×10^1 on the system, the acceptance ratio in the preselection process is 0.7343, this value is higher than that of the alanine dipeptide owing to the lower γ. That is, the sampling bias to the direction of the valley region is about 27%.

The principal torsion angles of the final conformation are compared with previous data [28] as represented in Table VII. We exclude the terminal group (Tyr^1 and Met^5), because the terminal group has high conformational

TABLE VII Comparison of principal torsion angles for Met-enkephalin

Residue	*Gly*^{2}		*Gly*^{3}		*Phe*^{4}	
Angle (degree)	ϕ	ψ	ϕ	ψ	ϕ	ψ
Our Results	86	68	89	−153	−76	−172
Previous Work [a]	121	60	114	−121	−101	−153

[a] Previous Work is Ref. [28].

flexibility. Since the principal torsion angles ϕ and ψ of Gly^3 and Phe^4 mainly determine the over-all structure of Met-enkephalin, we could confirm the structural similarity between our result and previous work. Figure 6 shows the energy equilibration processes for every 400 steps in the simulation. This figure clearly shows the increased power of VRMC compared to that of MMC. In the case of MMC the system may be trapped in local minima during the full simulation, while VRMC makes it possible to escape from the local minima. This fact indicates that MMC could not escape from the local minima within 8000 Ksteps. However, VRMC enables the system to overcome the energy barrier and avoid trapping in the local minima. Table VI shows the acceptance ratio and the averaged conformational change of MMC and VRMC. The acceptance ratio of VRMC is

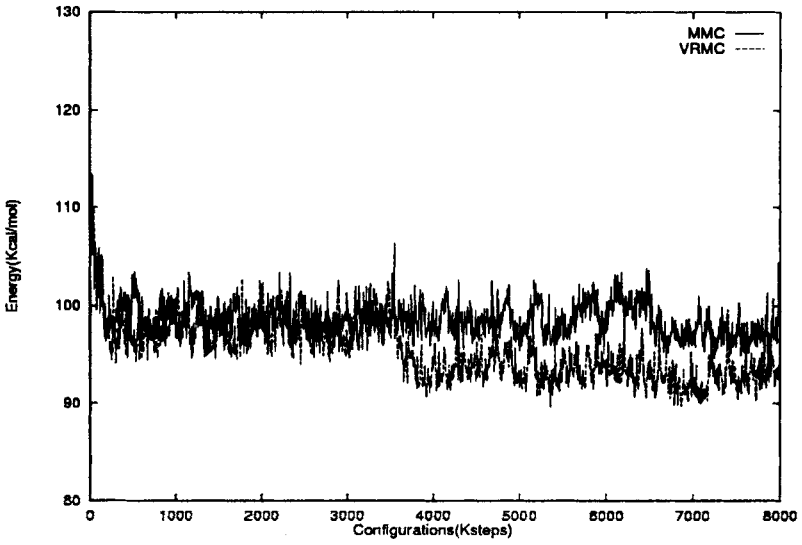


FIGURE 6 Comparison of the energy equilibration process (kcal/mol) between MMC and VRMC for Met-enkephalin during the process simulation, where the data is reported per every 400 steps.

higher than that of MMC by 0.1370 during the region with constraint. And we can find a remarkable increase of the averaged conformational change by VRMC. That is, the averaged conformational change of VRMC is increased about 125% as compared to that of MMC. This lower value of MMC probably stems from the trapping in local energy minimum.

IV. CONCLUSION

In this paper, we present the VRMC procedure, which takes into account the topology of the energy surface from *a priori* knowledge, then carries out a preferential sampling of the region near the energy minima and the valley between minima. Though it gives some bias in sampling process, supposing that the object of the simulation is twofold (finding the global energy minimum (GEM) and investigation of thermodynamic properties), VRMC is regarded as a good solution for the first-multiple minima problem. After finding the GEM, if we remove the preferential sampling, the second could be also achieved. We perform the VRMC without preferential sampling as the final stage of our simulations.

As mentioned in previous sections, our VRMC procedure is carried out on the model energy surface and the peptides. In both cases we can obtain successful results. The crossing number and the convergence rate show the efficiency of our new Monte Carlo procedure. From the comparison of the results between the alanine depeptide and the Met-enkephalin, we find that the power of VRMC would be increased when the system is larger. This method could be simply applied to commercial polymer systems, because it has only a few kinds of monomers unlike a protein and we could easily find the Valley Function.

Acknowledgements

This work was supported in part by the Korea Research Center for Theoretical Physics and Chemistry.

References

[1] Anfinsen, C. B. (1967). "The Formation of the Tertiary Structure of Proteins". *Harvey Lect.*, **61**, 95.

[2] Paine, G. H. and Scheraga, H. A. (1985). "Prediction of the Native Conformation of a Polypeptide by a Statistical–Mechanical Procedure. I. Backbone Structure of Enkephalin". *Biopolymers*, **24**, 1391.

[3] Paine, G. H. and Scheraga, H. A. (1985). "Prediction of the Native Conformation of a Polypeptide by a Statistical-Mechanical Procedure. II. Average Backbone Structure of Enkephalin". *Biopolymers*, **24**, 1547.
[4] Bruccoleri, R. E. and Karplus, M. (1987). "Prediction of the Folding of Short Polypeptide Segments by Uniform Conformational Sampling". *Biopolymers*, **26**, 137.
[5] Kostrowicki, J. and Scheraga, H. A. (1992). "Application of Diffusion Equation Method for Global Optimization to Polypeptides". *J. Phys. Chem.*, **96**, 7442.
[6] Karplus, M. and Petsko, G. A. (1990). "Molecular Dynamics Simulations in Biology". *Nature*, **347**, 631.
[7] Kim, T. K., Yoon, J. H., Shin, J. K. and Jhon, M. S. (1995). "Use of the High Directional Monte Carlo Methods to Predict the Low Energy Structures of Melittin". *Mol. Simulat.*, **15**, 177.
[8] Northrip, S. H. and McCammon, J. A. (1980). "Simulation Methods for Protein Structure Fluctuations". *Biopolymers*, **19**, 1001.
[9] Ripoll, D. R. and Scheraga, H. A. (1986). "On the Multiple-Minima Problem in the Conformational Analysis of Polypeptides. II. An Electrostatically Driven Monte Carlo Methods – Tests on Poly (L-Alanine)". *Proc. Natl. Acad. Sci. USA*, **1986**, 1283.
[10] Shin, J. K. and Jhon, M. S. (1992). "High Directional Monte Carlo Procedure Coupled with the Temperature Heating and Annealing as a Method to Obtain the Global Energy Minimum Structure of Polypeptides and Proteins". *Biopolymers*, **31**, 177.
[11] Yoon, J. H., Shin, J. K. and Jhon, M. S. (1995). "Determination of C-Terminal Structure of Human C-Ha-Ras Oncogenic Protein". *J. Comp. Chem.*, **16**, 478.
[12] Rao, M. and Gerne, B. J. (1979). "On the Force Bias Monte Carlo Simulation of Simple Liquids". *J. Chem. Phys.*, **71**, 129.
[13] Rossky, P. J., Doll, J. D. and Friedman, H. L. (1978). "Brownian Dynamics as Smart Monte Carlo Simulation". *J. Chem. Phys.*, **69**, 4628.
[14] Shin, J. K., Yu, J. Y. and Jhon, M. S. (1992). "Acceleration of Convergence Rates by High Directional Monte Carlo Sampling Procedure". *J. Chem. Phys.*, **97**, 9283.
[15] Frantz, D. D., Freeman, D. L. and Doll, J. D. (1992). "Extended J-Walking to Quantum-System-Applications to Atomic Clusters". *J. Chem. Phys.*, **97**, 5713.
[16] Hong, S. D. and Jhon, M. S. (1997). "Restricted Random Search Methods Based on Taboo Search in the Multiple Minima Problem". *Chem. Phys. Lett.*, **267**, 422.
[17] Allen, M. P. and Tildesley, D. J. (1989). Computer Simulation of Liquid, Clarendon Press, p. 24.
[18] Robson, B. and Garnier, J. (1988). Introduction to Proteins and Protein Engineering, Elsevier, 1988, p. 493.
[19] Choi, S. H., Shin, J. K., Yu, J. Y. and Jhon, M. S. (1994). "Molecular Dynamics Simulations of *trans*- and *cis*-*N*-Acetyl-*N'*-Methylamides of Xaa-Pro Dipeptides". *J. Molec. Strc.*, **323**, 233.
[20] Koch, I., Kaden, F. and Selbig, J. (1992). "Analysis of Protein sheet Topologies by Graph Theoretical Methods". *Proteins: Structure, Function, and Genetics*, **12**, 314.
[21] Boberg, J., Salakoski, T. and Vihinen, M. (1992). "Selection of a Representative Set of Structure From Brookhaven Protein Data Bank". *Proteins: Structure, Function, and Genetics*, **14**, 265.
[22] Grindley, H. M., Artymiul, P. J., Rice, D. W. and Willett, P. (1993). "Identification of tertiary Structure Resemblance in Proteins Using a Maximal Common Subgraph Isomorphism Algorithm". *J. Mol. Biol.*, **229**, 707.
[23] Weiss, N. A. and Hassett, A. J. (1987). Introductory Statistics, Addision-Wesley Publishing Company, p. 628.
[24] Burden, R. L. and Faires, J. D. (1989). Numerical Analysis, PWS-KENT Publishing Company, 1989, p. 430.
[25] Tejwani, G. A. and Rattan, A. K. (1997). "Met-Enkephalin in Alteration in the Rat during Chronic Injection of Morphine and/or Midazolam". *Brain Res.*, **775**, 119.
[26] Hagler, A. T., Stern, P. S., Sharon, R., Becker, J. M. and Naider, F. (1979). "Computer Simulation of the Conformational Properties of Oligopeptides. Comparison of Theoretical Methods and Analysis of Experimental Results". *J. Am. Chem. Soc.*, **101**, 6842.

[27] Noguti, T. and Gō, N. (1985). "Efficient Monte Carlo Method for Simulation of Fluctuating Conformations Native Protein". *Biopolymers*, **24**, 527.
[28] Graham, W. H., Varter II, E. S. and Hicks, R. P. (1992). "Conformational Analysis of Met-Enkephalin in Both Aqueous Solution and in the Presence of Sodium Dodecyl Sulfate Micelles Using Multidimensional NMR and Molecular Modeling". *Biopolymers*, **32**, 1755.