EFFICIENT STOCHASTIC GLOBAL OPTIMIZATION FOR PROTEIN STRUCTURE PREDICTION

Yingyao Zhou and Ruben Abagyan*

Skirball Institute of Bimolecular Medicine
Biochemistry Department
New York University Medical Center
540 1st Avenue, New York, NY 10016

INTRODUCTION: Why not MD or MC?

Biological macromolecules, large chain molecules with hundreds of torsion angles, adopt compact, uniquely folded and rigid conformations that correspond to their global free energy minimum. Predicting this unique conformation from a vast number of alternatives, for the whole protein or its parts, is the biggest challenge of computational biology. One of the difficulties is conceptual. To evaluate the free energy correctly we need to account for the dynamic nature of the entire system, including mobile water molecules, flexible side-chains and soft vibrational modes of a solute. Molecular Dynamics (MD, reviewed in Ref. 1-4) or Monte Carlo simulations (MC, reviewed in Ref. 4-8) in water can be applied to sample the conformational space and evaluate the free energy. However, these methods are still too slow to reach the biologically relevant folding times for proteins or even large peptides[2,9].

Fortunately, the free energy of surrounding water molecules can be implicitly evaluated through the electrostatic and surface effects[10], the side chain mobility contribution to the free energy can be roughly estimated through its solvent exposure, and the vibrational contribution can be considered comparable in different folded conformations. Therefore, the computationally expensive MC and MD methods, aimed at the generation of a Boltzmann ensemble, can be replaced by much more efficient stochastic global optimization methods aimed at identification of a unique global minimum in the smallest number of iterations. Global optimization methods can be classified into zero-order and first-order algorithms depending on whether a local minimization step is performed after each iteration[11]. Two reasons account for the clear superiority of the first-order methods for peptides and proteins[12]. The first reason is the energy improvement due to local minimization, which is often comparable to the variation of the energy values between different local minima. Second, an adequate standard local optimization method,

* Corresponding author

Rigidity Theory and Applications
Edited by Thorpe and Duxbury, Kluwer Academic/Plenum Publishers, 1999

using analytical energy derivatives, is the most efficient way to identify the nearest local minimum, and such algorithms as MC, MD, or random sampling will be far inferior in performing the same task.

Here we describe the principle of optimal sampling bias as an algorithm for generation of the random moves in a stochastic global optimization method and demonstrate a drastic improvement of the efficiency due to the optimal bias. The principle of optimal bias was first introduced in 1994[13] as a linear sampling bias; in this essay, we consider another optimization model and arrive at the square-root sampling bias rule. This algorithm is general and applicable to stochastic global optimization of any function, both continuous and discrete.

GLOBAL OPTIMIZATION: How to find a global minimum of a function?

In the Introduction we argued that free energy can be assigned to a single polypeptide chain conformation, and, therefore a unique native folded conformation can be predicted by global energy optimization. The global optimization algorithm is not bound by the trajectory continuity or Boltzmann ensemble generation requirements, and, therefore, has a larger potential to do what an optimization algorithm does best, i.e. find the minimum in the minimal number of function evaluations.

Global optimization is used in many fields[14,15], but in protein structure prediction it is additionally complicated by high dimensionality (the smallest protein has about 100 essential degrees of freedom), and small separation between the native energy minimum and the abundant false energy minima. The high dimensionality of the problem makes any systematic search impossible, a problem known as the Levinthal paradox[16]. To make matters worse, the optimization problem can not be considered at the discrete rotamer level since small 'continuous' angular adjustments are essential for favorable packing. Finally, the small energy difference between the correct and incorrect minima and the exponential growth of the density of the non-native states with energy impose strict requirements on the accuracy of energy evaluation (less than about 1 kcal/mol)[5].

Numerous approaches have been used to attack the global optimization problem in protein structure prediction, with some success[1-8] (Table 1). These methods are initially classified according to whether they are deterministic or not; stochastic methods are further subdivided according to the degree of similarity between conformations generated in consecutive iterations of the search algorithm.

Table 1. Classification of global optimization methods based on the degree of history dependence.

Global optimization Methods			
Deterministic search	Stochastic search		
	History independent	Intermediate history dependent	Maximum history dependent
Systematic search[17,18]*, build-up method[19-21]*, diffusion equation method (DEM)[24], packet annealing method (PA)[23].	Randomize all variables in a step: Local minimization from multiple random starting conformations.	Large change of one/a group of variables in a step: Genetic algorithm (GA)[36], lattice model MC[37], MC-Minimization (MCM)[11], standard Metropolis MC[26,27], electrostatically driven MC (EDMC)[38], extended scaled collective variable (ESCV) MC[33], restrictive MC[39], biased MC (BMC)[40,41], optimal-bias MC-Minimization (OBMCM)[13].	Changes all variables continuously in a step: Molecular dynamics (MD)[1-4,28-30], Local energy minimization[11,31], scaled collective variable (SCV) method[32], extended SCV MC[33]. High directional MC (HDMC)[34], some side chain MC[35].

* Only search the combinations of pre-calculated local minima.

In principle, deterministic methods are guaranteed to find the global minimum. In practice, however, such methods require the adoption of certain simplifying assumptions that compromise their accuracy. Systematic search[17,18] and the build-up methods[19-21] assume that the global minimum of a complete structure is a combination of a relatively small number of local minima of structural fragments. Both assumptions turn out to be wrong; many intramolecular interactions are nonlocal (about 50% by the contact area estimate[22]), the globally optimal conformation may contain strained fragments far from their local minima, and the number of local conformations to be retained is exceedingly large. The packet annealing (PA)[23] and the diffusion equation (DEM)[24] methods introduce an elegant concept of smoothing the probability distribution and the energy surface, respectively, and reduce the global optimization problem to a series of local minimizations. However, the deterministic character of these methods is something of an illusion. DEM procedure encounters numerous bifurcation points during the annealing process and a slight inaccuracy in the free energy function can lock the search into the wrong path[25].

We will distinguish between the MC or MD methods, which are designed to generate a Boltzmann ensemble, and global optimization algorithms (such as simulated annealing[26,27]) which attempt to identify a single conformation corresponding to the global minimum of a free energy function (in the pseudo-potential energy form).

Most of the MC-like stochastic global optimization strategies employ a three-step iteration: (i) modify the current conformation by means of a random move; (ii) evaluate its energy; (iii) accept or reject the new conformation according to an acceptance criterion. The random moves can be ranked by magnitude of change with respect to the current conformation (Table 1). The first group contains algorithms in which the generated conformations do not depend on the previous ones. The second group keeps maximum memory by changing all variables quasi-continuously according to certain rules or by some small amplitude random deviations. This category contains molecular dynamics (MD) [28-30], local energy minimization methods[31], scaled-collective-variable (SCV) method[32], extended SCV Monte Carlo (ESVC)[33], high directional MC (HDMC)[34], and some side chain MC methods[35]. The third group takes an intermediate approach by changing one variable or a group of variables (generally correlated variables) at a time. This group contains most of the global optimization methods including genetic algorithm (GA)[36] -based methods, lattice model MC[37] most other MC methods[4-8].

HISTORY-DEPENDENCE OF CONFORMATIONAL SEARCHES

Different history-dependent protocols inherit current structural information to varying degrees. Genetic-algorithm (GA) methods[36] make a single random change with each 'mutation', and conformational recombination extends the random change to a wider range. Various lattice MC methods[37] make local elemental jumps, which may involve modifying three to five bonds, and translation/rotation of a portion of the chain as well. In a global step of the MCM method[11], a random change of one angle is accompanied by a local minimization with respect to all torsion angles. Some methods[26] make sequential change to one variable at a time in standard Metropolis MC (MMC) implementation, the amplitude of randomization being tuned to ensure a sufficiently high acceptance ratio. Some other MMC methods[27] randomly change one angle with an amplitude of 90°. Electrostatically driven MC (EDMC) method[38] switches between a random prediction, where one dihedral angle is randomized with an amplitude of 180°, and an electrostatically driven move, where two coupled dihedral angles are changed with an amplitude estimated from the local electric field. Restricted MC methods[39] replace continuous side chain orientations by discrete rotamer values. Biased MC (BMC)[40,41] makes three- or four-residue backbone move at once, the statistical distributions of backbone dihedral angles and rotamer libraries for side

chain angles are taken into account in the conformation generation. Optimal-bias-MC-with-minimization (OBMCM, also referred to as Biased Probability MC[13]) modifies groups of correlated backbone or side chain variables according to optimal statistical distributions. MD, local energy minimization methods, SCV/ESCV, HDMC make small amplitude changes to all variables determined by dynamic equations or local energy landscape. Some side chain MC methods[35] change all side chain torsion angles simultaneously by 0^o or $\pm 10^o$.

How similar should the next conformation be to the previous one? Virtually identical as in a MD method, or totally unrelated as in a random search? In the following section we investigate this question.

COMPARISON OF GLOBAL OPTIMIZERS OF ZERO ORDER (WITHOUT MINIMIZATION)

The performance of the global optimization methods can be tested on small peptides. Met-enkephalin, the Tyr-Gly-Gly-Phe-Met pentapeptide, has been extensively studied and frequently used as a test peptide before[11,12,19,32,34,42], but it is too small and conformationally unusual for a good protein-like benchmark. Two other test peptides were used instead: an α-helix and a β-hairpin. The selected helix is a 12-residue synthetic peptide Acetyl-Glu-Leu-Leu-Lys-Lys-Leu-Leu-Glu-Glu-Leu-Lys-Gly-COOH crystallized and solved by Hill *et al*[43]. The second peptide is a 13-residue ubiquitin fragment (residue number 3-15) suggested to be an independent β-hairpin fold by circular dichroism and NMR studies[44,45].

We performed a series of Metropolis Monte Carlo (MMC) simulations without minimization from random starting conformations for four different move generation algorithms. (1) Change one randomly selected variable at each step, with amplitude of 30°, 90° and 180°. (2) Change two coupled variables such as backbone φ-ψ angles or χ_1-χ_2 angles in a randomly selected residue with 180° amplitude. (3) Change all variables of a randomly selected residue, (φ, ψ and χ), with 180° amplitude. (4) Randomize all variables with 2° amplitude after each step.

A simulation temperature of 600K was used for all simulations to ensure the same 'energetic accuracy' of 1.2 kcal/mol. Each type of simulation was repeated ten times and the conformational energies were recorded. Average angular RMSDs of conformations generated in adjacent steps represent the scale of a random move. The average best energies after a certain number of energy evaluation 1×10^5 for the α-helix and 5×10^4 for the β-hairpin), as well as their standard deviations and acceptance ratios ρ, are shown in table 2.

The result shows that neither smallest nor largest random moves result in good performance. In general, a good move is the one generating the largest change at a given temperature and acceptance ratio. That is exactly what a good biased move of several angles at a time allows to be accomplished. For the above two benchmarks, the optimal-bias MC algorithm (without minimization) reached E_{min} of -132 ± 5 kcal/mol, acceptance

Tabel 2. MMC simulations of the 12-residue α-helix and the 13-residue β-hairpin.

	12-residue α-helix			12-residue β-hairpin		
Test type	RMSD (°)	E_{min} (kcal/mol)	ρ	RMSD (°)	E_{min} (kcal/mol)	ρ
1	0.16	$\sim10^4$	0.09	0.15	$\sim10^4$	0.37
1	0.50	-93± 7	0.28	0.56	-41 ± 11	0.28
1	0.78	-122± 7	0.22	0.83	-76 ± 7	0.22
2	1.4	-111 ± 7	0.16	1.2	-71 ± 7	0.15
3	1.5	-89 ± 9	0.15	1.5	-65 ± 8	0.12
4	2.0	$\sim10^5$	0.00	2.0	$\sim10^6$	0.00

ratio of 0.28 for the α-helix, and E_{min} of –88 ± 8 kcal/mol, acceptance ratio of 0.28 for the β-hairpin while changing two variables at a time.

GLOBAL OPTIMIZERS WITH LOCAL MINIMIZATION ARE SUPERIOR

In 1997 Li and Scheraga introduced a new global optimization method in which each random step is followed by local energy minimization[11]. Even though they called it Monte Carlo-Minimization (MCM), the procedure did not obey the local balance condition and can only be considered as a stochastic global optimization algorithm. But how important is local energy minimization after each large random move? On the one hand, spending valuable energy evaluations on local energy optimization in basically the same conformational vicinity instead of more extensive sampling may sound wasteful. On the other hand, minimization algorithms using function derivatives are much more efficient than random sampling in finding the local energy minimum, and the unminimized values are not really representative because of the ruggedness of the energy landscape. The number of energy evaluations spent on local adjustments is typically hundreds of times larger than the number of random moves! Maybe we should use only a partial minimization thus saving the function evaluations for more random steps, given the fact that the energy drops much faster in the beginning of the minimization?

The above questions were systematically analyzed[12] and the conclusion was that allowing a full local optimization following each random step resulted in the best performance, with both partial and no minimization being clearly inferior under the constraint of the total number of energy evaluations. In other words, making 100,000 high quality moves is preferable over making 10,000,000 low quality moves.

In the MCM algorithm a randomly chosen angle was changed by a random, evenly distributed value. Introduction of the optimal bias into the random step resulted in another drastic increase of the global optimization performance[13].

OPTIMAL BIAS FOR STOCHASTIC GLOBAL OPTIMIZATION (OBMCM)

We know that the groups of torsion angles in peptides and proteins have certain preferences, i.e. some values are found more frequently than others. The preferences of the backbone angles (φ-ψ angles) as well as the side chain rotamer libraries have been described [46-50], and the correlations between the backbone and side chain angles have been studied as well[51]. How can we take advantage of these statistical preferences? We know that almost every protein or peptide contains some rare, unusual torsion angles; therefore, should one still use a flat probability distribution (as in the MCM method) to ensure that these rare values are sampled frequently enough? Or should we just use the discrete peaks of the distributions (the rotamers)[39] and hope that the rest will be taken care of by local minimization? The answers to these questions are important; as we will see later, the optimization efficiency is actually more sensitive to the answer to this question than to whether one uses simulated annealing or constant temperature, or whether one uses multiple independent runs or exchanges information between simulations.

There are basically two major alternatives: uniformly distributed random moves, and moves biased according to some statistical information. The statistical information may be sequence-independent *a priori* information[13] derived from the structures in the Protein Data Bank, or the statistical information accumulated during the simulation[39]. Configurational-bias Monte Carlo (CBMC) simulations have been introduced very early on (a good review of CBMC methods can be found in Chapter 13 of Frenkel and Smit's book[53]), but the ability to generate a Boltzmann ensemble, an appropriate concern for a Monte Carlo

algorithm, was the primary focus. However, the primary objective of a stochastic global optimization algorithm is identification of the global minimum in a minimal number of function evaluations, a different goal that is not necessarily compatible with the local balance principle. For example, the local minimization after each move violates the local balance but is necessary for efficient global optimization. Therefore, derivation of the bias which is optimal from the global optimization point of view, a problem addressed by OBMCM/BPMC algorithm[13], became an important objective.

The idea is to use the geometrical preferences of local groups of coupled torsion angles, preferences that can be pre-calculated, to guess their final values defined by all the interactions in a larger molecule, under the assumption that the global interactions are random with respect to the local preferences. Let us denote a group of coupled variables by vector $\mathbf{x}$ and its value corresponding to the global minimum state as $\mathbf{x}^0$. Therefore an arbitrary protein conformation can be represented by its n variable groups as $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n)$ and its lowest-energy conformation as $(\mathbf{x}_1^0, \mathbf{x}_2^0, \ldots, \mathbf{x}_i^0, \ldots, \mathbf{x}_n^0)$. We further assume that for all possible protein targets, $\mathbf{x}^0$ satisfies statistical ***distribution function*** $S(\mathbf{x}^0)$. Actually a separate distribution function for each type of amino acid can be generated, and the distribution function for the *jth* type of amino acid will be $S_j(\mathbf{x}^0)$. In MC-like algorithms, one randomly selects a vector, for instant x_i, to change during a global move, $\mathbf{x}_i$ is to be assigned a new value $\mathbf{x}_i'$ according to a probability function $f_j(\mathbf{x}_i')$ (assuming $\mathbf{x}_i$ belongs to the *jth* type of amino acid), which is to be called the ***sampling function*** later. The question is what are the sampling functions resulting in identification of the correct answer in the minimal number of energy evaluations.

Unfortunately, the question does not have a clear answer unless an analytical target function is specified to measure the performance of a global sampling. We propose here two target functions, (i) $f_j(\mathbf{x})$ maximize the probability of finding the lowest-energy conformation of a randomly given protein within a global sampling; (ii) $f_j(\mathbf{x})$ minimize the average number of global sampling steps required to successfully predict a randomly given protein. In order to simplify our analyses, it is assumed that all the n variable groups are randomly re-sampled according to their corresponding sampling function at a specific global step and there is no local minimization afterwards. The proofs for the most general case involving n continuously distributed variable groups are presented in Ref. 13 for the first target function and in the Appendix for the second target function.

We will try to guess the true value $\mathbf{x}^0$ of a vector $\mathbf{x}$, with the knowledge that $\mathbf{x}^0$ takes the value $\mathbf{x}_1$ with probability S_1, takes the value $\mathbf{x}_2$ with probability S_2, ..., takes the value $\mathbf{x}_n$ with probability S_n. In a biased guess, the sampling function f allows one to sample the value $\mathbf{x}_1$ with probability f_1, sample the value $\mathbf{x}_2$ with probability f_2, ..., sample the value $\mathbf{x}_n$ with probability f_n. The game and some possible strategies are illustrated in Figure 1, where the current target value of $\mathbf{x}$ is marked by a star and the distribution function 5 is resembled by the shaded bars. Random vectors are generated according to the sampling function until the true value $\mathbf{x}^0$ of x is hit. An additional condition is independence of each guess on the previous guess. This is counterintuitive in a simple guessing game, e.g., if there are only two states, and you gave the wrong answer, the next one will be right. However, in a real simulation with an MC-like calculation the global context of the same group of variables is constantly changing and the independence assumption can be justified.

Game 1: Find the optimal f_i so that the probability of correctly guessing the true value $\mathbf{x}^0$ in each guess is maximized.

If the actual value is $\mathbf{x}_i$, one will guess it correctly with the probability f_i in a step. Since such an event happens with a probability of S_i, the overall probability to be maximized is $P = \sum_i S_i f_i$, under the normalization condition $\sum_i f_i = 1$.

This is equivalent to maximize $P + \lambda\left(\sum_i f_i - 1\right)$, where λ is the Lagrange multiplier and f_i can be treated as n independent variables. It is then straightforward to derive the

optimal sampling function by setting the derivatives of this target function with respect to f_i equal to zero. The conclusion is $\boxed{S_i / f_i = \lambda}$, i.e., the optimal sampling function equals the original distribution function.

Game 2: Find the optimal f_i so that the average number of unsuccessful guesses is minimized.

Let us note that if the true value is $\mathbf{x}_i$, and the probability of sampling of this particular state is f_i in each trial, therefore, it will take $1/f_i$ trials on average to find the true value. Since such an event occurs with expected probability S_i, the 'ensemble average' of the average numbers of required guesses is $\bar{N} = \sum_i S_i f_i^{-1}$, under the normalization condition $\sum_i f_i = 1$.

Optimizing $\bar{N} + \lambda\left(\sum_i f_i - 1\right)$, we arrive at $\boxed{S_i / f_i^2 = \lambda}$, i.e., square-root sampling functions minimize the cost of global minimization.

As mentioned before, the same conclusions can be generalized for any arbitrary number of vectors with continuous distributions $S(\mathbf{x})$. The linear bias $f_j(\mathbf{x}) = S_j(\mathbf{x})$ maximizes the correct guessing probability[13], and square-root bias $f_j(\mathbf{x}) \propto \sqrt{S_j(\mathbf{x})}$ minimizes the average number of guesses required.

SUPERIOR PERFORMANCE OF THE OPTIMAL-BIAS-MCM

Comparison between the zero-order MMC and OBMC (with both the linear and the square-root bias) show that both biased sampling algorithms out-performed the uniform random sampling scheme. Both linear and square-root bias result in comparable performance on both previous benchmarks. However, because the square-root bias allows sampling of the rarely populated zones of the torsion space much more frequently than the linear biasing functions, we expect that less standard benchmarks would reveal a better performance of the square-root bias.

We also compared the first-order method such as unbiased MCM and linear-bias MCM algorithms using the 12-residue α-helix[13] and a more realistic ββα peptide (results are not shown) as a benchmark. The performance increase due to the optimal bias varies

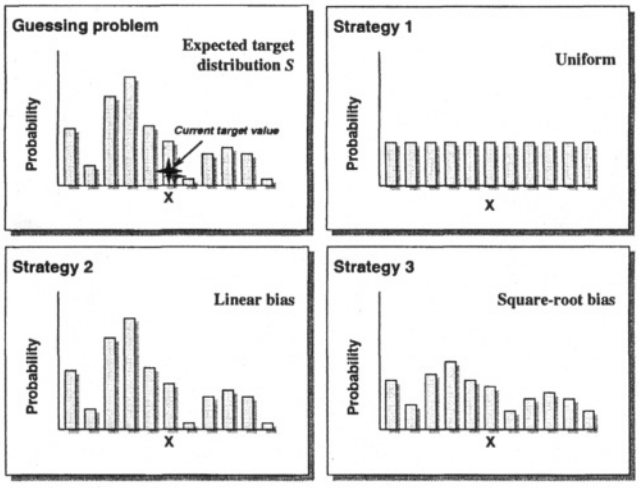

Figure 1. Schematic diagram of various sampling strategies.

but on average is about ten fold for a mixed α/β topology. However, these calculations take several days even for the OBMCM algorithm and we were not able to reach the solution with the MCM algorithm in a reasonable time.

Waiting until each algorithm reaches its global minimum may take a lot of time, and this time varies strongly between simulations. Previously, we used a more stable performance criterion, which was a fraction of the set of many low energy minima visited after the fixed number of function evaluations[12]. Here we returned to the old measure of the number of iterations until the global minimum was reached, but we softened the minimum identification criterion and averaged this number with up to 10 independent simulations. $R(n)$ is the fraction of systems that have reached the global minimum after n energy evaluations. By reaching the global minimum, we mean that a simulation hits a conformation of correct secondary structure and also has energy within 3 kcal/mol above the lowest energy found by pre-simulations. Success rate, also called cumulative distribution function (CDF), has been used before to study the folding time of the simulated annealing algorithm[54].

$R(n)$ can be approximately described by a Poisson distribution[54]. Taking the simulation cost for the early stage of forming compact globular conformations into account, we use the following expression to describe the success rate:

$$R(n) = 1 - e^{-q(n-n_0)}, \text{ with } n > n_0,$$

where q is a constant. $1\text{-}e^{-q}$ ($\approx q$, for $q << 1$) can be interpreted as the probability of hitting a global minimum conformation per energy evaluation. Since n_0 is the average number of energy evaluations required to lower the system energy to a plateau and $1/q$ is the mean value of the Poisson distribution, $n_0 + 1/q$ is the measurement of overall simulation cost including both early and latter stages in a simulation.

The benchmarks used here are the 12-residue α-helix[43] and a 12-residue β-hairpin[55]. Their global minimum energies were –185.0 kcal/mol and -198.6 kcal/mol, respectively. Three algorithms were analyzed: (i) Lee *et al.*[40], biased MC (BMC) with linear sampling function but without minimization. We used the distributions derived in Ref. 13 for

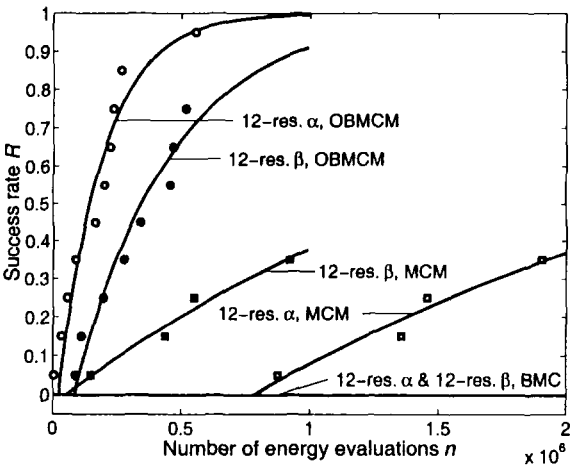

Figure 2. The success rate of BMC (Lee et al., 1993), MCM (Li & Scheraga, 1987) and OBMCM (Abagyan & Totrov, 1994) simulations. Results for the BMC simulations are denoted by the horizontal line, since no successful simulation was found within 4×10^6 energy evaluations under this scheme.

Table 3. The performance measurements of MMC and OBMCM methods

	12-residue α-helix		12-residue β-hairpin	
	MCM	OBMCM	MCM	OBMCM
n_0	7.80×10^{5}	2.22×10^{4}	5.06×10^{4}	8.20×10^{4}
$1\text{-}e^{-q} \approx q$, for $q << 1$	3.77×10^{-7}	5.81×10^{-6}	5.02×10^{-7}	2.65×10^{-6}
$n_0 + 1/q$	3.43×10^{6}	1.94×10^{5}	2.04×10^{6}	4.60×10^{5}

backbone sampling, but rotamer libraries for the side chain sampling; (ii) Li & Scheraga[11], MCM with uniform sampling and minimization; (iii) Abagyan & Totrov[13], OBMCM with the linear-bias and minimization.

Ten simulations for each case were initiated under constant simulation temperature 600K. q and n_0 values were then derived from the data. The results are shown in Figure 2 and Table 3. We found that OBMCM is 18 times faster compared to the unbiased MCM in the alpha-helix simulation, and 4.4 times faster in the beta-hairpin simulation. No successful simulations were found for the BMC case, the lowest energies reached by this protocol within 4×10^{6} functional calls were –151.1 kcal/mol and –171.2 kcal/mol for the α-helix and the β-hairpin, respectively (therefore $n_0 > 4\times10^{6}$).

SUMMARY

The native structure of a protein may be described with reasonable accuracy as the global minimum of the free energy (in the pseudo-potential energy form), only as a function of free torsion angles. Therefore, global optimization methods might be preferable over methods designed to create dynamic ensembles, such as MD or MC that are bound by the trajectory continuity requirement or the local balance requirement.

The Monte Carlo Minimization (MCM) method outperforms zero order MC-like stochastic global optimization protocols.

The Optimal-Bias-MCM method further improves the sampling efficiency by an order of magnitude by incorporating the optimal-bias into MC conformation generation. The square-root bias derived in this work and the linear bias[13] are two possible strategies.

The OBMCM algorithm can predict a 23-residue ββα peptide[56], with 70 essential torsion angles and 385 atoms, starting from completely random conformations. (Figure 3).

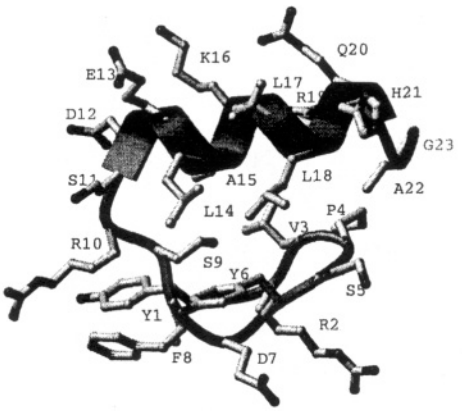

Figure 3. The predicted structure of a 23-residue ββα peptide (Ref. 56) with OBMCM method.

APPENDIX: OPTIMAL CONTINUOUS SAMPLING FUNCTIONS IN GLOBAL SAMPLING

If a group of coupled torsion angles of a residue (φ-ψ, and/or χ_1-χ_2) is denoted by vector $\mathbf{x}$, the lowest-energy conformation of protein consisting of n such groups of variables can be denoted as $(\mathbf{x}_1^0, \mathbf{x}_2^0, \ldots, \mathbf{x}_i^0, \ldots, \mathbf{x}_n^0)$, where $\mathbf{x}_i^0$ is the values of the ith group of torsion angles in the lowest-energy protein conformation. $\mathbf{x}_i^0$ has an *a priori* continuous probability distribution $S_j(\mathbf{x}_i^0)$ in the subspace formed by the vector, where j denotes the type of amino acid $\mathbf{x}_i$ belongs to. We sample each variable group $\mathbf{x}_i$ according to a sampling function $f_j(\mathbf{x})$. Randomly given a representative protein, we consider the problem of finding the optimal sampling functions that minimize the average number of energy evaluations required for successful structure prediction.

Following the same assumption made in Ref.13, i.e., the probability of finding the true value $\mathbf{x}_i^0$ of variable group $\mathbf{x}_i$ is proportional to $f_j(\mathbf{x}_i^0)$, when a global sampling is made for $\mathbf{x}_i$ according to the sampling function f_j. Since the probability of finding the true conformation at this specific conformation generation step reads

$$P = c\prod_{i=1}^{n} f_j(\mathbf{x}_i^0),$$

where c is a constant, it takes $1/P$ steps to find the true conformation on average. The S-ensemble average is the mean number of iterations:

$$\overline{N} = \frac{1}{c}\int_{\mathbf{x}_1^0}\int_{\mathbf{x}_2^0}\cdots\int_{\mathbf{x}_n^0}\prod_{i=1}^{n} S_j(\mathbf{x}_i^0)\prod_{i=1}^{n} f_j^{-1}(\mathbf{x}_i^0)d\mathbf{x}_i^0,$$

$$\overline{N} = \frac{1}{c}\prod_{i=1}^{n}\int S_j(\mathbf{x}_i^0) f_j^{-1}(\mathbf{x}_i^0)d\mathbf{x}_i^0.$$

Since $\overline{N}$ is always positive, maximizing $\overline{N}$ is equivalent to maximize

$$\ln\overline{N} = \sum_{i=1}^{n}\ln\int S_j(\mathbf{x}_i^0) f_j^{-1}(\mathbf{x}_i^0)d\mathbf{x}_i^0 - \ln c$$

to find the optimal sampling functions, we set $\delta\ln\overline{N}$ to zero:

$$\delta\ln\overline{N} = -\sum_{i=1}^{n}\int S_j(\mathbf{x}_i^0) f_j^{-2}(\mathbf{x}_i^0)\delta f_j(\mathbf{x}_i^0)d\mathbf{x}_i^0 \cdot \left(\int S_j(\mathbf{x}_i^0) f_j^{-1}(\mathbf{x}_i^0)d\mathbf{x}_i^0\right)^{-1} = 0.$$

Given the normalization conditions for the sampling functions, $\int\delta f_j(\mathbf{x}_i^0)d\mathbf{x}_i^0 = 0$, in order for the above equation to hold for any arbitrary function δf_j, we have

$$f_j(\mathbf{x}) = \frac{1}{c'}\sqrt{S_j(\mathbf{x})},$$

where c' is the normalization constant equal to $\int\sqrt{S_j(\mathbf{x})}d\mathbf{x}$.

ACKNOWLEDGEMENTS

We thank NIH (Grant R01 GM55418-01) and DOE (Grant DE-FG02-96ER62268) for financial support (this does not constitute an endorsement by both agencies of the views expressed in the article). We also thank Alex Morrill and Sheila Silvestein for careful reading of the manuscript.

REFERENCES

1. R. Elber, *Curr. Opin. Struct. Biol.*, 6:232 (1996).
2. C.L. Brook III, *Curr. Opin. Struct. Biol.*, 5:211 (1995).
3. T. Schlick, E. Barth, and M. Mandzink, *Ann. Rev. Biophys. Biolmol. Struct.,* 16:179 (1997).
4. B.J. Berne and J.E. Straub, *Curr. Opin. Struct. Biol.,* 7:181 (1997).
5. R.A. Abagyan, Eds. W.F. van Gunsteren, P.K. Weiner, and A.J. Wilkinson, *Computer Simulation of Biomolecuar Systems, Theoretical and Experimental Applications*, Kluwer Academic Publisher, vol. 3 pp.363-394 (1997).
6. H.A. Scheraga, *Biophys. Chem.*, 59:329 (1996).
7. M. Karplus and A. Sali, *Curr. Opin. Struct. Biol.*, 5:58 (1995).
8. G. Nemethy and H.A. Scheraga, *FASEB J.,* 4:3189 (1990).
9. E. Demchuk, D. Bashford, and D.A. Case, *Fold. Des.*, 2:35 (1997).
10. B. Honig and A. Nicholls. *Science,* 268:1144 (1995).
11. Z. Li and H.A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, 84:6611 (1987).
12. R. Abagyan and P. Argos, *J. Mol. Biol.*, 225:519 (1992).
13. R. Abagyan and T. Totrov, *J. Mol. Biol.*, 235:983 (1994).
14. B. Berg, *Nature,* 361:708 (1993).
15. S. Kirkpatrick, C.D. Gellatt, and M.P. Vecchi, *Science*, 220:671 (1983).
16. C. Levinthal, Eds. P. Debruner, J.C.M. Tsibris, and E. Munck, *Mossbauer Spectroscopy in Biological Systems,* Univ. Illinois Press, pp. 22-24 (1969).
17. R.E. Bruccoleri and M. Karplus, *Biopolymers,* 26:137 (1987).
18. T. Schaumann, W. Braun, and K. Wüthrick, *Biopolymers*, 29:679 (1990).
19. M. Vásquez and H.A. Scheraga, *Biopolymers*, 24:1437 (1985).
20. S. Vajda and C. DeLisi, *Biopolymers*, 29:1755 (1990).
21. I. Simon, L. Glasser, and H.A. Scheraga, *Proc. Nat. Acad. Sci., U.S.A.*, 88:3661 (1991).
22. R.A. Abagyan and M.M. Totrov, *J. Mol. Biol.,* 268:678 (1997).
23. D. Shalloway, C.A. In Floudas, and P.M. Pardalos, Eds. *Recent Advances in Global Optimization,* Princeton, Vol. 1, pp. 433-648 (1991).
24. J. Kostrowicki and H.A. Scheraga, *J. Phys. Chem.*, 96:7442 (1992).
25. R.A. Abagyan, *FEBS Letters*, 325:17 (1993).
26. H. Kawai, T. Kikuchi, and Y. Okamoto, *Protein Eng.,* 3:85 (1989).
27. S.R. Wilson and W. Cui, *Biopolymers*, 29:225 (1990).
28. J.A. McCammon, B.R. Gelin, and M. Karplus, *Nature*, 267:585 (1977).
29. M. Levitt and R. Sharon, *Proc. Nat. Acad. Sci., U.S.A.*, 85:7557 (1988).
30. R.E. Bruccoleri and M.A. Karplus, *Biopolymers*, 29:1847 (1990).
31. M.J.D. Powell, *Math. Programming*, 12:241 (1997).
32. T. Noguti and N. Go, *Biopolymers*, 24:527 (1985).
33. A. Kidera, *Proc. Natl. Acad. Sci. U.S.A.*, 92:9886 (1995).
34. J.K. Shin and M.S. Jhon, *Biopolymers*, 31:177 (1991).
35. C. Lee and S. Subbiah, *J. Mol. Biol.*, 217:373 (1991).
36. R. Unger and J. Moult, *J. Mol. Biol.*, 231:75 (1993).
37. J. Skolnick, A. Kolinski, and R. Yaris, *Biopolymers*, 28:1059 (1989).
38. D.R.Pipoll and H.A. Scheraga, *Biopolymers*, 27:1283 (1988).
39. L. Holm and C. Sander, *J. Mol. Biol.*, 218:183 (1991).
40. H.S. Kang, N.A. Kurochkina, and B. Lee, *J. Mol. Biol.,* 229:448 (1993).
41. B. Lee, N. Kurochkina, and H.S. Kang, *FASEB J.*, 10:119 (1996).
42. Purisima and Scheraga, *J. Mol. Biol.*, 196:697 (1987).
43. C.P. Hill, D.H. Anderson, L., Wesson, W.F. DeGrado, and D. Eisenberg, *Science*, 249:543 (1990).
44. M.S. Briggs and H. Roder, *Proc. Natl. Acad. Sci. U.S.A.*, 89:2017 (1992).
45. J.P. Cox, P.A. Evans, L.C. Packman, D.H. Williams, aznd D.N. Woolfson, *J. Mol. Biol.*, 234:483 (1993).
46. B. Robson and R.H. Pain, *Biochem. J.,* 141:869 (1975).

47. M.J. Rooman, J.P.A. Kocher, and S.J. Wodak, *J. Mol. Biol.,* 221:961 (1991).
48. R. Abagyan and T. Totrov, *J. Mol. Biol.*, 235:983 (1994).
49. J.W. Ponder and F.M. Richards, *J. Mol. Biol.*, 193:775 (1987).
50. J. Janin, S. Wodak, M. Levitt, and B. Maigret, *J. Mol. Biol.*, 125:357 (1978).
51. R.L. Roland and M. Karplus, *J. Mol. Biol.*, 230:543 (1993).
52. M.N. Rosenbluth and A.W. Rosenbluth, *J. Chem. Phys.,* 23:356 (1955).
53. D. Frenkel and B. Smit, *Understanding Molecular Simulation from Algorithms to Applications,* Academic Press, (1996).
54. R.C. Brower, G. Vasmatzis, M. Silverman, and C. DeLisi, *Biopolymers,* 33:329 (1993).
55. M. Ramirez-Alvarado, F.J. Blanco, H. Niemann, and L. Serrano, *J. Mol. Biol.*, 273:898 (1997).
56. M.D. Struthers, J.J. Ottesen, B. Imperiali, *Fold. & Des.*, 3:95 (1998).