New Stochastic Strategy to Analyze Helix Folding

M. A. Moret,*† P. M. Bisch,† K. C. Mundim,‡ and P. G. Pascutti†

*Departamento de Física, Universidade Estadual de Feira de Santana, Feira de Santana, †Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, and ‡Instituto de Química, Universidade de Brasília, Brasília, Brazil

ABSTRACT We propose an alternative stochastic strategy to search secondary structures based on the generalized simulated annealing (GSA) algorithm, by using conformational preferences based on the Ramachandran map. We optimize the search for polypeptide conformational space and apply to peptides considered to be good α-helix promoters above a critical number of residues. Our strategy to obtain conformational energies consist in coupling a classical force field (THOR package) with the GSA procedure, biasing the $\Phi \times \Psi$ backbone angles to the allowed regions in the Ramachandran map. For polyalanines we obtained stable α-helix structures when the number of residues were equal or exceeded 13 amino acids residues. We also observed that the energy gap between the global minimum and the first local minimum tends to increase with the polypeptide size. These conformations were generated by performing 2880 stochastic molecular optimizations with a continuum medium approach. When compared with molecular dynamics or Monte Carlo methods, GSA can be considered the fastest.

INTRODUCTION

It is well known that the biological activity depends on the spatial conformation acquired by the macromolecules in the physiological medium. The action of hormones and drugs is also dependent on the molecular three-dimensional structure of the target molecules. In recent years, the atomic description of biological molecules in computational simulations have promoted significant advances in the comprehension of the biological process as well as proposed new insights in the design of molecules to satisfy specific properties.

Biological macromolecules have a large number of degrees of freedom leading to several local minima in the molecular energy hyper-surface. Concerning protein functionality, it is presumed that these molecules express their biological activity when they are close to the global minimum of energy (Yon, 1997). A generalized concern is how to predict the lowest conformational energy through simulation procedures. In that sense, we have developed a computational code to perform stochastic molecular optimization (Moret et al., 1998b) hoping to find the lowest conformational energy.

The available literature on peptide conformational analysis is enormous, since Corey and Pauling (1953) determined the ideal values for all backbone bond lengths and bond angles. The allowed values for the pairs of dihedral angles about the C_α atoms, which are limited by steric constraints, were determined by Ramachandran et al. (1963) and are summarized in the so-called Ramachandran map.

Important advances are found in the theoretical field of chain topologies prediction (Rooman et al., 1991) and also regarding the folding patterns determined by experimental studies (Richards, 1991; Dobson, 1992; Elöve et al., 1992; Haynie and Freire, 1993; Kiefhaber et al., 1992; Radford et al., 1992; Khorazanizadeh et al., 1993; Evans and Radford, 1994; Clarke et al., 1999) as well as by theoretical ones (Dill, 1990, 1993; Shakhnovich, 1994; Sali et al., 1994; Wolynes et al., 1995; Bryngelson et al., 1995; Pande et al., 1998; Hiltpold et al., 2000).

Helices are the most prevalent secondary structural motif observed in proteins with known structure. Several recent theoretical studies of the helix-coil transitions have shown the relation between the unfolded state, in random coil conformations, and the helical conformational states, by using molecular dynamics (Daggett et al., 1991; Tobias and Brooks, 1991; Wang et al., 1995; Brooks, 1996, Doruker and Bahar, 1997; Bertsch et al., 1998, Hiltpold et al., 2000), Monte Carlo (Sung, 1994, 1995; Wu and Wang, 1998), or other theoretical approaches (Jun and Weaver, 2000; Park and Goddard, 2000). Both standard molecular dynamics and Monte Carlo simulation methods require extended computational time to obtain helical secondary structures.

To search peptide conformational space, we proposed a very fast stochastic procedure consisting of a simulated annealing methodology coupled to a biased search of the energy hyper-surface according to the allowed regions in the Ramachandran map. We applied this procedure to investigate the helical propensities of polyalanines as a function of the polypeptide size, in a continuum medium approach with a dielectric constant $\epsilon_r = 2$. To illustrate the method applicability, a more complex system is presented briefly, where we search for the secondary structures in a small protein.

Simulated annealing methods have been applied successfully for the description of a variety of global optimization problems. The power of the simulated annealing methods is due to their suitability for large-scale optimization problems, especially for those in which a desired global mini-

Submitted May 2, 2000, and accepted for publication December 17, 2001.

Address reprint requests to Dr. P. M. Bisch, Instituto de Biofisica, Universidade Federal do Rio de Janeiro, CCS, Bloco G, Ilha do Fundao, 21919-900 Rio de Janeiro, Brazil. Tel.: 55-21-2562-6575; Fax: 55-21-2280-8193; E-mail: pmbisch@biof.ufrj.br.

0006-3495/02/03/1123/10 $2.00

TABLE 1 Seven possible conformations for several amino acid sequences

Φ	Ψ	Conformation
−65	−40	A
−89	−1	C
−117	142	B
−69	140	P
78	20	G
103	−176	E
−83	133	O

mum is hidden among many local minima. The first nontrivial solution in this sense was proposed to solve combinatorial problems (Kirkpatrick et al., 1983). Their algorithm strictly follows the quasi-equilibrium Boltzmann-Gibbs statistics using a Gaussian visiting distribution. Based on generalized thermostatistics (Tsallis, 1988, 1995; Curado and Tsallis, 1991), Tsallis and Stariolo (1996) proposed the generalized simulated annealing (GSA), or Tsallis machine, approach. Tsallis machine was applied in several problems, such as molecular optimization using classical methods (Moret et al., 1998b) or semi-empirical methods (Mundim and Tsallis, 1996), geophysical problems (Mundim et al., 1998), traveling salesman problem (Penna, 1995a), numerical data fitting (Penna, 1995b), and genetic algorithm (Moret et al., 1998a). GSA has been proven to be the most effective simulated annealing method when considered for optimizing combinatorial problems (Penna, 1995a) and seems to be the most effective one in simulated annealing when considered for optimizing problems in real space (Tsallis and Stariolo, 1996). Our strategy to obtain molecular conformational energies is to couple a classical force field with the GSA procedure and use a Ramachandran map as a tendency for the probable dihedral angles.

We considered as allowed regions of the Ramachandran map, those corresponding to the Φ and Ψ angles values proposed in the PRELUDE software package (Rooman et al., 1991). These values were computed from comparative statistics of the backbone secondary structure for several amino acid sequences. Table 1 shows the seven possible conformations proposed by this method to predict main-chain topology. These specific angles describe the average conformation of a wide range of proteins with known backbone topology and were also employed by Dandekar and Argos (1994) to perform a polypeptide conformational search using a genetic algorithm. In our strategy we used these seven values (Table 1) to define the center of each allowed region for the dihedral coordinates. Dandekar and Argos (1994) used an empirical potential based on the hydrophobic contribution of each amino acid. Our procedure uses a classical force field that describes all atoms in the system whereas conformations are obtained by a stochastic method, i.e., GSA. The stochastic procedure is used to scan the peptide main chain oriented by Ramachandran map and also to analyze side-chain conformations.

In developing the computational code, for the GSA biased by the Ramachandran map procedure (GSARM), four aspects were considered to create an efficient calculation system. First, the energy function $E(\varphi)$ is defined in an N-dimension continuous space, where $\varphi \in R^N$; second, transpose easily the conformational barrier to avoid trapping the system in local minima; third, the search is biased considering seven conformational regions that are defined in the Ramachandran map as proposed by Rooman et al. (1991); and, fourth, the method gives a rapid analysis of the energy hyper-surface, increasing information about Ramachandran maps in addition.

Here we present the GSARM procedure used for recovering the global minimum. We then discuss the results obtained for some molecular structures of simple polypeptides (polyalanines), and the conclusions of this research are presented.

METHODS

The THOR program (Arêas et al., 1995; Moret et al., 1998b; Pascutti et al., 1999a,b) was developed to be a comprehensive and a flexible tool to investigate macromolecular structures of biological interest such as proteins and membranes. The computational code is based on a classical force field and considers the GROMOS parameters (van Gusteren and Berendsen, 1987) as well as the corrections from the GROMOS96 version; however, other force fields can be easily implemented. Both molecular dynamics and optimization methods are available in this program. The choice of either of these methods depends on the user needs. For example, if dynamic properties are needed, the software can perform molecular dynamics simulations. To analyze systematically the conformational energy hyper-surface or to map the global and local minima we use stochastic methods.

In the THOR program, the conformational energy of the molecule is made up of a sum of bonded and nonbonded terms (Arêas et al., 1995; Pascutti et al., 1999a,b). In this approach, only hydrogen atoms covalently bonded to oxygen or to nitrogen are considered explicitly, whereas CH_1, CH_2, and CH_3 groups are assumed to be an atomic unit. In our stochastic procedure we use a simplified version of the conformational energy, which maintains fixed bond lengths and bond angles within their ideal values. We focus our search on the dihedral angle space (Moret et al., 1998b). Therefore, we analyze the changes of the following energy function:

$$E = E_\varphi + E_{\mathrm{VdW}} + E_{\mathrm{el}}, \tag{1}$$

where E_φ is the dihedral angle potential, E_{VdW} is the van der Waals potential, and E_{el} is the Coulomb potential term (see definitions and used parameters in Moret et al., 1998b).

The search of local and global minima and the mapping of the energy hyper-surface involves the comparison of the energies of two consecutive random conformations. The molecular geometry at steps $t + 1$ and t, are given by φ_{t+1}, and φ_t, respectively, where φ_t is an N-dimensional vector that contains all dihedral coordinates (N) to be optimized. Then, for two consecutive steps, they are related by

$$\varphi_{t+1} = \varphi_t + \Delta\varphi_t, \tag{2}$$

where $\Delta\varphi_t$ is a random perturbation of the dihedral angles.

This random perturbation $\Delta\varphi_t$ allows visiting the potential hyper-surface. Each point in Table 1 defines a central point of seven allowed square regions. To define the extension of these regions we have initially taken large Φ and Ψ angular variations centered on each point. By testing the

methodology on polyalanines, we have verified that most of the optimized structures present variations of the Φ and Ψ values that did not exceed 3.6° according to the expected central points of Table 1. Therefore, we have considered the conformational search in the GSARM procedure restricted to these seven square areas of side equal to 7.2. To compare the number of cycles necessary for convergence, we have also performed conformational search on the entire $\Phi \times \Psi$ space (full GSA procedure).

To generate the random vector φ_{t+1} we use the visiting distribution function $g(\Delta\varphi_{t,i})$ for each component $\Delta\varphi_{t,i}$ of the perturbation vector $\Delta\varphi_t$, defined as follows:

$$g_{q_v}(\Delta\varphi_{t,i}) = \left(\frac{q_v - 1}{\pi}\right)^{1/2} \frac{\Gamma\left(\frac{1 - \frac{1}{2}(q_v - 1)}{q_v - 1}\right)}{\Gamma\left(\frac{1}{q_v - 1} - \frac{1}{2}\right)} \times \frac{[T_{q_v}(t)]^{1/(3-q_v)}}{\left[1 + (q_v - 1)\frac{(\Delta\varphi_{t,i})^2}{[T_{q_v}(t)]^{2/(3-q_v)}}\right]^{1/(q_v-1)-1/2}}, \quad (3)$$

where q_v is the visiting index, T_{qv} is the visiting temperature, and Γ is the gamma function. To obtain the $\Delta\varphi_t$ vector we took a random value for the random vector $a = \{a_i\}$ $(0 < a_i < 1)$ and performed a numerical integration of the visiting distribution probability:

$$a_i = \int g(\Delta\varphi_{t,i}) d\varphi \quad (4)$$

The integral of $g_{qv}(\Delta\varphi_t)$ has an analytical solution only for $q_v = 1$ or $q_v = 2$. In the GSA package, the integral is calculated by using a series expansion and by taking the inverse function of a polynomial series, whose expansion has a cutoff at the 17th order, as proposed in Moret et al. (1998b).

The following step is to compare the energy of the new conformation related to the old one; if this energy is lower than the previous one, the new conformation is accepted. To test the acceptance of a new conformation with higher energy we use the generalized Tsallis statistic P_{qA}, given by Tsallis and Stariolo (1996):

$$P_{q_A} = [1 - (1 - q_A)[E(\varphi_t) - E(\varphi_{t+1})]/T_{q_A}(t)]^{1/(1-q_A)}, \quad (5)$$

where q_A is acceptance index and T_{qA} is the acceptance temperature. At the limit, where $q_A \to 1$ we recover the Boltzmann-Gibbs statistics used in the conventional Metropolis criterion (Metropolis et al., 1953).

In summary, the algorithm for searching the peptide secondary structure, using GSA or GSARM procedure, is as follows. 1) Fix the parameters (q_A, q_v). Start, at $t = 1$, with arbitrary internal coordinates (arbitrary dihedral angles) and a high enough value T_0 for the initial visiting and acceptance temperatures: $T_{qv}(1)$ and $T_{qA}(1)$. 2) Randomly choose a region of the Ramachandran map for each residue using the visiting distribution probability $g_{qv}(\Delta\varphi_t)$ (Eq. 5). 3) Randomly generate the vector φ_{t+1}, given by the visiting distribution probability $g_{qv}(\Delta\varphi_t)$ (Eq. 5). 4) Calculate the conformational energy $E(\varphi_{t+1})$ by using the THOR program and the acceptance criterion as follows: if $E(\varphi_{t+1}) < E(\varphi_t)$, replace φ_t by φ_{t+1}; if $E(\varphi_{t+1}) \geq E(\varphi_t)$, run a random number $r \in [0, 1]$; if $r > P_{qA}$ (acceptance probability) retain φ_t; otherwise, replace φ_t by φ_{t+1}. 5) Cool the system by decreasing the temperatures $T_{qv}(t)$ and $T_{qA}(t)$, assuming, for the sake of simplicity, that $T_{qv}(t) = T_{qA}(t)$:

$$T_{q_A}(t) = T_{q_v}(t) = T_{q_v}(1)\frac{2^{q_v-1} - 1}{(1 + t)^{q_v-1} - 1}, \quad (6)$$

where t is the discrete computational step. 6) Go back to step 2 until the energy $E_{min}(\varphi)$ is reached, i.e., when a lower energy value is no longer obtained in several steps. A more general test for the convergence is whether the same final conformation is obtained starting from different initial conditions.

For a sufficiently large number of steps, this procedure assures that the system can escape from any local minimum and explore the entire allowed energy hyper-surface. In the following section the applications for the GSA and GSARM approaches is shown.

To study the helix formation and to compare the two methodologies (GSA and GSARM) we performed a search of minima in the energy hyper-surface of the polyalanines with 5–20 alanine residues. To obtain the dihedral minima structures we have used both GSA and GSARM approaches taking as initial conformations the completely extended ((Φ, Ψ) = (180°, 180°)) conformation for all polyalanines. In the calculations we have fixed all bond lengths and bond angles within their ideal values, as proposed by Corey and Pauling (1953). We have performed a set of simulations with different initial parameters, i.e., $T_0 = [1, 2, 5, 10, 50, 100]$ and q_A and q_v values in the interval [1.1, 2.5] with 0.1 as a step. We observed that global minimum is reached in more than 50% of the simulations using $T_0 = 100$, $q_A = 1.1$, and q_v values larger or equal to 2. For each peptide system we performed 180 simulations (90 with GSA and 90 with GSARM). Therefore we performed 2880 simulations (1440 to GSA and 1440 to GSARM). Using this procedure we obtained sets of configurations in local and global minima.

RESULTS AND DISCUSSION

The α-helix is the classic element of protein structure. Pauling et al. (1951) were the first ones to describe the α-helix. They predicted it as a stable and favorable structure in proteins. All the hydrogen bonds of the α-helix backbone are aligned along the helical axis with the same orientation. Because a peptide bond has a dipole moment arising from the different polarity of the NH and CO groups, these dipole moments are also aligned along the helical axis. The overall effect is a significant macro-dipole that has the positive pole at the amino end and the negative pole at the carboxyl end of the α-helix. The overall energy that stabilizes the α-helix came from the attractive contributions due to hydrogen bonds and/or by charge-helix dipole interactions (Shoemaker et al., 1985) as well as from the van der Waals contributions. The α-helix formation is a cooperative process where the electrostatic energy has an important role in stabilizing this type of structure, as was shown by Park and Goddard (2000) through ab initio quantum mechanics calculations. These authors found that extending the length of an α-helix by adding additional residues increasingly favors the α-helix formation. A critical number of amino acids, however, are necessary to stabilize this α-helix structure, and an upper limit may also be imposed by the entropy effect. Isolated α-helix structures, in fact, would have to be longer than 13 residues to be stabilized by the attractive interactions (Shoemaker et al., 1987; Rogers, 1989; Voet and Voet, 1994).

Different amino acids have been found to present weak though definite preference in favor or against being in α-helix structure, and the intrinsic helical propensity of some amino acids has been demonstrated to be position dependent (Petukhov et al., 1999). In this sense, Ala, Glu,

Leu, and Met are considered to be good α-helix promoters whereas Pro, Gly, Tyr, and Ser are considered to be poor ones (Branden and Tooze, 1991). Such preferences were the main considerations in all early attempts to predict secondary structures from amino acid sequences, but they were not strong enough to obtain accurate predictions.

Short polypeptides and individual protein fragments generally do not form helices in water. The first example resulting in a significant α-helix formation in water, near 0°C, was obtained with the C- and S-peptide fragments of ribonuclease A (Kim and Baldwin, 1984). Stable α-helix formation was observed in 16-residue alanine-based peptides and was attributed to the high helix-forming potential of alanines (Marqusee et al., 1989). Other experimental measurements on alanine-based peptides have shown that helix nucleation takes place on the millisecond time scale (Clarke et al., 1999).

Molecular dynamics simulations, considering aqueous or implicit solvent, have also tested the stability of polyalanines. Doruker and Bahar (1997) studied the α-helix stability in homopeptides of 13 amino acid residues, and they proposed a rank for the amino acids involved where Ala $<$ Val $<$ Ser $<$ Gly. They observed that polyalanine unfolds within few hundreds of picoseconds at 350 K. Starting from all-coil conformations, Hiltpold et al. (2000) observed the helical stabilization of alanine-based polypeptides of $\sim$30 residues, in an implicit solvent model, within the first 30 ns, at 360 K, with an average folding time of 10 ns. By means of torsion-coordinates molecular dynamics, a method that eliminates bond and angle degrees of freedom, Bertsch et al. (1998) observed the α-helix formation in a 20-residue alanine peptide in trajectories of 0.5 ns, with a half-life of 210 ps for the helix formation. A shorter α-helix folding time of 100 ps was found by Wu and Wang (1998) for a 16-residue polyalanine, with self-guided molecular dynamics, based on the motion guided by an introduced external force.

Stochastic methods have also been applied in polyalanine helix studies (Sung, 1994; Hoffmann and Knapp, 1996). Sung (1994), with the Metropolis-Monte Carlo method and the solvent-referenced interaction, observed the α-helix formation in a 16-residue alanine peptide, with different initial conditions, after 15×10^6 steps. In this work we applied the stochastic method GSA, and demonstrated that polyalanines are stable in the α-helix structure when the peptide has 13 or more amino acid residues, in a low dielectric constant medium. Furthermore, we show that this stable conformation can be reached in a few thousand steps.

In Tables 2 and 3 are shown the 10 lowest-energy conformations obtained for most of the studied peptides, using both GSARM and GSA procedures. It can be observed that, unless a critical number of residues is attained ($\sim$13 residues), the lowest-energy state obtained corresponds to a nearly random conformation, and only few residues are in the α-helix region whereas the majority of them is out of this region. Furthermore, for peptides with less than 13 residues, the energy gap between the lowest-energy state and any other lower-energy states is of the order of the available thermal energy at room temperature. On the contrary, for peptides with more than 13 residues, the lowest-energy state corresponds to a state where most of the residues are in the α-helix conformation, and the energy gap between the lowest state and the following lower one tends to be very large.

We summarized our results for polyalanines in Fig. 1, showing the minimum energy conformations of some polyalanines with 5 residues up to 20 residues using these methodologies. In this figure, we note that no conformational preference is observed for peptides with less than 13 residues (Fig. 1, *A–F*). Peptides with 13 or more residues tended to be stabilized in an α-helix structure (Fig. 1, *G–J*). We observe in Table 2 that the peptide with 13 alanine residues presented an α-helix structure that corresponds to the lowest-energy minimum. The peptides with 6, 7, 8, 9, and 11 alanine residues presented an α-helix structure corresponding only to its secondary local minimum. On the other hand, in Table 3, all peptides with more than 13 residues presented an α-helix conformation corresponding to the lowest-energy minimum.

In Fig. 2 is shown the Ramachandran map for all minima energy configurations obtained with both GSARM and GSA methods for the pentalanine and for the icoalanine. It was possible to discriminate the most populated regions on the Ramachandran map. We observed that pentalanine (Fig. 2 *A*) populates more the β-region than the α-region. On the contrary, we observed that icoalanine (Fig. 2 *B*) populates more the α-region than the β-region. In Fig. 2, *C–H*, more details of the icoalanine simulations are shown. We noted that N-terminus residues have not been stabilized even around an α-region, in accordance with the literature (Munoz and Serrano, 1995). On the other hand, the presented residues showed an α-helix structure preference.

In Fig. 3 *A* is shown the energy gap between the global (lowest) minimum and the first local minimum for all polyalanines. In this figure we observe that for peptides with a number of residues less than 12 residues the energy gap tends to increase with the enhancement of the peptide size. On the other hand, peptides with 13–16 residues do not present a large energy gap. Peptides with 17 or more residues presented a large energy gap, and therefore, these peptides are very stable in an α-helix conformation. In Fig. 3 *B* is shown the energy gap between the lowest-energy state and the minimum with the largest number of the residues in the α-region. In this figure we also observe that peptides with 13 or more residues tend to have an α-helix conformation.

Although we have used a simple version for the atomic representation of the peptide in a continuum electrostatic medium of dielectric constant $\epsilon_r = 2$, considering the united atom model for aliphatic carbons and the GROMOS force field, the general features, i.e., polyalanine α-helix prefer-

TABLE 2 Conformations obtained with GSA and GSARM

Conformations	ΔE (kcal/mol)
AACBA	0.00
AAAAA	0.03
AAAAP	1.53
ABAAB	1.56
AAAAC	1.66
AAACP	2.11
PAABB	2.84
AABEB	3.83
BAAAA	3.96
BABPB	4.09
BPAACB	0.00
AAAAAA	0.94
AAAAAP	1.56
AAAAAB	1.93
AAAACP	2.51
ACBPAB	2.83
PPGBPB	3.84
AAAAPP	3.95
PBBPPP	4.57
BABBAB	4.97
AAACPP	4.97
ABABAAB	0.00
BACBPPA	3.54
AABPABP	4.92
AAAAAAP	5.12
AAAAAPB	6.82
AAAAAPP	6.84
ABPPPBA	8.44
BAPGAAA	8.80
PAAAAAP	8.86
AAAACBB	9.64
ABAABPBA	0.00
AAAAAAAB	2.53
AAAAAAPP	3.51
AAAAAPPB	5.86
AAAAACBP	6.35
PAAAAAAP	6.70
AAAACPPB	7.26
PAAAAAPP	7.32
AAAAACCP	7.78
PPPPPBPB	8.16
AAAABAAACB	0.00
AAAABACCCB	4.27
BBAPGCPAAA	13.27
AAAAACCCAA	14.91
BPAABPPPPB	15.11
BPACAAAAAA	15.23
BPACBPPPPB	15.36
PAAAACABPP	16.90
AAAAACPCAP	17.70
PPBAAAACAP	17.77
PPBBBPGBBB	17.77
AAAAAAACAAAA	0.00
AACBPBAPAPAA	1.92
AACBPBAPABAA	4.30
BPAAAAAAACAB	6.77
AAAAAACACBAB	8.99
ABBAAAAAAAPP	11.89
PAPAAAAAABAB	12.93
ACPAAACACCBB	16.87
BPPGBPBPPPPA	17.47
BPPGBPBBPPPA	18.09

TABLE 3 Conformations obtained with GSA and GSARM

Conformations	ΔE (kcal/mol)
AAAAAAAAAAACAA	0.00
AAAAAACPAAAAAA	5.93
AAAAAAAAAPPAAA	10.80
AAAAAAAAAPBBPB	11.88
ABABBAAAAAAAAA	14.69
PGACBPAACBBBEB	16.02
AAPBPBAAAAAAAA	17.55
AABAABPPBPBPPB	17.64
PAPAAACAAAABPP	18.64
BPABPAPAAAAAAA	19.40
AAAAAAAAAAAAAPCPB	0.00
ABAAACCAAACAACPBA	17.09
AAAAAABBAACPPAACA	17.21
PBPPAPAAAAACBCAAP	18.76
PPPBPPBAAAACBAAPB	19.14
AAACCACPAAAAPPPAB	20.23
AAAAPPBACPABPPPPB	24.93
BBAABPAAABPBPPPPB	26.39
BBPACBCPPAAACPAAP	26.40
BPBBPBCBAAPPAPPPP	27.14
BAAAAAAAAAAAAAAAAA	0.00
BBAAAAAAAAAGBABAAA	13.25
BBAAAAEAAAAEBABAAAA	15.63
AAAAACAAAAAAACPBCCA	32.47
PAAAAACAAPPAAAAAACP	34.39
PBCPAPAAAAAAAAAPBPP	38.94
AACCAAAAAAACAPBACBB	40.82
BBPPPPAAAAAAAAAPBPP	41.10
PPPBCACABAAAAAAAAPP	42.16
AABAAGAAGAEAAAAAAE	46.52

ence above a critical residue number of ~13 amino acids, the energy gap from the α-helix to other conformations, and the gap enhancement showing the crescent stabilization as the number of residues increases, are in good accordance with experiments (Shoemaker et al., 1987; Marqusee et al., 1989; Clarke et al., 1999) and previous theoretical predictions (Rogers, 1989; Sung, 1994; Hoffmann and Knapp, 1996; Doruker and Bahar, 1997; Bertsch et al., 1998; Wu and Wang, 1998; Park and Goddard, 2000; Hiltpold et al., 2000).

To compare both approaches, GSARM and GSA, we have analyzed the number of cycles necessary to have convergence as a function of the peptide size (Fig. 4). It is shown in Fig. 4 that the GSARM approach that took less than 1000 steps to converge is a fast method to determine peptide conformations. All energy minima obtained by both GSA and GSARM procedures for the studied polyalanines were recorded, and both procedures lead to the same results, although differences were observed in the number of interactions and consequently the computational time expended in each simulation.

The time step in molecular dynamics simulations is of the order of 0.0005–0.0020 ps. When considering the latter, 50,000 steps are necessary for a 100-ps trajectory and millions of steps to simulate events at a nanosecond time scale. So far, it was reported in the literature that

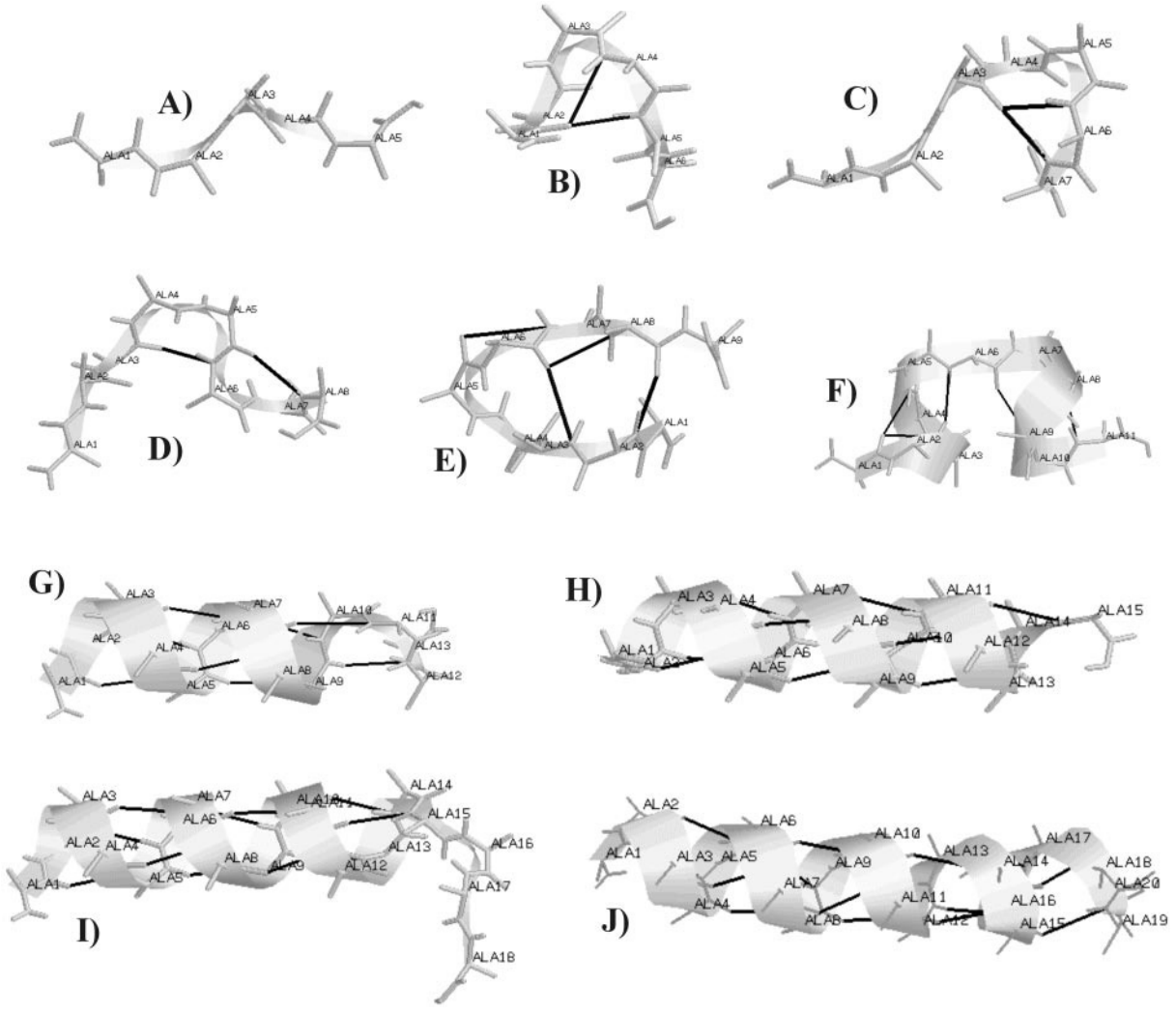

FIGURE 1 The structures correspond to the lowest minimum of the polyalanines with 5 (*A*), 6 (*B*), 7 (*C*), 8 (*D*), 9 (*E*), 11 (*F*), 13 (*G*), 15 (*H*), 18 (*I*), and 20 (*J*) residues. It is clearly shown that polyalanines with 13 or more residues have α-helix conformations.

simulations of the α-helix folding for a polyalanine had taken 0.1–30 ns (Wu and Wang, 1998; Bertsch et al., 1998; Hiltpold et al., 2000). Usual stochastic methods, such as Metropolis-Monte Carlo, can take millions of steps to simulate the α-helix polyalanine folding (Sung, 1994). Applying the GSARM procedure, peptides as long as 20 monomers initially in an extended conformation can adopt an α-helical structure in less than 1000 steps. The fact that the global energy minimum is attained in so few steps is probably due to the fractal structure of peptide and protein energy hyper-surfaces, as we have recently suggested (Moret et al., 2001).

To test our procedure on protein fragments we studied the secondary structures of the insect defensin A (Protein Data Bank code 1ICA, Cornet et al., 1995). It is a basic 4-kDa protein that in vivo is excreted in the hemolymph of the flesh fly *Phormia terramovae* larvae in response to bacterial challenge or tissue injury (Lambert et al., 1989) and is principally active against Gram-positive bacteria. Insect defensin A presents one α-helix and two β-strands stabilized by three disulfide bridges. The energy hyper-surface of the α-helix and the two β-sheet structures were analyzed, by means of the GSA procedure, randomly searching all the main chain Φ and Ψ angles and all of the side-chain dihedral angles of these structures starting from the native conformation. The remaining residues, including the main-chain and lateral groups, were kept at the native positions.

The results for the α-helix domain (HIS13-ARG23) showed that the lowest energy for the unfolded state at a random coil conformation is $\sim$10 kcal/mol higher than the conformational energy of the α-helix global minimum. However, when we performed the search in the region of β-sheet (ARG26-ARG39), we found that the lowest-energy conformation corresponds to a random coil, which is only 0.8 kcal/mol less than the corresponding β-sheet conformational energy.

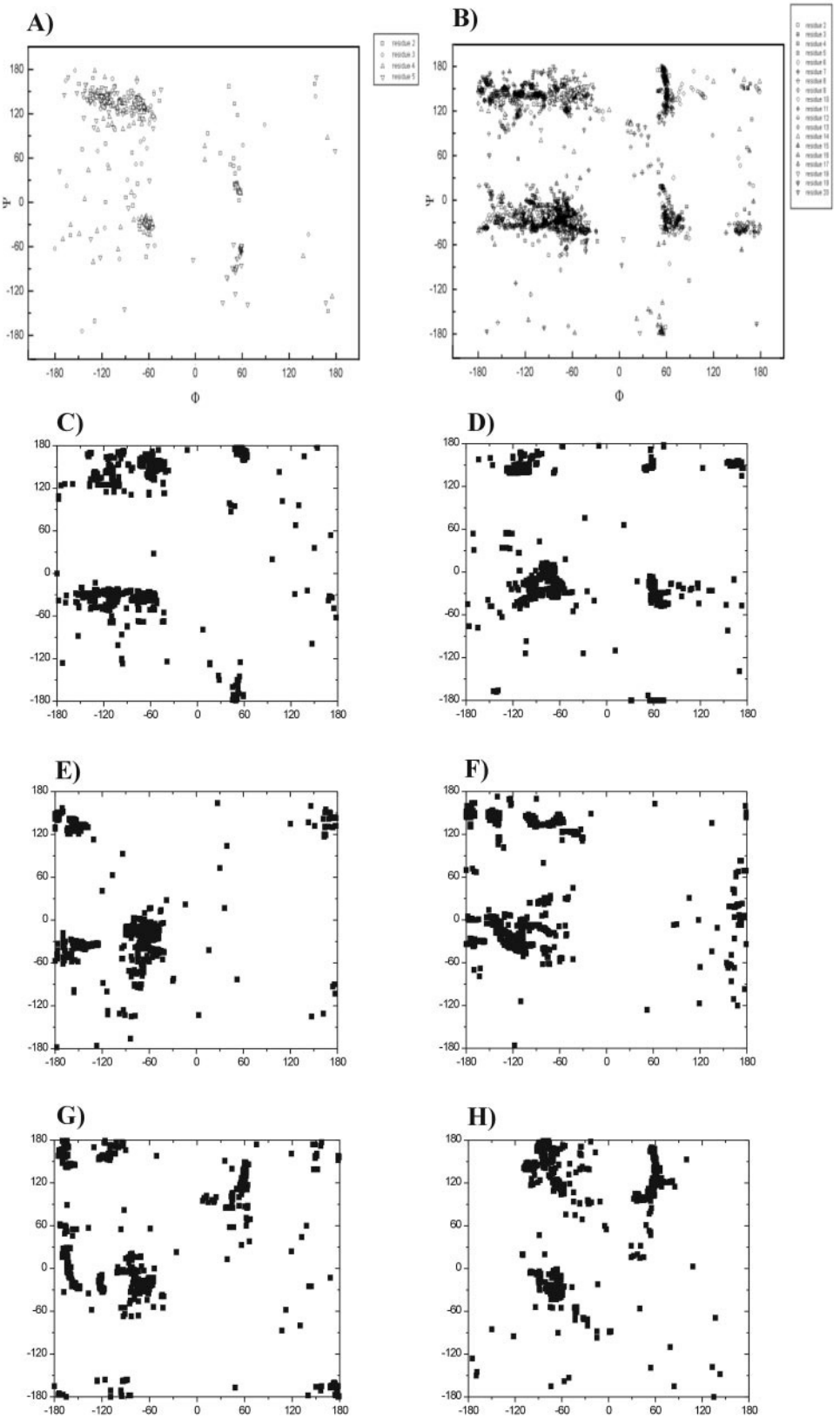

FIGURE 2 (*A* and *B*) The Φ and Ψ coordinates of each residue of the minimum energy structures found by using both GSA and GSARM procedures, for pentalanine (*A*) and icoalanine (*B*). (*C–H*) Ramachandran plot of some residues of the icoalanine are shown in detail: N-terminus (*C*), ALA3 (*D*), ALA8 (*E*), ALA12 (*F*), ALA17 (*G*), and C-terminus (*H*).

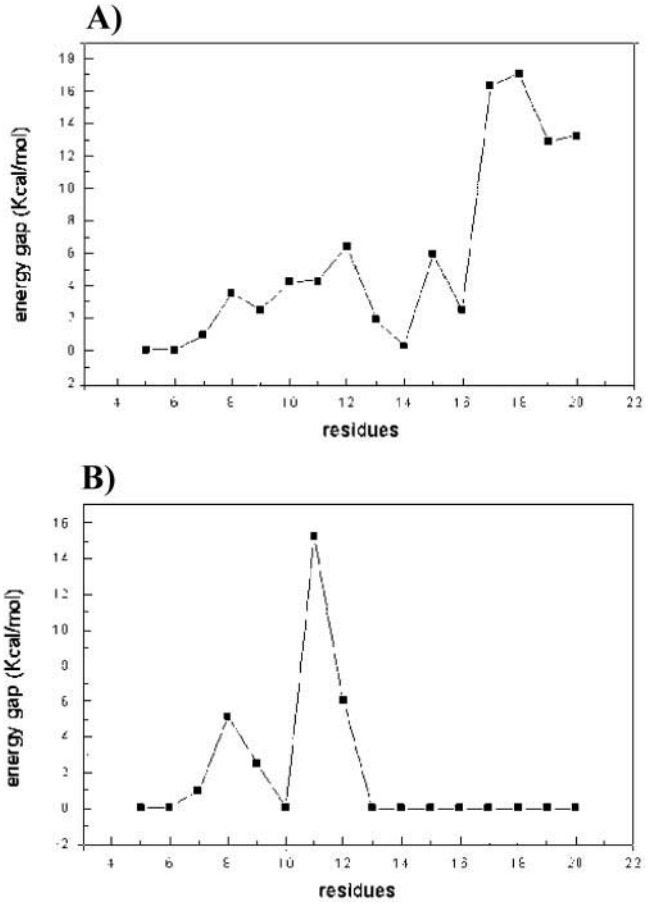

FIGURE 3 (*A*) Energy gap between the first local minimum and the global minimum for all polyalanines; (*B*) Energy gap between the structures with a largest number of residues in the α-region and the lowest minimum.

CONCLUSIONS

The GSA method was used to explore the energy hypersurface of peptides. This method, based on generalized thermostatistics, is a fast stochastic approach used to determine the allowed global and local energy minima. We have applied this strategy to predict secondary structures of polyalanines. An α-helix conformation was found as corresponding to the lowest conformational energy for peptides of 13 or more residues, as suggested in the literature. When the search of the peptide backbone conformations is performed with the GSARM approach, where certain values of dihedral angles Φ and Ψ in the Ramachandran map are more frequently visited, the simulation generally expends less than 1000 steps to find the global minimum energy for a 5–20-residue polyalanine. Compared with other theoretical approaches based on molecular dynamics or Monte Carlo methodologies, GSARM can be considered as the fastest strategy.

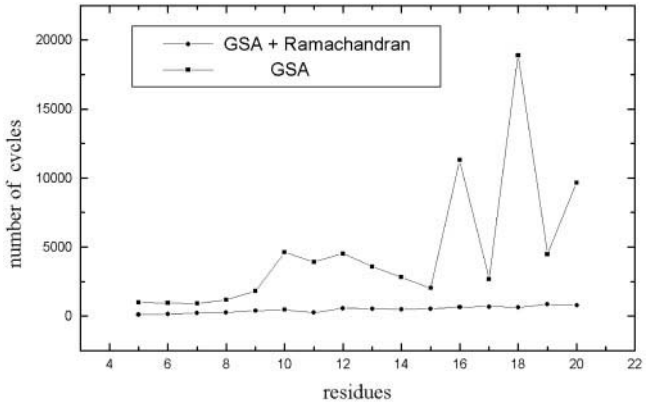

FIGURE 4 The lowest number of cycles necessary to obtain a conformation of minimum energy as function of polypeptide size.

The electrostatic interactions between adjacent dipoles in the peptide backbone are repulsive, so a critical number of residues is needed to be counterbalanced by the attractive interactions, such as H-bonds. Then, short peptides of less than seven residues are flexible, and they show no conformational preference. When the number of residues of the chain increases (8–13 residues) the structures tend to collapse into H-bonded turns, as shown in Fig. 1. Above a critical number of 13 amino acid residues the enhancement of the hydrogen bond number stabilizes the polyalanine in an α-helix structure. In fact, for long-chain peptides, most of the possible H-bonds of the backbone tend to be formed. Then, an energy gap arises from the lowest-energy conformation to the next low-energy conformation. This energy gap enlarges with the number of residues.

The helical conformations presented in Fig. 1 were also obtained when we performed a restricted search to one quadrant of the Ramachandran map. In this case, the backbone angles Φ and Ψ were restrained to the third quadrant, i.e., $-180^{o} < \Phi < 0^{o}$ and $-180^{o} < \Psi < 0^{o}$, as proposed by Wang et al. (1995) and Tobias and Brooks (1991) to analyze helical propensities.

Finally, we performed a search considering a large value for the dielectric constant ($\epsilon_r = 80$) to obtain the lowest-energy minimum for two systems: a 13-alanine peptide and a 14-alanine one. According our simulations these peptides tend to adopt values for their Φ and Ψ dihedral angles corresponding to the α-helix region in the Ramachandran map (i.e., in the third quadrant), but the H-bond formation is poorly maintained. This example as well as the simulations of the defensin show the importance of having a good description of the solvent effect, mainly when we are interested in searching for the protein conformations in its appropriated environment. This environment changes from the low dielectric constant, typical of the protein interior, to the higher one, for residues in contact with the solvent.

According to the Ramachandran map, proteins have forbidden values for the Φ-Ψ pairs of dihedral angles. Performing the search for the peptide backbone conformations, exclusively in the allowed regions, the number of steps to

achieved GSA convergence decreases drastically. Both approaches, the GSA and the GSARM, have been shown to be applicable to helix folding studies; however, the method proposed here is sufficiently general to be extended to analyze any conformational state of polypeptides and proteins in general.

The partial financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Fundaćão de Amparo a Pesquisa do Estado do Rio de Janeiro, and Fundação Universitária José Bonifácio is acknowledged.

REFERENCES

Arêas, E. P. G., P. G. Pascutti, S. Schreier, K. C. Mundim, and P. M. Bisch. 1995. Molecular dynamics simulations of signal sequences at a membrane/water interface. *J. Phys. Chem.* 99:14882–14892.

Branden, C., and J. Tooze. 1991. Introduction to protein structure. Garland Publishing, New York.

Bertsch, R. A., N. Vaidehi, S. I. Chan, and W. A. Goddard. 1998. Kinetic steps for alpha-helix formation. *Protein Struct. Funct. Genet.* 33: 343–357.

Brooks, C. L. 1996. Helix-coil kinetics: folding time scales for helical peptides from a sequential model. *J. Phys. Chem.* 100:2546–2549.

Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Protein Struct. Funct. Genet.* 21:167–195.

Clarke, D. T., A. J. Doig, B. J. Stapley, and G. R. Jones. 1999. The alpha-helix folds on the millisecond time scale. *Proc. Natl. Acad. Sci. U.S.A.* 96:7232–7237.

Corey, R. B., and L. Pauling. 1953. Fundamental dimensions of polypeptide chains. *Proc. R. Soc.* 141:10.

Cornet, B., J.-M. Bonmatin, C. Hetru, J. A. Hoffmann, M. Ptat, and F. Vovelle. 1995. Refined three-dimensional solution structure of insect defensin A. *Structure.* 3:435–448.

Curado, E. M. F., C. and Tsallis. 1991. Generalized statistical mechanics: connection with thermodynamics. *J. Phys. A.* 24:L69 (Corrigenda: 1991, *J. Phys. A* 24:3187, and 1992, *J. Phys. A* 25:1019).

Daggett, V., P. A. Kollman, and I. D. Kuntz. 1991. A molecular-dynamics simulation of polyalanine: an analysis of equilibrium motions and helix coil transitions. *Biopolymers.* 31:1115–1134.

Dandekar, T., and P. Argos. 1994. Folding the main chain of small proteins with genetic algorithm. *J. Mol. Biol.* 236:844–861.

Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29: 7133–7155.

Dill, K. A. 1993. Folding proteins: finding a needle in a haystack. *Curr. Opin. Struct. Biol.* 3:99–103.

Dobson, C. M. 1992. Unfolded proteins, compact states and molten globules. *Curr. Opin. Struct. Biol.* 2:6–12.

Doruker, P., and I. Bahar. 1997. Role of water on unfolding kinetics of helical peptides studied by molecular dynamics simulations. *Biophys. J.* 72:2445–2456.

Elöve, G. A., A. F. Chaffotte, H. Roder, and M. E. Goldberg. 1992. Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy. *Biochemistry.* 31:6876–6883.

Evans, P. A., and S. E. Radford. 1994. Probing the structure of folding intermediates *Curr. Opin. Struct. Biol.* 4:100–106.

Haynie, D. T., and E. Freire. 1993. Structural energetics of the molten globule. *Proteins.* 16:115–140.

Hiltpold, A., P. Ferrara, J. Gsponer, and A. Caflisch. 2000. Free energy surface of the helical peptide Y(MEARA)(6) *J. Phys. Chem. B.* 104: 10080–10086.

Hoffmann, D., and E. W. Knapp. 1996. Polypeptide folding with off-lattice Monte Carlo dynamics: the method. *Eur. Biophys. J. Biophys. Lett.* 24:387–403.

Khorazanizadeh, S., I. D. Evans, T. R. Butt, and H. Roder. 1993. Folding and stability of a tryptophan containing mutant of ubiquitin. *Biochemistry.* 32:7054–7063.

Kiefhaber, T., F. X. Schmid, K. Willaert, Y. Engelborghs, and A. Chaffotte. 1992. Structure of a rapidly formed intermediate in ribonuclease T1 folding. *Protein Sci.* 1:1162–1172.

Kim, P. S., and R. L. Baldwin. 1984. A helix stop signal in the isolated S-peptide of ribonuclease A. *Nature.* 307:329–334.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science.* 220:671–686.

Jun, B., and D. L. Weaver. 2000. Diffusion-controlled kinetics of the helix-coil transition with square barrier hydrogen bonds. *J. Chem. Phys.* 112:4394–4401.

Lambert, J., E. Keppi, J.-L. Dimarq, C. Wicker, J.-M. Reichhart, B. Dunbar, P. Lepage, A. Van Dorsselaer, J. Hoffmann, J. Fothrgill, and D. Hoffmann. 1989. Insect immunity: isolation from immune blood of the depteran *Phormia terranovae* of two insect antibacterial peptides with sequence homology to rabbit lung macrophage bactericidal peptides. *Proc. Natl. Acad. Sci. U.S.A.* 86:262–266.

Marqusee, S., V. H. Robins, and R. L. Baldwin. 1989. Unusually stable helix formation in short alanine based peptides. *Proc. Natl. Acad. Sci. U.S.A.* 86:5286–5290.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. J. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.

Moret, M. A., P. M. Bisch, and F. M. C. Vieira. 1998a. Algorithm for multiple minima search. *Phys. Rev. E.* 57:R2535–R2538.

Moret, M. A., P. G. Pascutti, P. M. Bisch, and K. C. Mundim. 1998b. Stochastic molecular optimization using generalized simulated annealing. *J. Comp. Chem.* 19:647–657.

Moret, M. A., P. G. Pascutti, K. C. Mundim, P. M. Bisch, and E. Nogueira, Jr. 2001. Multifractality, Levinthal paradox and energy hyper-surface, *Phys. Rev. E.* 63:20901–20904.

Mundim, K. C., T. Lemaire, and A. Bassrei. 1998. Optimization of nonlinear gravity models through generalized simulated annealing. *Physica A.* 252:405–416.

Mundim, K. C., and C. Tsallis. 1996. Geometry optimization and conformational analysis through generalized simulated annealing. *Int. J. Quantum Chem.* 58:373–381.

Munoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* 245:275–296.

Pande, V. S., A. Y. Grosberg, T. Tanaka, and D. S. Rokhsar. 1998. Pathways for protein folding: is a new view needed? *Curr. Opin. Struct. Biol.* 8:68–79.

Park, C., and W. A. Goddard. 2000. Stabilization of alpha-helices by dipole-dipole interactions within alpha-helices. *J. Phys. Chem. B.* 104: 7784–7789.

Pascutti, P. G., L. J. El-Jaik, K. C. Mundim, A. S. Ito, and P. M. Bisch. 1999a. Molecular dynamics simulation of α-melanocyte stimulating hormone in a water-membrane model interfaces. *Eur. Biophys. J.* 28: 499–509.

Pascutti, P. G., K. C. Mundim, A. S. Ito, and P. M. Bisch. 1999b. Polarization effect on peptide conformations at water-membrane interface by molecular dynamics simulation. *J. Comp. Chem.* 20:971–982.

Pauling, L., R. B. Corey, and H. R. Branson. 1951. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* 37:205–211.

Penna, T. J. P. 1995a. Traveling salesman problem and Tsallis statistics, *Phys. Rev. E.* 51:R1–R2.

Penna, T. J. P. 1995b. Fitting Curves by Simulated Annealing. *Comp. Phys.* 9:341–343.

Petukhov, M., K. Uegaki, N. Yumoto, S. Yoshikawa, and L. Serrano. 1999. Position dependence of amino acid intrinsic helical propensities. II.

Non-charged polar residues: Ser, Thr, Asn, and Gln. *Protein Sci.* 8:2144–2150.

Radford, S. E., C. M. Dobson, and P. A. Evans. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature.* 358:302–307.

Ramachandran, G. N., C. Ramakrishnan, and V. Sasisekharan. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7:95.

Richards, F. M. 1991. The protein folding problem. *Sci. Am.* 264:34–41.

Rogers, N. C. 1989. The role of electrostatic interactions in the structure of the globular proteins. *In* Prediction of Protein Structure and Principles of Protein Conformations. Gerald D. Fasman, editor. Plenum Press, New York.

Rooman, M. J., J.-P. Kocher, and S. J. Wodak. 1991. Prediction of protein backbone conformation based on seven structure assignments: influence of local interactions. *J. Mol. Biol.* 221:961–970.

Sali, A., E. I. Shakhnovich, and M. Karplus. 1994. How does a protein fold? *Nature.* 369:248–251.

Shakhnovich, E. I. 1994. Proteins with selected sequence fold into unique native conformation. *Phys. Rev. Lett.* 24:3907–3910.

Shoemaker, K. R., P. S. Kim, D. N. Brems, S. Marqusee, E. J. York, I. M. Chaiken, J. M. Stewart, and R. L. Baldwin. 1985. Nature of the charged-group effect on the stability of the C-peptide helix. *Proc. Natl. Acad. Sci. U.S.A.* 82:2349–2353.

Shoemaker, K. R., P. S. Kim, E. J. York, J. M. Stewart, and R. L. Baldwin. 1987. Tests of the helix dipole model for stabilization of α-helices. *Nature.* 326:563–567.

Sung, S.-S. 1994. Helix folding simulations with various initial conditions. *Biophys. J.* 66:1796–1803.

Sung, S.-S. 1995. Folding simulation of alanine-based peptides with lysine residues. *Biophys. J.* 68:826–834.

Tobias, D. J., and C. L. Brooks. 1991. Thermodynamics and mechanism of α helix initiation in alanine and valine peptides. *Biochemistry.* 30: 6059–6070.

Tsallis, C. 1988. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* 52:479.

Tsallis, C. 1995. Comment on "Thermodynamic stability conditions for the Tsallis and Renie entropies". *Phys. Lett. A.* 206:389–396.

Tsallis, C., and D. Stariolo. 1996. Generalized simulated annealing. *Phys. A.* 233:395–404.

van Gusteren, W. V., and H. J. C. Berendsen. 1987. Groningen Molecular Simulation (GROMOS) Library Manual. Biomos, Groningen, The Netherlands.

Voet, D., and J. G. Voet. 1994. Biochemistry. John Wiley and Sons, New York.

Wang, L., T. O'Connel, A. Tropsha, and J. Hermans. 1995. Thermodynamic parameters for the helix-coil transition of oligopeptides: molecular dynamics simulation with the peptide growth method. *Proc. Natl. Acad. Sci. U.S.A.* 92:10924–10928.

Wolynes, P. G., J. N. Onuchic, and D. Thirumalai. 1995. Navigating the folding routes. *Science.* 267:1619–1620.

Wu, X. W., and S. M. Wang. 1998. Self-guided molecular dynamics simulation for efficient conformational search. *J Phys. Chem. B.* 102: 7238–7250.

Yon, J. M. 1997. Protein folding: concepts and perspectives. *Cell. Mol. Life Sci.* 53:557–567.