Generalized Simulated Annealing Applied to Protein Folding Studies

FLAVIA P. AGOSTINI,[1] DIOGO DE O. SOARES-PINTO,[2] MARCELO A. MORET,[3] CARLA OSTHOFF,[1] PEDRO G. PASCUTTI[4]

[1]*Laboratório Nacional de Computação Científica—LNCC, Petrópolis, Rio de Janeiro, Brazil*
[2]*Centro Brasileiro de Pesquisas Físicas, CBPF, Rio de Janeiro, Brazil*
[3]*CEPPEV, Fundação Visconde de Cairú, Salvador, Bahia, Brazil, and Departamento de Física,UEFS, Feira de Santana, Bahia, Brazil*
[4]*Instituto de Biofísica Carlos Chagas Filho, IBCCF/UFRJ, Rio de Janeiro, Brazil*

Received 16 June 2005; Accepted 19 February 2006
DOI 10.1002/jcc.20428
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: During the last few years, computational simulations based on the atomic description of biological molecules have resulted in significant advances in the comprehension of biological processes. It is well known, however, that a molecular system may have a great number of conformations due to the large number of rotation of degrees of freedom around chemical bonds, leading to several local minima on the energy hypersurface. It has been proposed though, that proteins express their biological function when their structure is close to a conformation with energy global minimum. To help solve the protein-folding problem, we use a new strategy based on Simulated Annealing methods. These methods have been well suited for a large extent of optimization problems, especially those containing many local minima. In fact, this work applies the Generalized Simulated Annealing method (GSA) coupled to the GROMOS96 Molecular Force Field to research the minimum energy conformation of 18-alanine. We show that the q_T GSA parameter can be used to control the freezing process during the annealing procedure, and to avoid polypeptide chains to be trapped in energy local minima. We scanned the q-values for visiting (q_V) and accepting (q_A) functions for q_T values ranging from 1 to 3, and found the best values to obtain an α-helix conformation for the polyalanine peptide, which is the conformation with energy global minimum. Global optimization methods also exemplify a class of applications that requires a large amount of computational resources, being suitable for Grid computing. To implement a Grid computing platform, we developed and tested a Grid environment based on MYGRID middleware, which is a technology that can employ all machines accessed by the user to run the application.

 J Comput Chem 27: 1142–1155, 2006

Key words: protein folding; global optimization; generalized simulated annealing; Tsallis' statistics; Grid-based platform

Introduction

The protein-folding problem is an important research area, as the three-dimensional structure determined by the sequence of amino acids dictates the protein functions. Through this study we intend to contribute to advances in this area.

For the purpose of this work, we assume that protein folding is under thermodynamic control.[1] However, protein degrees of freedom limit their active conformation prediction through conventional molecular optimization methods when an extended or random conformation is taken as the initial state. To solve this kind of problem, we use the Generalized Simulated Annealing method (GSA), introduced by Tsallis and Stariolo,[2,3] to extract information about the energy conformations. Simulated Annealing methods have been successfully applied in a great variety of optimization problems, especially those involving large-scale problems containing many local minima.[4]

Simulated Annealing methods are based on the metallurgy process, where a molten metal is gradually cooled until it reaches a crystalline structure—its global minimum, in the same way an artificial temperature is introduced in the optimization method, which is gradually cooled, being extremely convenient for eventually detrapping from local minima and reaching the global minimum in the end of the process. The GSA procedure is a generalization of the classical[4] and the fast simulated annealing,[5] based on the Generalized Statistical Mechanics, and it can be even faster than both methods, depending on

Correspondence to: F. P. Agostini; e-mail: flavia@lncc.br
Contract/grant sponsor: FAPERJ, CAPES and CNPq

the right choice of parameters. The Generalized Simulated Annealing (GSA),[2,3] has been applied to a great variety of problems such as genetic algorithm,[6] molecule optimization using classical methods,[7] or semiempirical methods,[8] geophysical problem,[9] material science,[10–13] Traveling Salesman Problem,[14] and numerical data fitting.[15] Although GSA is a general procedure, its efficiency depends not only on the right choice of the parameters involved but also on the application of a general strategy.

To investigate the GSA procedure parameters, we used an 18-alanine small peptide with a known helical structure, low dielectric environment, and initial random conformation. According to the q_V parameter visiting the distribution function, new conformations are obtained by random displacement of torsion angles. Each new conformation can be accepted or not according to q_A parameter for acceptance probability. Recent work presents a successful investigation for polyalanine helical propensities, where a 3D structure is obtained through only a few steps compared with conventional methods.[7] But there, a limited q_V and q_A parameter set was explored, so that the annealing procedure was interrupted when the structure got trapped in an energy local minimum in most of the conformational search.

In this work we systematically scanned q_V versus q_A values and found that it is necessary to introduce a new parameter to specifically control the temperature decreasing. In the original GSA procedure the cooling function is also defined by the q_V parameter, giving only a moderate efficiency for the GSA procedure when it is applied to protein folding studies. In this work we introduced a q_T parameter for the cooling function, decoupling it from q_V, to better control the temperature decreasing. As a result, we found the energy global minimum, and the α-helix structure for polyalanine, by using several q_T, q_V, and q_A values. A great number of helix folded structures are found for q_T values ranging from 1.7 to 2.5, whereas the best q_V values range from 1.1 to 1.9, and q_V decreases while q_T increases for folded structures.

To make possible the global optimization of a protein, a Grid environment has been developed for the execution of folding calculations through the middleware MYGRID.[16] MYGRID is a user- centric Grid middleware that provides a global execution environment, allowing the remote execution of a large amount of parallel tasks through the machines to which the user has access.

In a previous work,[17] we evaluated GSA algorithm best parameters for speedup scenarios including a Grid composed of a 20-node PC cluster, and a 40 heterogeneous nondedicated node Grid system, using MYGRID's scheduling algorithm solution (WQR), which sends replication task to idle machines to execute a replica of an unfinished task. When a task is replicated, the first replica that finishes is considered as the valid execution of the task and the other replicas are cancelled. Therefore, overloaded or slow machines do not slow down the average execution time for the total task. Based on previous work results, we conclude that the MYGRID test bed effectively extends local computational capabilities, and that it is an important tool to further improve the new method strategy based on the Simulated Annealing.

For future experiments we have plans to apply GSA to the protein folding of larger peptides. In that case, the number of possible conformations increases dramatically, therefore increasing each task computational time.

Figure 1. Sequence of steps for 18-alanine folding in ribbon representation.

In the next section we explain the GSA algorithm used for the investigation of 18-alanine structure best parameters associated to a Grid infrastructure. Then we present the results obtained in this work and concluding remarks.

Materials and Methods

We use the THOR program to obtain conformational structures with the GSA method.[18] The GSA procedure is a powerful tool, with the advantage that it can reach the peptide conformation with the energy global minimum in much fewer steps than conventional stochastic methods. THOR is based on the GROMOS96 classical force field and considers explicitly only hydrogen atoms covalently bonded to oxygen or nitrogen, and not CH_1, CH_2, and CH_3 groups, which are assumed to be an atomic unit.

For the energy conformation of the molecule we consider the torsional term and the nonbonded terms representing van der Waals and electrostatic interactions, as most of the variation in structure and relative energies is due to the complex interplay between torsional and nonbonded contributions. Therefore, we have the following energy function:

$$E(r_i,\varphi_n) = \sum_{n=1}^{N_\varphi} K_{\varphi n}(1 + cos(k_n\varphi_n - \delta_n)) + \sum_{i=1i<j}^{N_{i+1}} \left\{ \frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \right\}.$$

In our simulations we consider a continuum medium approach with a dielectric constant $\varepsilon_r = 2$. Consequently, electrostatic interactions with the solvent are minimized, enhancing protein interactions, and thus, accelerating the folding process. Moret et al.[18] compares simulations with $\varepsilon_r = 2$ and $\varepsilon_r = 80$.

The search for the conformation with the energy global minimum involves comparison of the energies of two consecutive random conformations. From a given conformation, $\{\varphi_t\}$, a new coordinate set, $\{\varphi_{t+1}\}$, of torsion angles is obtained through a random displacement, $\{\varphi_{t+1} = \varphi_t + \Delta\varphi_t\}$, where $\Delta\varphi_t$ is computed from the visiting distribution function:

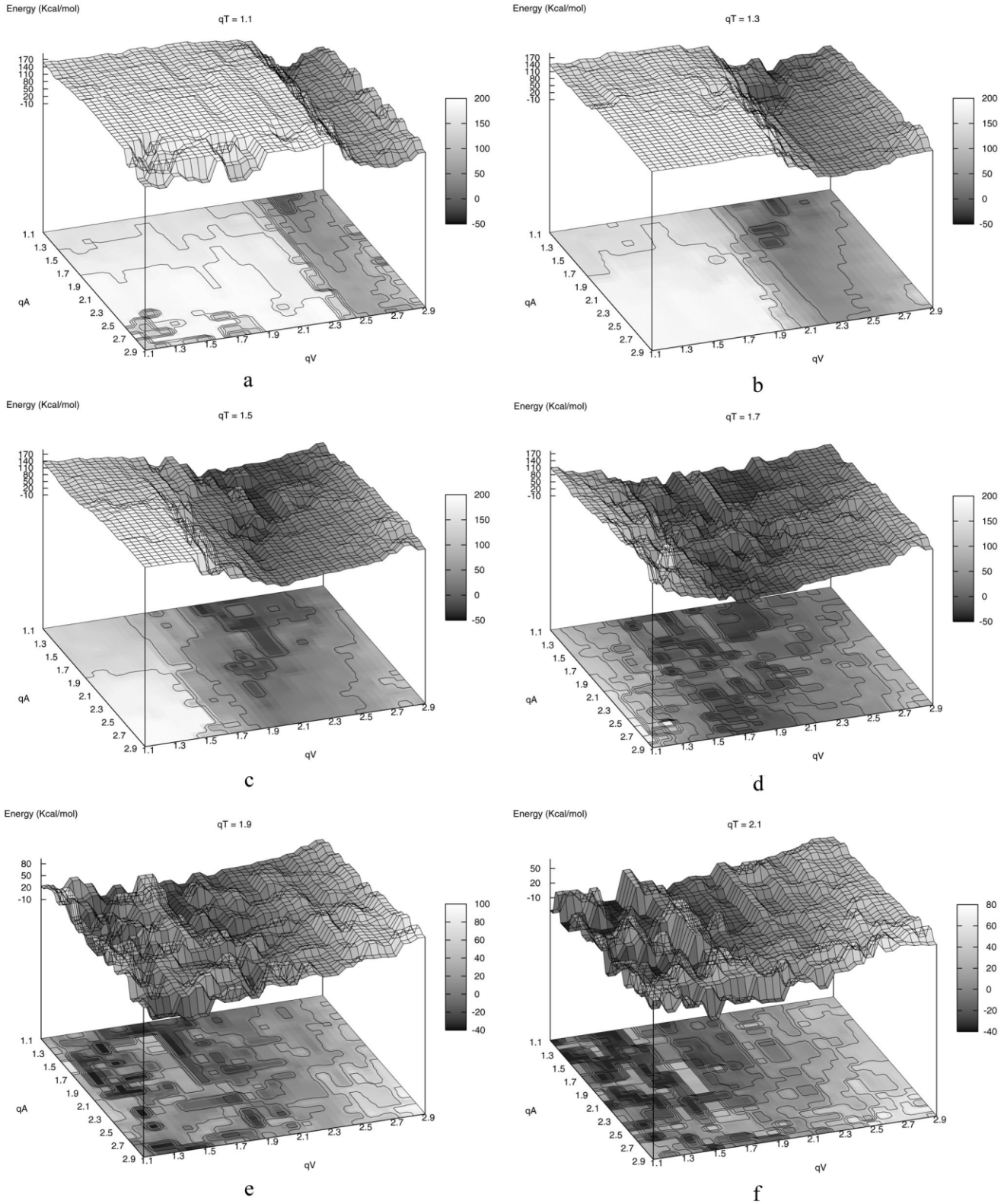

Figure 2. (a–j) Minimum energy diagram of q_A vs. q_V for $q_T = 1.1$ to 2.9.

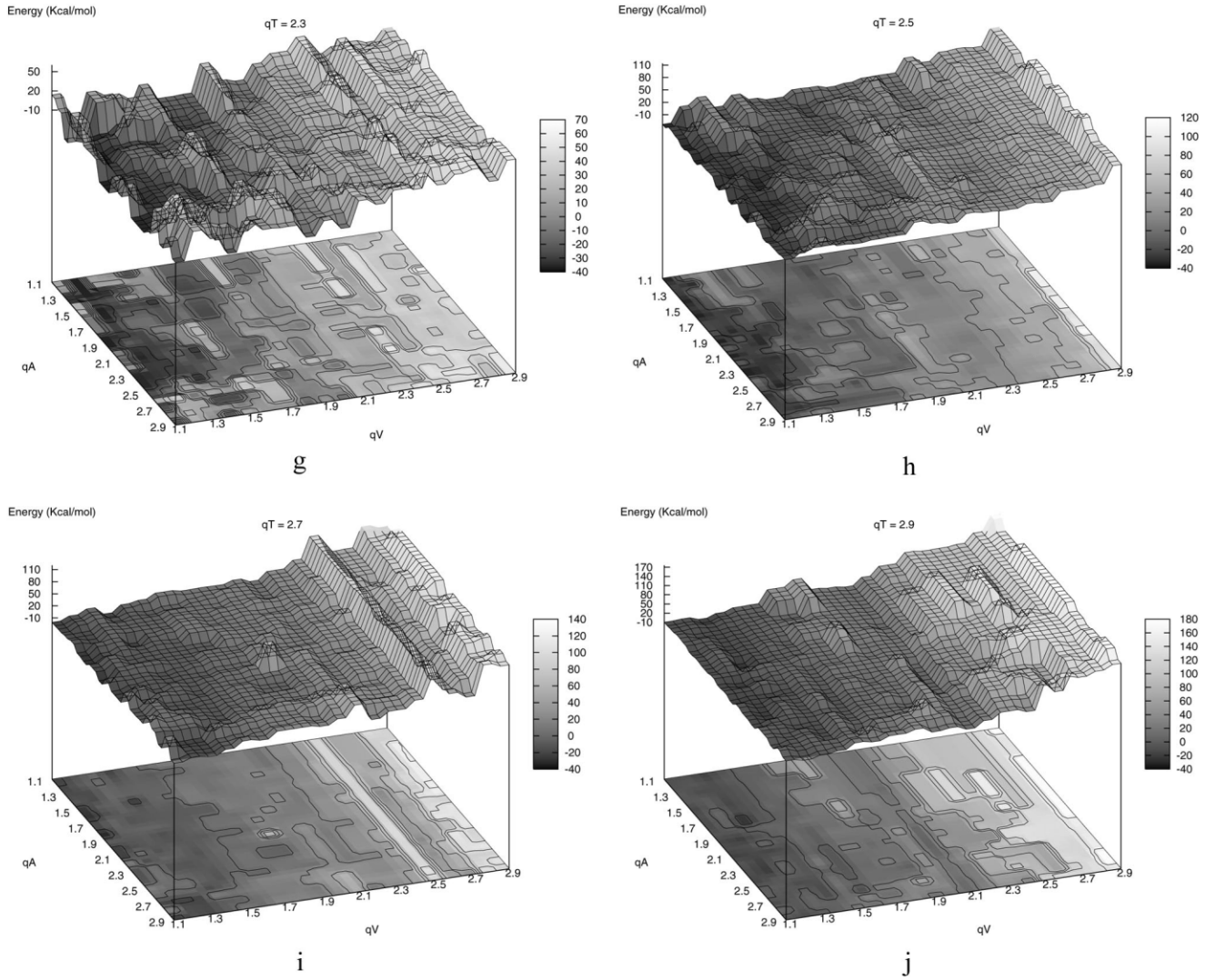

Figure 2. (*continued*)

$$g_{qv}(\Delta\varphi_{1,i}) = \left(\frac{q_V - 1}{\pi}\right)^{\frac{1}{2}} \frac{\Gamma\left(\dfrac{1 - 1/2(q_V - 1)}{q_V - 1}\right)}{\Gamma\left(\dfrac{1}{q_V - 1} - \dfrac{1}{2}\right)} \frac{(T_{qT}(t))^{1/3 - q_V}}{\left\{1 + (q_V - 1)\dfrac{(\Delta\varphi_{1,i})^2}{(T_{q_T}(t))^{\frac{2}{(3-q_V)}}}\right\}^{(1/(q_V-1))-(1/2}}.$$

Note from the equation above that it is necessary to invert this function to obtain the value of $\Delta\varphi_{t,i}$. $\Delta\varphi_{t,i}$ computed this way is not limited. Thus, we use a simplification on this equation, introduced by Dall'Igna et al.,[19] where $\Delta\varphi_{t,i}$ is now given by this own function, and not by its inverse. We consider

$$\Delta\varphi_{i,t} \propto g_{qv}(r_i),$$

with r_i randomly chosen, greatly reducing the computational cost. This simplification also implies in $\Delta\varphi_{t,i}$ limited in an infinite domain.

The jumps in the structure can be accepted or not according to the acceptance probability:

$$P_{qA} = \left[1 + (q_A - 1)\frac{E(\varphi_{t+1}) - E(\varphi_t)}{T_{qT}(\mathrm{t})}\right]^{1/q_A - 1}.$$

If $E(\varphi_{t+1}) < \mathrm{E}(\varphi_t)$, φ_t is replaced by φ_{t+1}; else, if $E(\varphi_{t+1}) \geq E(\varphi_t)$, a random number r &insin; [0,1] is generated and, if $r > Pq_A$, φ_t is retained; otherwise, φ_t is replaced by φ_{t+1}.

On the expression written above, $Tq_T(t)$ is the visiting temperature, which is given by:

$$T_{qT}(t) = T_{qT}(1)\frac{2^{q_T - 1} - 1}{(1 + t)^{q_T - 1} - 1}$$

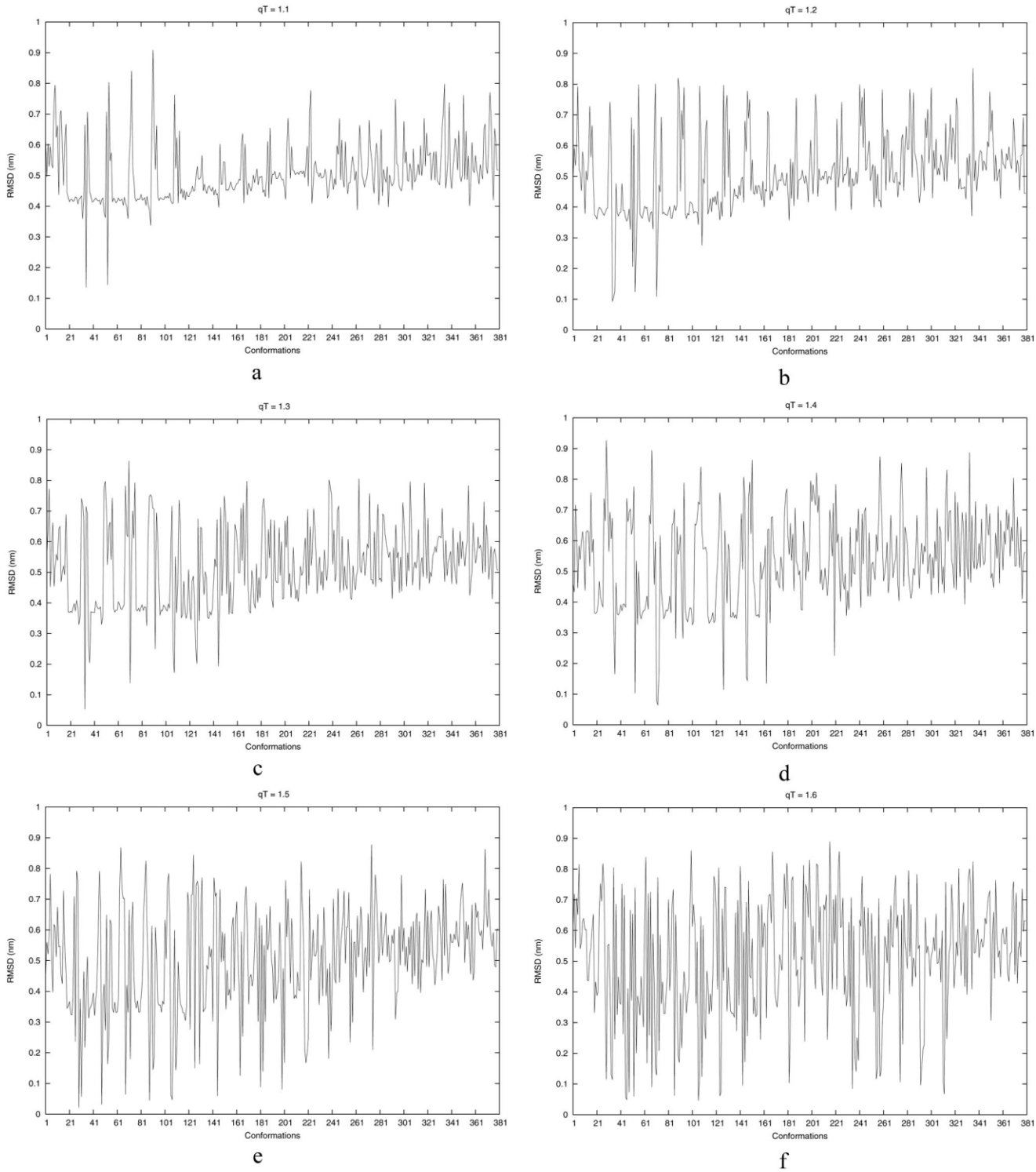

Figure 3. (a–t) RMSD diagram for $q_T = 1.1$ to 3.0 using α-helix confomration for 18-alanine as a reference.

At each step, t, a new temperature is calculated. It is expected that, with the right set of parameters, the temperature decreases as quickly as possible, but still assuring that the global minimum is reached and guaranteeing that no trapping in a local minimum occurs. For the 18-alanine simulations, an initial temperature $Tq_T(1) = 1.10$ was used.

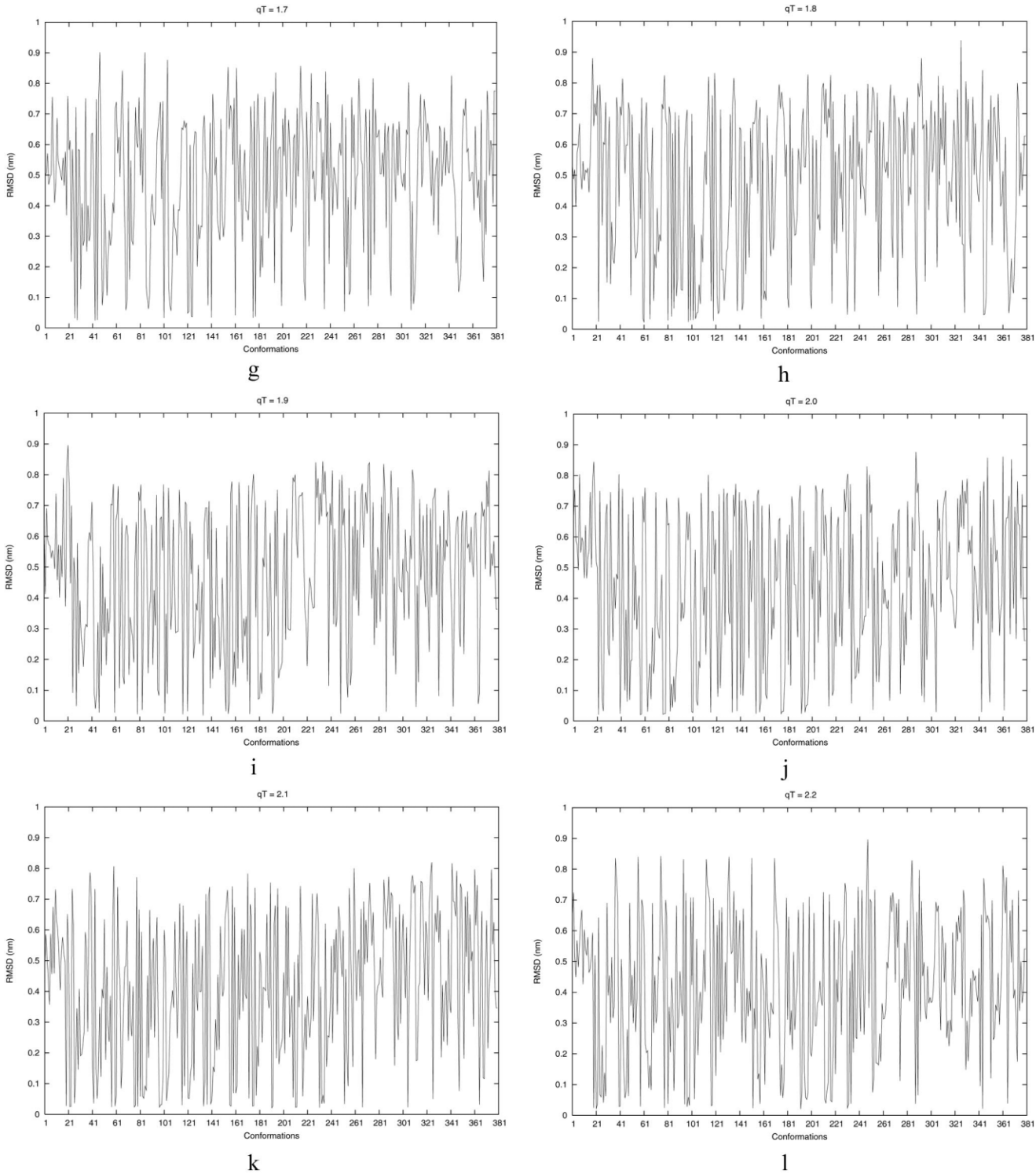

Figure 3. (*continued*)

Here, the q_T parameter for visiting temperature is considered independent from q_A or q_V, differently from the Tsallis and Stariolo[3] proposal, which considers $q_T = q_V$. The q_T definition was included in an attempt to slowdown the temperature decay and, as a result, to visit other areas of the energy hypersurface, and to avoid structural trapping in local minima. It was an important step

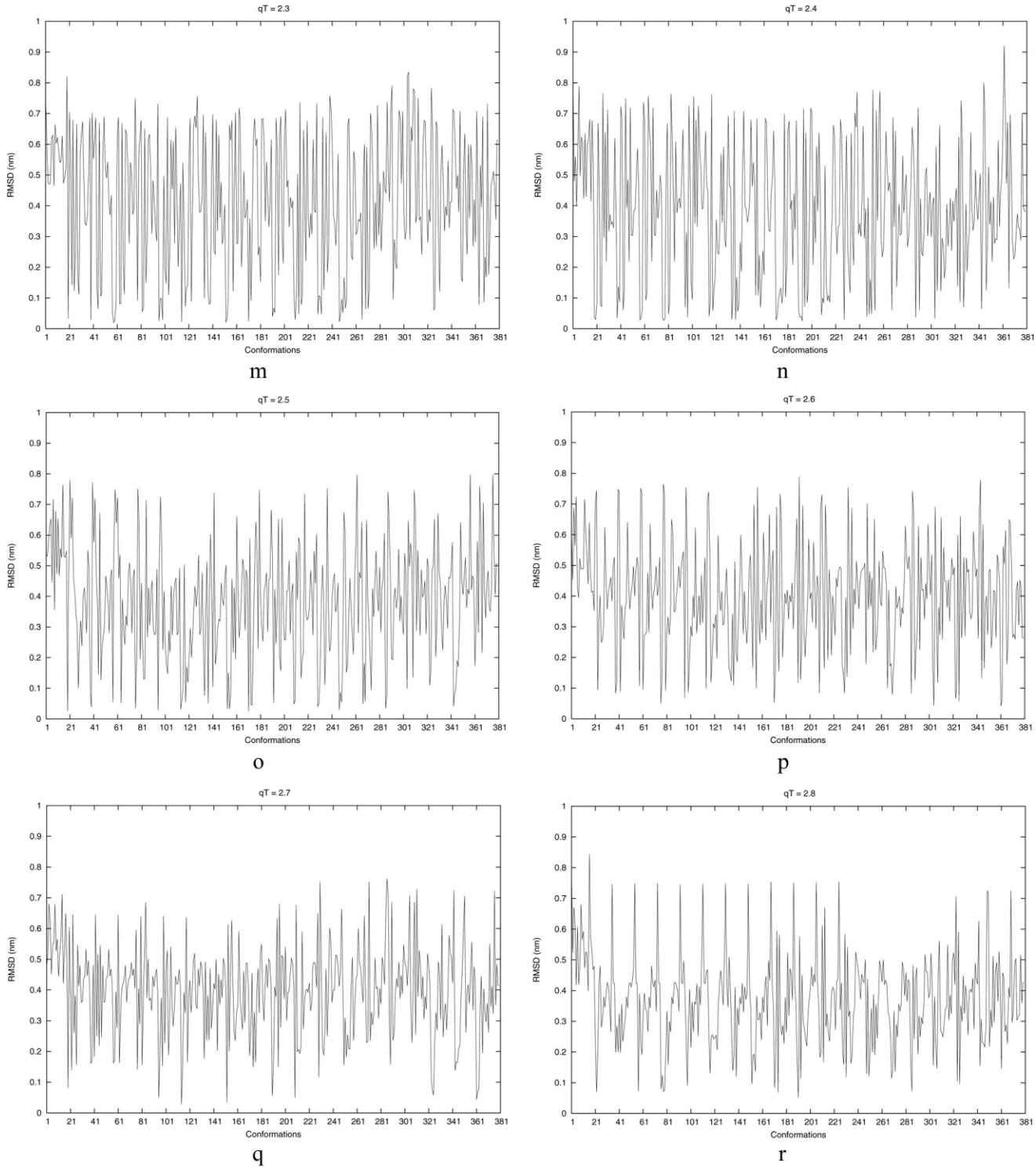

Figure 3. (*continued*)

for the solution of problems introduced by the temperature dependence on q_A or q_V parameters.

The cooling is adjusted by the parameter q_T, while the visiting distribution function and the acceptance probability function are adjusted by the parameters q_V and q_A, respectively. The efficiency of the GSA procedure consists on the right choice of these parameters. Hence, we could explore the q_V, q_A, and q_T values independently.

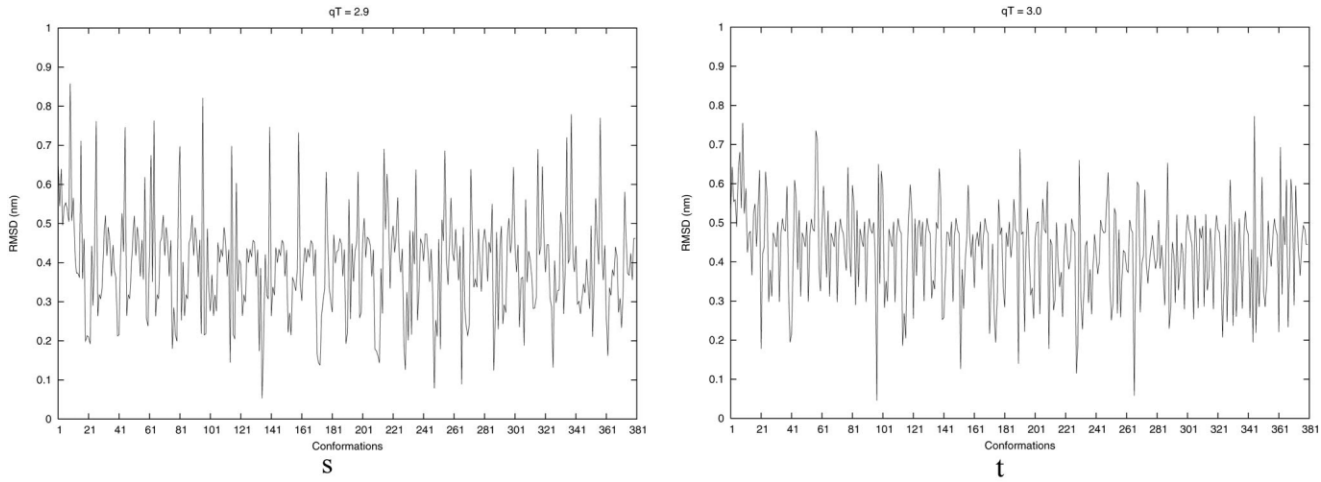

Figure 3. (*continued*)

Moret et al.[7] presents a detailed analysis about GSA implementation on THOR program. GSA simulations are an example of a class of applications that requires a large amount of computational resources, being suitable for a Grid computing environment. The aim of Grid computing projects is to aggregate multiple geographically dispersed machines as a distributed computing platform. In particular, we use MYGRID, which is a Grid middleware that supports the execution of Bag-of-Tasks applications and provides a global execution environment that allows the remote execution of several tasks in parallel through the machines accessed by the user. MYGRID is an open source software and is available at http://dsc.ufcg.edu.br/ourgrid.

In this work the evaluation of the best parameters for GSA algorithm is a search through the q_A, q_V, and also q_T parameter space for the 18-alanine folding. For each parameter search the other two parameters were kept fixed. The q_T search ranges from 1.1 up to 3.0, while q_A and q_V parameters range from 1.1 up to 2.9. For variations on the three parameters we used 0.1 steps, and also used 500,000 annealing steps for each parameter set values.

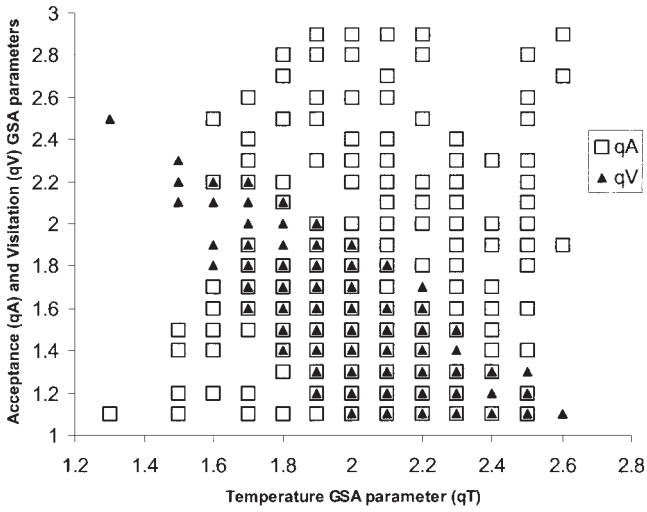

Figure 4. GAS parameters temperature (q_T) vs. visiting (q_V) and acceptance (q_A) from Table 1.

Results

As shown in recent work,[17] 18-alanine is stabilized in the α-helical structure. Figure 1 presents a folding sequence calculated by GSA steps for the 18-alanine, using proper q_V, q_A, and q_T parameters, starting from a random conformation. These selected intermediate conformations represent a possible folding pathway. Through GSA results, not only the final configuration can be achieved, but also the intermediate results can be analyzed to better understand the protein folding problem. The same result can be attained starting from an extended *all-trans* conformation.

The temperature decays very slowly for $q_T < 1.5$, such that much more steps would be necessary for reaching a conformation with global minimum of energy. On the other hand, the temperature parameter causes a very fast cooling when q_T presents the greater values ($q_T > 2.5$). In this case, no helix was obtained because all structures froze in an energy local minimum. The effect of temperature parameter is shown in Figures 2. For the parameter value $q_T = 1.1$ (Fig. 2a), an energy valley can be observed on the $q_V = 2.4$ position for almost all the q_A values. This energy valley is shifted to the lower q_V value positions when q_T is increased (Fig. 2b–e), but disappear when q_T is greater than 2.0 (Fig. 2f–j). Although the energy valley is no more observed for large q_T values, the energy minima stay around the lowest q_V values in most cases (Fig. 2f–i).

To check if an α-helix structure was obtained we used the root-mean-square deviation (RMSD) calculation taking an α-helix structure for the 18-alanine polymer as the reference. Until $q_T = 1.4$, a small number of structures presented RMSD values below 0.1 nanometer (Fig. 3a–d), the RMSD value where both structures appear as an α-helix. By increasing the q_T value a growing number

Table 1. Data from Figure 3 Showing All q_T, q_A, and q_V for α-Helix Folded Structures, Which Correspond to Energies Below −30 kcal/mol and RMSD Below 0.1 Nanometers.

q_T	q_A	q_V	Energy (Kcal/mol)	RMSD (nm)
1.20	1.10	2.60	−27.403294	0.0935
1.30	1.10	2.50	−30.545525	0.0528
1.40	1.30	2.50	−20.455280	0.0772
	1.30	2.60	−11.282080	0.0650
1.50	1.10	2.10	−35.927271	0.0214
	1.10	2.30	−28.973676	0.0569
	1.20	2.10	−34.244301	0.0313
	1.30	2.20	−29.298610	0.0639
	1.40	2.30	−31.827246	0.0447
	1.50	2.20	−32.403345	0.0625
	1.50	2.30	−31.018296	0.0470
	1.70	2.30	−22.978290	0.0592
	1.90	2.10	−25.698092	0.0879
	2.00	2.00	−27.298236	0.0805
1.60	1.20	1.80	−33.387568	0.0564
	1.20	1.90	−32.779888	0.0495
	1.20	2.10	−31.141793	0.0731
	1.20	2.50	−9.171311	0.0598
	1.30	2.10	−29.633462	0.0901
	1.40	2.10	−31.535804	0.0623
	1.50	2.20	−34.512387	0.0463
	1.60	2.10	−33.768345	0.0612
	1.60	2.20	−29.935944	0.0758
	1.70	2.10	−32.585868	0.0955
	2.20	1.80	−31.928983	0.0847
	2.50	1.80	−30.028761	0.0964
	2.60	1.90	−20.177728	0.0667
1.70	1.10	1.80	−35.030410	0.0317
	1.10	2.00	−34.604101	0.0260
	1.20	1.60	−36.088330	0.0246
	1.20	1.80	−36.595675	0.0262
	1.20	2.20	−25.630487	0.0746
	1.30	2.30	−28.260824	0.0580
	1.30	2.40	−11.654982	0.0932
	1.40	2.20	−20.623416	0.0954
	1.40	2.30	−24.619951	0.0628
	1.50	1.70	−36.351983	0.0326
	1.50	2.20	−27.424968	0.0714
	1.50	2.30	−28.684049	0.0575
	1.60	1.80	−34.701465	0.0483
	1.60	1.90	−29.634565	0.0545
	1.60	2.10	−34.265215	0.0382
	1.60	2.20	−32.935950	0.0369
	1.70	1.60	−33.061342	0.0741
	1.70	1.90	−34.912081	0.0341
	1.80	2.00	−31.972583	0.0412
	1.90	1.60	−36.332007	0.0332
	1.90	1.80	−35.086590	0.0372
	2.00	2.10	−27.632767	0.0728
	2.10	2.20	−20.598078	0.0896
	2.20	1.90	−32.528401	0.0625
	2.30	1.70	−33.583372	0.0542
	2.40	1.80	−29.688532	0.0712
	2.60	1.60	−29.783785	0.0583
	2.60	1.80	−30.592459	0.0791
1.80	1.10	1.50	−36.462916	0.0247
	1.30	1.40	−37.027149	0.0346
	1.30	1.50	−36.914265	0.0239
	1.30	2.00	−32.790466	0.0326
	1.30	2.30	−22.182473	0.0927
	1.40	1.60	−36.876153	0.0287
	1.40	1.90	−32.555359	0.0417
	1.40	2.10	−28.295366	0.0681
	1.50	1.40	−35.760352	0.0239
	1.50	1.60	−36.090209	0.0296
	1.50	1.80	−35.163509	0.0303
	1.50	2.00	−33.601068	0.0341
	1.50	2.10	−26.889086	0.0525
	1.50	2.20	−27.896810	0.0559
	1.50	2.40	1.065660	0.0818
	1.60	1.40	−29.837758	0.0922
	1.60	1.60	−36.540339	0.0275
	1.60	2.00	−30.885742	0.0515
	1.60	2.10	−30.542225	0.0618
	1.60	2.50	−13.470402	0.0932
	1.70	1.70	−33.282143	0.0593
	1.70	2.10	−29.839996	0.0621
	1.70	2.20	−27.821046	0.0866
	1.80	1.80	−34.399104	0.0353
	1.80	2.00	−27.022705	0.0973
	1.80	2.20	−22.069021	0.0935
	1.90	2.20	−12.382545	0.0699
	2.00	2.20	−22.518684	0.0671
	2.10	2.00	−29.140648	0.0838
	2.20	1.40	−36.145481	0.0478
	2.20	2.00	−30.325134	0.0581
	2.40	1.80	−27.418600	0.0728
	2.50	1.50	−34.649064	0.0484
	2.70	1.70	−29.446080	0.0536
	2.80	1.40	−28.179898	0.0465
	2.80	1.50	−29.445125	0.0505
	2.90	1.60	−27.278163	0.0526
1.90	1.10	1.70	−29.161014	0.0914
	1.10	2.00	−25.701628	0.0488
	1.20	1.60	−29.873226	0.0868
	1.20	1.70	−35.776345	0.0402
	1.20	1.80	−27.594612	0.0980
	1.20	2.00	−33.019655	0.0281
	1.30	1.40	−36.884396	0.0275
	1.40	1.40	−36.739778	0.0222
	1.40	1.80	−34.976582	0.0361
	1.50	1.30	−32.252569	0.0829
	1.50	1.80	−34.648597	0.0258
	1.50	2.00	−27.143191	0.0885
	1.60	1.40	−36.823459	0.0219
	1.60	1.80	−35.240852	0.0318
	1.70	1.20	−36.654687	0.0184
	1.80	1.20	−36.080846	0.0319
	1.80	1.40	−37.134105	0.0236
	1.80	1.50	−33.663737	0.0623

(*continued*)

Table 1. (*Continued*)

q_T	q_A	q_V	Energy (Kcal/mol)	RMSD (nm)
	1.90	1.30	−36.674633	0.0230
	1.90	2.00	−28.603883	0.0713
	1.90	2.10	−27.208227	0.0739
	1.90	2.30	−12.770251	0.0892
	2.00	1.30	−36.210786	0.0243
	2.00	1.40	−36.796224	0.0685
	2.30	1.30	−36.685573	0.0247
	2.30	1.90	−25.739300	0.0770
	2.50	1.30	−36.678486	0.0309
	2.60	1.90	−31.126293	0.0456
	2.70	1.90	−26.197779	0.0815
	2.80	1.20	−36.375594	0.0472
	2.90	1.40	−36.017745	0.0555
	2.90	1.50	−30.530045	0.0886
2.00	1.10	1.40	−36.056806	0.0213
	1.10	1.70	−28.091684	0.0661
	1.10	1.80	−34.312607	0.0326
	1.20	1.30	−36.689212	0.0243
	1.20	1.70	−28.153496	0.0998
	1.20	1.90	−31.018395	0.0642
	1.20	2.10	−22.450598	0.0691
	1.30	1.10	−36.918744	0.0211
	1.30	1.20	−36.671681	0.0237
	1.30	1.70	−32.371404	0.0311
	1.30	2.00	−20.609890	0.0714
	1.40	1.10	−37.198550	0.0222
	1.40	1.20	−36.923769	0.0241
	1.40	1.30	−36.891893	0.0243
	1.40	1.70	−34.892810	0.0315
	1.40	1.90	−29.565370	0.0447
	1.40	2.10	−20.570211	0.0617
	1.50	1.60	−35.661943	0.0315
	1.50	1.70	−34.466648	0.0279
	1.50	2.00	−19.026906	0.0696
	1.50	2.10	−14.560176	0.0514
	1.60	1.30	−36.584020	0.0284
	1.60	1.90	−24.691902	0.0793
	1.70	1.20	−36.855056	0.0285
	1.70	1.60	−35.680179	0.0326
	1.70	1.90	−23.318290	0.0866
	1.70	2.10	−15.546049	0.0834
	1.80	1.30	−36.817121	0.0240
	1.80	1.60	−35.370945	0.0293
	1.80	1.70	−30.901819	0.0660
	1.80	2.10	−19.464141	0.0877
	1.80	2.20	−9.687189	0.0735
	1.90	1.50	−36.373118	0.0239
	1.90	1.60	−36.321682	0.0319
	1.90	1.70	−35.287163	0.0319
	2.00	1.30	−36.478337	0.0240
	2.00	1.40	−30.147959	0.0896
	2.00	1.60	−36.190772	0.0291
	2.00	1.70	−32.446499	0.0516
	2.00	1.80	−27.369519	0.0531
	2.10	1.70	−29.308220	0.0486
	2.10	2.00	−27.578448	0.0833

(*continued*)

Table 1. (*Continued*)

q_T	q_A	q_V	Energy (Kcal/mol)	RMSD (nm)
	2.20	1.70	−29.490609	0.0578
	2.30	1.60	−33.563454	0.0369
	2.40	1.10	−33.420810	0.0667
	2.40	2.00	−19.939207	0.0891
	2.50	1.90	−26.162438	0.0825
	2.50	2.10	−18.787248	0.0620
	2.60	1.20	−36.771537	0.0298
	2.80	1.20	−32.340709	0.0301
	2.80	1.90	−26.069839	0.0616
	2.90	1.20	−32.828994	0.0353
2.10	1.10	1.10	−36.112986	0.0275
	1.10	1.40	−34.752383	0.0250
	1.10	1.50	−35.732432	0.0259
	1.10	1.80	−31.996634	0.0363
	1.20	1.50	−35.982292	0.0356
	1.20	1.80	−23.676700	0.0512
	1.30	1.10	−37.069748	0.0240
	1.30	1.40	−36.085180	0.0271
	1.30	1.50	−31.619557	0.0650
	1.30	1.90	−16.073689	0.0851
	1.40	1.10	−36.390710	0.0238
	1.40	1.20	−36.031714	0.0293
	1.40	1.40	−33.189406	0.0260
	1.40	1.60	−32.363624	0.0583
	1.40	1.80	−26.287530	0.0584
	1.40	1.90	−23.202985	0.0522
	1.40	2.00	−13.091722	0.0923
	1.40	2.10	−24.751764	0.0788
	1.50	1.30	−35.557893	0.0233
	1.50	1.40	−34.900132	0.0306
	1.50	1.50	−29.343230	0.0324
	1.50	1.90	−27.940834	0.0447
	1.50	2.00	−14.579892	0.0926
	1.60	1.30	−35.865587	0.0306
	1.60	1.50	−29.244945	0.0924
	1.60	1.80	−28.685495	0.0531
	1.60	1.90	−23.109307	0.0517
	1.70	1.20	−36.486494	0.0267
	1.70	1.50	−36.112284	0.0325
	1.70	1.80	−30.344005	0.0341
	1.70	1.90	−26.866662	0.0467
	1.80	1.50	−27.574836	0.0300
	1.80	1.60	−33.183982	0.0271
	1.80	1.80	−17.897885	0.0687
	1.80	2.00	−27.091897	0.0688
	1.80	2.10	−18.244301	0.0941
	1.90	1.20	−33.792988	0.0522
	1.90	1.50	−35.716396	0.0271
	1.90	1.60	−35.291426	0.0285
	1.90	1.80	−23.647232	0.0611
	2.00	1.20	−36.294439	0.0212
	2.00	1.30	−36.511501	0.0265
	2.10	1.10	−33.864210	0.0556
	2.10	1.30	−36.007052	0.0231
	2.10	1.40	−35.843906	0.0238
	2.10	1.90	−20.508499	0.0588

(*continued*)

Table 1. (*Continued*)

q_T	q_A	q_V	Energy (Kcal/mol)	RMSD (nm)
	2.20	1.40	−35.187088	0.0221
	2.20	1.60	−33.601710	0.0348
	2.20	1.70	−27.353634	0.0617
	2.20	1.80	−27.156265	0.0359
	2.30	1.60	−34.165655	0.0339
	2.30	1.90	−19.849015	0.0922
	2.40	1.20	−36.550572	0.0231
	2.60	1.20	−36.862286	0.0232
	2.70	1.40	−31.932809	0.0504
	2.80	1.70	−24.296987	0.0764
	2.90	1.50	−35.602816	0.0322
2.20	1.10	1.10	−37.115869	0.0243
	1.10	1.30	−35.984220	0.0222
	1.10	1.40	−34.501209	0.0350
	1.10	1.60	−27.007434	0.0679
	1.10	1.70	−23.581820	0.0599
	1.10	1.90	−26.098663	0.0415
	1.10	2.10	−11.955363	0.0626
	1.20	1.30	−35.536764	0.0290
	1.20	1.40	−34.081161	0.0295
	1.20	1.80	−28.039345	0.0560
	1.20	1.90	−22.929531	0.0777
	1.20	2.10	−11.955249	0.0624
	1.30	1.20	−35.965024	0.0288
	1.30	1.90	−14.358269	0.0887
	1.30	2.10	−20.422970	0.0828
	1.40	1.30	−36.094697	0.0244
	1.40	1.60	−34.302358	0.0373
	1.40	2.10	−22.761477	0.0806
	1.50	1.10	−34.103286	0.0364
	1.50	1.30	−36.097400	0.0282
	1.50	1.40	−31.717641	0.0596
	1.50	1.50	−32.831473	0.0327
	1.50	1.90	−12.702578	0.0527
	1.60	1.40	−35.469400	0.0319
	1.60	1.50	−31.912972	0.0329
	1.80	1.10	−36.782604	0.0234
	1.80	1.70	−30.805382	0.0374
	1.80	2.10	−11.385098	0.0999
	1.90	1.50	−26.519726	0.0643
	1.90	1.70	−25.050647	0.0574
	2.00	1.30	−35.885033	0.0209
	2.00	1.40	−28.803523	0.0810
	2.00	1.70	−18.269715	0.0583
	2.00	1.80	−16.136032	0.0516
	2.00	1.90	−21.877517	0.0735
	2.10	1.40	−30.081084	0.0527
	2.10	1.50	−33.315363	0.0402
	2.10	1.60	−26.856807	0.0818
	2.10	1.90	−23.855124	0.0484
	2.10	2.00	−10.926472	0.0442
	2.20	1.40	−35.385605	0.0239
	2.20	1.50	−31.225173	0.0373
	2.20	1.70	−33.373047	0.0378
	2.30	1.60	−21.964290	0.0748
	2.50	1.40	−30.894365	0.0375

(*continued*)

Table 1. (*Continued*)

q_T	q_A	q_V	Energy (Kcal/mol)	RMSD (nm)
	2.50	1.60	−23.441904	0.0660
	2.80	1.10	−28.285041	0.0955
	2.80	1.30	−35.225555	0.0215
	2.90	1.50	−31.587157	0.0390
2.30	1.10	1.20	−36.458935	0.0327
	1.20	1.20	−36.377508	0.0285
	1.20	1.80	−18.952144	0.0648
	1.30	1.10	−32.162446	0.0504
	1.30	1.20	−36.224863	0.0199
	1.30	1.30	−33.115554	0.0284
	1.30	1.40	−21.814475	0.0816
	1.30	1.70	−25.246935	0.0807
	1.30	1.80	−18.248446	0.0817
	1.40	1.30	−23.816371	0.0908
	1.40	1.70	−27.543728	0.0550
	1.40	1.80	−23.733647	0.0717
	1.50	1.20	−36.335151	0.0253
	1.50	1.30	−29.229434	0.0991
	1.50	1.40	−26.447157	0.0994
	1.50	1.50	−30.442776	0.0294
	1.60	1.20	−35.893353	0.0221
	1.60	1.50	−23.170798	0.0716
	1.60	2.00	−2.409806	0.0886
	1.70	1.20	−29.341834	0.0987
	1.70	1.60	−12.151106	0.0799
	1.70	1.70	−26.645821	0.0829
	1.80	1.10	−36.474679	0.0205
	1.80	1.20	−30.623803	0.0349
	1.90	1.10	−36.468513	0.0246
	1.90	1.30	−28.222654	0.0917
	2.00	1.20	−32.452925	0.0394
	2.00	1.30	−29.705226	0.0665
	2.00	1.40	−27.700133	0.0522
	2.10	1.20	−35.479531	0.0284
	2.10	1.50	−30.133661	0.0479
	2.10	1.70	−12.628410	0.0779
	2.20	1.20	−30.787481	0.0470
	2.20	1.50	−29.345978	0.0471
	2.30	1.10	−35.234990	0.0226
	2.30	1.20	−31.920088	0.0485
	2.30	1.30	−29.560031	0.0967
	2.30	1.40	−26.962583	0.0524
	2.30	1.60	−12.338435	0.0669
	2.30	1.70	−14.893874	0.0944
	2.40	1.10	−35.258969	0.0297
	2.40	1.40	−32.074786	0.0639
	2.40	1.60	−14.666472	0.0663
	2.50	1.80	−15.945011	0.0959
	2.70	1.40	−25.579318	0.0614
	2.70	1.50	−26.107084	0.0695
	2.90	1.40	−23.503316	0.0779
	2.90	1.80	−16.890885	0.0867
2.40	1.10	1.10	−35.465178	0.0371
	1.10	1.20	−35.659742	0.0302
	1.10	1.30	−19.440604	0.0826
	1.10	1.60	−11.039973	0.0779

(*continued*)

Table 1. (*Continued*)

q_T	q_A	q_V	Energy (Kcal/mol)	RMSD (nm)
	1.10	1.70	−22.059088	0.0722
	1.20	1.10	−34.058446	0.0347
	1.20	1.30	−24.870105	0.0874
	1.20	1.60	−26.657996	0.0607
	1.30	1.10	−33.162114	0.0282
	1.30	1.20	−33.000287	0.0354
	1.30	1.30	−24.927960	0.0622
	1.30	1.60	−18.613315	0.0929
	1.40	1.10	−34.303902	0.0339
	1.40	1.20	−34.637731	0.0271
	1.40	1.30	−32.659207	0.0339
	1.40	1.60	−24.775961	0.0475
	1.40	1.70	−14.728947	0.0906
	1.50	1.10	−28.661461	0.0699
	1.50	1.30	−34.018697	0.0366
	1.50	1.70	−15.277833	0.0945
	1.60	1.20	−32.929948	0.0407
	1.60	1.30	−23.823381	0.0928
	1.60	1.50	−21.812758	0.0587
	1.70	1.20	−34.744247	0.0299
	1.70	1.50	−31.933262	0.0301
	1.70	1.70	−23.959992	0.0561
	1.70	1.80	−11.030234	0.0859
	1.80	1.70	−22.311277	0.0700
	1.90	1.10	−34.024558	0.0290
	1.90	1.20	−35.206712	0.0409
	1.90	1.30	−23.147225	0.0927
	1.90	1.60	−20.319081	0.0824
	2.00	1.10	−31.291452	0.0554
	2.00	1.20	−33.800347	0.0379
	2.00	1.30	−33.022303	0.0394
	2.00	1.40	−28.376152	0.0257
	2.00	1.60	−25.150444	0.0720
	2.00	1.80	−19.001227	0.0854
	2.10	1.10	−32.118168	0.0448
	2.10	1.20	−24.803908	0.0985
	2.10	1.30	−29.846457	0.0855
	2.10	1.60	−1.638251	0.0924
	2.10	1.80	−10.744386	0.0860
	2.20	1.10	−35.141533	0.0297
	2.30	1.10	−33.251799	0.0391
	2.30	1.30	−29.261389	0.0464
	2.30	1.50	−21.377245	0.0480
	2.30	1.80	−22.965302	0.0592
	2.40	1.30	−20.782684	0.0598
	2.50	1.40	−23.409985	0.0367
	2.50	1.70	−27.566404	0.0603
	2.60	1.10	−35.614391	0.0328
	2.70	1.30	−25.968418	0.0867
	2.70	1.60	−20.758107	0.0694
	2.80	1.20	−33.269779	0.0649
2.5	1.10	1.10	−33.211154	0.0262
	1.20	1.10	−23.181241	0.0858
	1.20	1.20	−30.926229	0.0388
	1.30	1.10	−26.075598	0.0548
	1.30	1.80	−19.469142	0.0515

(*continued*)

Table 1. (*Continued*)

q_T	q_A	q_V	Energy (Kcal/mol)	RMSD (nm)
	1.40	1.10	−33.724794	0.0345
	1.50	1.10	−27.244986	0.0289
	1.60	1.10	−34.566421	0.0313
	1.60	1.20	−24.243920	0.0603
	1.60	1.50	−22.095125	0.0541
	1.70	1.20	−24.354027	0.0758
	1.70	1.50	−24.482464	0.0501
	1.80	1.20	−31.782017	0.0325
	1.80	1.40	−26.545958	0.0322
	1.80	1.50	−16.556582	0.0862
	1.90	1.10	−34.680855	0.0246
	1.90	1.30	−31.470938	0.0459
	1.90	1.40	−24.171618	0.0452
	2.00	1.30	−32.139461	0.0528
	2.10	1.10	−34.389134	0.0479
	2.10	1.20	−24.512433	0.0597
	2.20	1.20	−27.645879	0.0409
	2.20	1.30	−29.889405	0.0500
	2.30	1.10	−34.674889	0.0293
	2.30	1.20	−29.595157	0.0860
	2.30	1.30	−23.921979	0.0541
	2.40	1.20	−27.352437	0.0489
	2.40	1.40	−16.826819	0.0577
	2.40	1.90	−11.375487	0.0940
	2.50	1.20	−33.719819	0.0342
	2.50	1.30	−20.616729	0.0761
	2.60	1.10	−29.167026	0.0982
	2.80	1.20	−29.056932	0.0414
	2.80	1.30	−20.587840	0.0845
	2.80	1.40	−20.182878	0.0988
2.60	1.10	1.50	−27.104539	0.0938
	1.20	1.10	−28.966773	0.0832
	1.20	1.50	−23.711687	0.0894
	1.30	1.50	−21.916086	0.0955
	1.40	1.10	−25.563903	0.0500
	1.40	1.50	−28.381449	0.0922
	1.50	1.20	−26.796001	0.0678
	1.50	1.50	−22.605738	0.0863
	1.90	1.10	−33.093359	0.0521
	2.10	1.10	−27.629540	0.0840
	2.20	1.30	−25.281155	0.0830
	2.40	1.50	−12.380604	0.0786
	2.50	1.20	−28.570701	0.0835
	2.60	1.20	−28.435876	0.0428
	2.70	1.10	−30.024864	0.0666
	2.70	1.20	−21.373838	0.0951
	2.70	1.40	−16.044090	0.0575
	2.90	1.10	−33.039948	0.0436
	2.90	1.20	−25.605568	0.0543
2.70	1.10	1.10	−20.833448	0.0808
	1.50	1.10	−25.305289	0.0495
	1.60	1.10	−27.837058	0.0278
	1.80	1.10	−31.275296	0.0347
	2.00	1.10	−22.105883	0.0561
	2.10	1.10	−22.169439	0.0501
	2.70	1.20	−20.730631	0.0728

(*continued*)

Table 1. (*Continued*)

q_T	q_A	q_V	Energy (Kcal/mol)	RMSD (nm)
	2.70	1.30	−19.768336	0.0592
	2.90	1.10	−21.680708	0.0442
	2.90	1.20	−14.328725	0.0696
	2.90	1.30	−18.326209	0.0780
2.80	1.10	1.40	−13.703398	0.0698
	1.30	1.10	−18.486695	0.0728
	1.40	1.10	−18.374451	0.0797
	1.40	1.30	−11.869615	0.0721
	1.40	1.40	−18.587488	0.0767
	1.50	1.40	−22.869017	0.0905
	1.80	1.10	−23.388418	0.0961
	1.90	1.10	−25.676865	0.0837
	1.90	1.40	−15.660610	0.0686
	2.00	1.20	−20.316915	0.0525
	2.50	1.20	−24.907898	0.0719
	2.70	1.40	−6.805331	0.0956
2.90	1.70	1.30	−23.578941	0.0529
	2.30	1.20	−19.876439	0.0784
	2.40	1.10	−14.955900	0.0892
3.00	1.50	1.30	−16.319216	0.0459
	2.40	1.10	−16.651384	0.0582

of helix structures is obtained, causing several RMSD values below 0.1 nanometer (Figure 3e to Figure 3p). The q_T parameter takes an upper limit value to generate the helix conformation. When the temperature parameter is greater than 2.5, a smaller amount of RMSD values below 0.1 nm can be observed (Fig. 3q–t).

In Table 1, the RMSD values give more information about the folded α-helix structures than the energy values. All RMSD values below 0.1 nm are listed in this table, accompanied by their related q_T, q_A, and q_V values. From this table and Figure 3 we observe, therefore, that there is an optimum interval for q_T values to reach the folded structure, which ranges from 1.6 to 2.5. As foreseen, the greater is the q_T limit, the fast the cooling, causing the trapping of the structure in an energy local minimum. The existence of an upper limit for q_T values is due to the temperature dependence of the visiting and acceptance functions, as exposed in both equations in the methods section.

From the results in Figures 2 we observed an inverse tendency of the q_T values in relation to the q_v values to find the global minimum: while the temperature parameter increases, the visiting one decreases. This inverse tendency can be better observed in Figure 4, where we show the plot of q_T vs. q_A and q_V. In this figure, the q_V values decrease linearly with the enhancement of the q_T values, showing also the broadening of possible values that conduce to the folded α-helix structure. On the other hand, the acceptance q_A parameter, ranging from 1.1 to 2.9, shows no value preferences, mainly in the q_T interval from 1.6 to 2.5.

Discussion

There are many studies in the protein folding area with many different techniques being used, such as molecular dynamics,[20,21] homology modeling, genetic algorithms,[22] and others. In the area of molecular dynamics, the protein folding presents recent works[23,24] that combine task and data parallelism to speed up protein folding in a Grid platform. As their applications are not completely Bag of Tasks,[25] they do not have the same potential to exploit Grid architecture benefits the same way this work does. Also, other stochastic methods, such as Metropolis–Monte Carlo, can take millions of steps for the polyalanine α-helical folding.[18] Therefore, GSA seems to be a reasonable method for a more accurate folding of proteins. The GSA algorithm has been used in many problems involving molecular systems optimization and protein folding.[26–30]

As an initial step we use known structures to which results obtained from GSA can be compared, and 18-alanine is taken as the first step. Hence, validity of the results is studied, such that latter on unknown structures can be determined with a certain reliability. In our study we choose the α-helix formed by 18-alanine as a starting point because it is a stable and favorable structure in proteins. In the future, we intend to apply the method to larger and heterogeneous polypeptides.

Finally, massive calculations for the prediction of protein structure in a Grid may be an important step to fill in the huge gap between the knowledge of the amino acid sequence that compose proteins and the determination of three-dimensional structures through experimental methods, such as X-ray crystallography and nuclear magnetic resonance. This work searches for the best q_T, q_V, and q_A parameters for the global optimization of a polyalanine containing 18 residues. As summarized in Figure 4, we found that these parameters have a better interval of values ranging from 1.6 to 2.5 for q_T, 1.1 to 1.9 for q_V, and 1.1 to 2.9 for q_A.

The GSA application in protein folding studies is feasible, and presents a great advantage because the number of computational cycles to reach the global minimum is extremely small, in contrast with annealing procedures based on Boltzmann statistics. However, we found that a new temperature parameter should be implemented, decoupling the freezing process from the visiting distribution. By using q_T equal q_V, as proposed by Tsallis and Stariolo,[2,3] we found that it is not possible to have an ideal parameter set for protein folding studies, because the variance of these parameters present a linearly inverse tendency.

Because standard simulated annealing works, there was an attempt to define optimal cooling schedules using physical characteristics of the system.[4] Straub and coworkers approached this matter in the context of GSA using the ergodic measure.[31] However, we have found no parallel between physical characteristics of the studied system with the best interval of q_T, q_V, and q_A values observed in this work. We have achieved good results by applying these parameter set intervals to molecular docking in protein–ligand systems and folding of other polypeptide chains. By scanning q_T vs. q_V vs. q_A for several classes of molecular optimization problems we intend to find the patterns of parameter sets and try to find its physical meaning.

Acknowledgments

We acknowledge support from CNPq, CAPES, and LNCC. We also thank the computing facilities support from the XML com-

putation center of HP—Hewlett-Packard Computadores Ltda in Petrópolis, Rio de Janeiro, and the Carcara Project at LNCC.

References

1. Yon, J. M. Cell Mol Life Sci 1997, 53, 557.
2. Tsallis, C.; Stariolo, D. A. Physica A 1996, 233, 395.
3. Stariolo, D. A.; Tsallis, C. In Annual Reviews of Computational Physics; Stauffer, D., Ed.; World Scientific: Singapore, 1995, p. 343, vol. II.
4. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Science 1983, 220, 671.
5. Szu, H.; Hartley, R. Phys Lett A 1987, 122, 157.
6. Moret, M. A.; Bisch, P. M.; Vieira, F. M. C. Phys Rev E 1998, 57, R2535.
7. Moret, M. A.; Pascutti, P. G.; Bisch, P. M.; Mundim, K. C. J Comp Chem 1998, 19, 647.
8. Mundim, K. C.; Tsallis, C. Int J Quantum Chem 1996, 58, 373.
9. Mundim, K. C.; Lemaire, T.; Bassrei, A. Physica A 1998, 252, 405.
10. Dorfman, S.; Liubich, V.; Fuks, D.; Mundim, K. C. J Phys Condens Matter 2001, 13, 6719.
11. Fuks, D.; Dorfman, S.; Mundim, K. C.; Ellis, D. E. Int J Quantum Chem 2001, 85, 354.
12. Fuks, D.; Mundim, K. C.; Malbouisson, L. A. C.; Berner, A.; Dorfman, S.; Ellis, D. E. J Mol Struct 2001, 539, 199.
13. Berner, A.; Mundim, K. C.; Ellis, D. E.; Dorfman, S.; Fuks, D.; Evenhaim, R. Sensors Actuators 1999, 74, 86.
14. Penna, T. J. P. Phys Rev E 1995, 51, R1.
15. Penna, T. J. P. Comp Phys 1995, 9, 341.
16. Cirne, W.; Marzullo, K. Proceedings of the 13th Symposium on Computer Architecture and High Performance Computing, 2001.
17. Agostini, F. P.; Osthoff, C.; Vassalo, A.; Pinto, D. S.; Pascutti, P. G. Rev Technol Inform Bras 2003, 3, 103.
18. Moret, M. A.; Bisch, P. M.; Mundim, K. C.; Pascutti, P. G. Biophys J 2002, 82, 1123.
19. Dall'Igna, A., Jr.; Silva, R. S.; Mundin, K. C.; Dardenne, L. E. Genet Mol Biol 2004, 27, 616.
20. Zhou, Y.; Karplus, M. J Mol Biol 1994, 293, 917.
21. Zhou, Y.; Karplus, M. Nature 1999, 401, 400.
22. Dandekar, T.; Argos, P. J Mol Biol 1994, 236, 844.
23. Uk, B.; Taufer, M.; Stricker, T.; Settanni, G.; Cavalli, A.; Caflisch, A. Darthmouth University Technical Report TR2003–38, 2003.
24. Uk, B.; Taufer, M.; Stricker, T.; Settanni, G.; Cavalli, A. Darthmouth University Technical Report TR2003–385, 2003.
25. Cirne, W.; Paranhos, D.; Costa, L.; Santos–Neto, E.; Brasileiro, F.; Sauvé, J.; Silva, F.; Osthoff, C. 5th International Conference in Parallel Processing, Taiwan, 2003.
26. Andricioaei, I.; Straub, J. E. Phys Rev E 1996, 53, R3055.
27. Hansmann, U. H. E. Physica A 1997, 242, 250.
28. Hansmann, U. H. E. Comp Phys Commun 2002, 147, 604.
29. Hansmann, U. H. E. Phys Rev E 2004, 70, 01290.
30. Hansmann, U. H. E.; Okamoto, Y. Bras J Phys 1999, 29, 187.
31. Whitfield, L. B.; Straub, J. E. Physica A 2002, 305, 157.