Poster I-89

Tsallis Statistics Applied to Protein Folding, Protein Model Refinement and Molecular Complex Optimization

Authors:
Pedro G. Pascutti (*Instituto de Biofísica Carlos Chagas Filho - Universidade Federal do Rio de Janeiro - Brasil*)
Tacio V. Amorim Fernandes (*Instituto de Biofísica Carlos Chagas Filho - Universidade Federal do Rio de Janeiro - Brasil*)
Flávia P. Agostini (*Laboratório Nacional de Computação Científica/CNPq - Petrópolis-RJ - Brasil*)

Short Abstract: We applied a proposal of generalization for Statistical Mechanics using stochastic simulated annealing coupled to a molecular Force Field in protein folding studies, and show that in this strategy is necessary a new parameter to control the freezing process, to avoid polypeptide chains to be trapped in energy local minima.

Long Abstract:
Statistical Physics assumes that energy is an extensive variable, meaning that the total energy of a system is proportional to the system size. Similarly, the entropy is also supposed to be an extensive variable in such way that for two related systems “A” and “B” the total entropy is:

$$S(AB) = S(A) + S(B)$$

Generally this is justified by appealing to the short-range nature of the interactions which hold matter together, form chemical bonds, etc. But suppose we deal with long-range interactions, most prominently gravity and electrostatic, we can then find that entropy is not extensive. Constantino Tsallis proposed a generalization for Statistical Mechanics saying that the total entropy for two related system can be described by:

$$S(AB) = S(A) + S(B) + [1 – q]S(A)S(B)$$

The “q” parameter is a measure of how correlated is the systems “A” and “B”. For low-correlated systems, “q” tends to “1” and the total entropy can be described by the Boltzmann entropy in standard Statistical Mechanics.

The Generalized Simulated Annealing (GSA) method proposed by Tsallis and Stariolo (1986), taking into account Tsallis Statistics, has been well suited for a large extent of optimization problems, especially those containing many local minima. In this method the energy surface is scanned using a visiting function and new configurations are accepted using an acceptance function, both functions are adjusted by q-values: the visiting, qV, and the accepting, qA, respectively. For special conditions (qV = 2, and qA = 1) the GSA becomes the standard Monte Carlo Simulated Annealing method. In the GSA algorithm the temperature decreasing is also controlled by qV.

Unfolded proteins have a great number of accessible conformations due to the large number of rotation degrees of freedom around chemical bonds, leading to several local minima on the energy surface. It has been proposed though, that proteins express their biological function when their structure is close to a conformation with energy global minimum, the folded state.

We applied the GSA method coupled to the GROMOS96 Molecular Force Field in order to research the minimum energy conformation of several molecular systems, from molecular docking to protein folding problem (Moret, et al., 1998, 2001, 2002, 2005, 2006; Agostini et al., 2006; Shida et al. 2006). In this work we show that a new qT GSA parameter can be used to more efficient 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 (qV), and accepting (qA) functions for qT values ranging from 1 to 3, and we found that these parameters have a better interval of values ranging from 1.6 to 2.5 for qT, 1.1 to 1.9 for qV , and 1.1 to 2.9 for qA.

The GSA application in flexible molecular docking, protein model refinement after Homology Modeling and 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 qT equal qV, as proposed by Tsallis and Stariolo, 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.

Key words: protein folding, molecular docking,, global optimization, Generalized Simulated Annealing, Tsallis' statistics.

References:

Tsallis, C., and Stariolo, D. A. (1996), Generalized Simulated Annealing, Physica A - Statistical and Theoretical Physics 233:395-404.

Moret, M. A., Pascutti, P. G., Bisch, P. M., Mundim, M. S. P., Mundim, K. C. (2006), Classical and quantum conformational analysis using Generalized Genetic Algorithm. Physica A - Statistical and Theoretical Physics, 363, 260-268

Moret, M. A., Bisch, P. M., Nogueira-Jr, E., Pascutti, P. G., (2005), Stochastic strategy to analyze protein folding, Physica A - Statistical and Theoretical Physics, 353, 353-364.

Moret, M. A., Bisch, P. M., Mundim, K. C., Pascutti, P. G. (2002), A new stochastic strategy to analyze helix folding, Biophysical Journal, 82(3), 1123-1132.

Moret, M. A., Pascutti, P. G., Mundim, K. C., Bisch, P. M., Nogueira-Jr, E. (2001), Multifractality, Levinthal paradox, and energy hypersurface, Physical Review E, 63, 20901-04.

Moret, M. A., Pascutti, P. G., Mundim, K. C., Bisch, P. M., (1998), Stochastic molecular optimization using generalized simulated annealing, J. Computational Chemistry, 19(6), 647-657.

Agostini, F. P., Soares-Pinto, D. O., Osthoff, C., Moret, M. A., Pascutti, P. G. (2006), Protein Folding Studies by Generalized Simulated Annealing, Journal of Computational Chemistry, in press.

Shida, C. S., Sousa, G. L. S. C., Bish , P. M, and Pascutti, P. G. (2006), Study of heparin disaccharide and lisossomal protease cathepsin B complex using docking by Generalized Simulated Annealing and Molecular Dynamics Simulation, in preparation.