Effects of constraints in general branched molecules: A quantitative ab initio study in HCO-L-Ala-NH$_2$

Pablo Echenique[*,†], J. L. Alonso[*,†] and Iván Calvo[*,†]

[*]*Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, Pedro Cerbuna 12, 50009, Zaragoza, Spain.*
[†]*Instituto de Biocomputación y Física de los Sistemas Complejos (BIFI), Edificio Cervantes, Corona de Aragón 42, 50009, Zaragoza, Spain.*

Abstract. A general approach to the design of accurate classical potentials for protein folding is described. It includes the introduction of a meaningful statistical measure of the differences between approximations of the same potential energy, the definition of a set of Systematic and Approximately Separable and Modular Internal Coordinates (SASMIC), much convenient for the simulation of general branched molecules, and the imposition of constraints on the most rapidly oscillating degrees of freedom. All these tools are used to study the effects of constraints in the Conformational Equilibrium Distribution (CED) of the model dipeptide HCO-L-Ala-NH$_2$. We use ab initio Quantum Mechanics calculations including electron correlation at the MP2 level to describe the system, and we measure the conformational dependence of the correcting terms to the naive CED based in the Potential Energy Surface (PES) without any simplifying assumption. These terms are related to mass-metric tensors determinants and also occur in the Fixman's compensating potential. We show that some of the corrections are non-negligible if one is interested in the whole Ramachandran space. On the other hand, if only the energetically lower region, containing the principal secondary structure elements, is assumed to be relevant, then, all correcting terms may be neglected up to peptides of considerable length. This is the first time, as far as we know, that the analysis of the conformational dependence of these correcting terms is performed in a relevant biomolecule with a realistic potential energy function.

Keywords: constraints, protein folding, ab initio, conformational equilibrium, dipeptide, mass metric tensor, Fixman potential
PACS: 02.50.Cw, 31.15.Ar, 87.14.Ee, 87.15.Aa, 87.15.Cc

INTRODUCTION

Proteins are long chains comprised of twenty different amino acidic monomers and they are central elements in the biological machinery of all known living beings. They perform most of the catalytic tasks that are vital in the many coupled chains of chemical reactions occurring in the cells, they are found as structural building blocks in the cytoskeleton or in organelles, such as the ribosome, and they also play a very important role as membrane receptors. Their absence or malfunctioning is related to many diseases such as Creutzfeldt-Jakob's or Cancer [1, 2] and the proteins involved in the biology of pathogens are often the preferred target of newly designed drugs (see the talks by E. Freire and C. Cavasotto in this meeting).

Despite their complexity and the many opposing forces that determine their behaviour, these molecules swiftly acquire a unique three-dimensional *native* structure in the physiological milieu. Some details of this process are still not clear (see the talk by J. M. Sánchez-Ruiz), such as the relative proportion of the naturally occurring proteins that fold co- or post-translationally (i.e., during or after biosynthesis at the ribosome) [3], or the role played by molecular chaperones such as GroEL (see M. Karplus' talk), the Prolyl-peptidyl-isomerase, or the Protein disulfide isomerase, among others. However, since the pioneering work of Anfinsen [4], it is known that a large number of water-soluble globular proteins are capable of reaching their native structure *in vitro* after being unfolded by changes in their environment, such as a raise of the temperature, the addition of denaturing agents or a change in the pH. It is the prediction of the native structure in these cases (only from the amino acid sequence and the laws of physics) that has become paradigmatic and receives the name of *protein folding problem*.

In 2005, in a special section of the Science magazine entitled "What don't we know?" [5], a selection of the hundred most interesting yet unanswered scientific questions was presented. What indicates the importance of the protein folding problem is not the inclusion of the question *Can we predict how proteins will fold?*, which was a must, but the large number of other questions which were related to or even dependent on it, such as *Why do humans have so few genes?*, *How much can human life span be extended?*, *What is the structure of water?*, *How does a single somatic cell become a whole plant?*, *How many proteins are there in humans?*, *How do proteins find their partners?*, *How do prion diseases work?*, *How will big pictures emerge from a sea of biological data?*, *How far can we push chemical self-assembly?* or *Is an effective HIV vaccine feasible?*, to quote just a few of them.

Some authors [6] divide the problem in two parts: the prediction of the three-dimensional, biologically functional, native state of a protein and the description of the actual folding process that takes the protein there from the unfolded state. The first part, which is more pressing and more technologically oriented, is included in the second part and it is, therefore, easier to tackle, as the relative success of knowledge-based methods suggests [7, 8]. However, we believe that, not only much theoretical insight may be gained from a solution of the more general second part of the problem, but also much engineering and design power, as well as new comprehension about so distinct topics as the ones quoted in the preceding paragraph. This is why our approach is one of bottom-top and ab initio flavor.

POTENTIAL ENERGY FUNCTIONS

The fundamental theory of matter that is nowadays accepted as correct by the scientific community is Quantum Mechanics. For the study of the conformational behaviour of a molecule consisting of n atoms, with atomic numbers Z_α and masses M_α, $\alpha = 1,\ldots,n$, one typically assumes that relativistic effects are negligible[1] and that, according to the Born-Oppenheimer scheme [9], the great differences in mass between electrons and nuclei allows to consider that the former are described by a Hamiltonian which

[1] Which, in organic molecules, is approximately correct for all the particles involved, except, maybe, for some core electrons in the heaviest atoms.

is adiabatically decoupled from the nuclear one and that depends only parametrically on the positions of the nuclei. Hence, the behaviour of the system *in vacuum* may be extracted from the non-relativistic time-independent nuclear Schrödinger equation:

$$\left(-\sum_{\alpha=1}^{n} \frac{\hbar^2}{2M_\alpha} \nabla_\alpha^2 + \sum_{\beta>\alpha} \left(\frac{e^2}{4\pi\varepsilon_0} \right) \frac{Z_\alpha Z_\beta}{|\vec{R}_\beta - \vec{R}_\alpha|} + E_{\mathrm{e}}^0(R) \right) \Psi_{\mathrm{N}}(R) = E\,\Psi_{\mathrm{N}}(R) \ , \qquad (1)$$

where $E_{\mathrm{e}}^0(\vec{R})$ denotes the effective potential due to the electronic cloud in the fundamental energy state[2] and R is shorthand for $\vec{R}_1,\ldots,\vec{R}_n$.

Despite the exponential growth in computing power that has been taking place in the last decades (see, for example, the talks by A. Perczel and I. Campos), a precise description of the behaviour of any biologically interesting system derived from the solution of (1) remains far from being even imaginable. Not to mention the huge complications that arise when the unavoidable inclusion of solvent is considered. This is why, omitting a myriad of possible intermediate descriptions, the most popular choice for the *in silico* prediction of the protein folding process has become the use of the so-called *force fields* [10–13], in which one assumes that the behaviour of the macromolecule (omitting again the solvent, to compare with (1)) is *classical* and may be described via a very simple potential energy function which, typically, has the form

$$\begin{aligned} V_{\mathrm{ff}} \ := \ & \frac{1}{2}\sum_{\alpha=1}^{N_r} K_{r_\alpha}(r_\alpha - r_\alpha^0)^2 + \frac{1}{2}\sum_{\alpha=1}^{N_\theta} K_{\theta_\alpha}(\theta_\alpha - \theta_\alpha^0)^2 + \sum_{\alpha=1}^{N_\phi} A_\alpha \cos(B_\alpha \phi_\alpha + \phi_\alpha^0) + \\ + \ & \sum_{\beta>\alpha} \left(\frac{C_{12}^{\alpha\beta}}{R_{\alpha\beta}^{12}} - \frac{C_6^{\alpha\beta}}{R_{\alpha\beta}^6} \right) + \sum_{\beta>\alpha} \left(\frac{e^2}{4\pi\varepsilon_0} \right) \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \qquad (2) \end{aligned}$$

where r_α are bond lengths, θ_α are bond angles, ϕ_α are dihedral angles[3] and $R_{\alpha\beta}$ denotes the interatomic distances. Finally, all the parameters entering (2) (which may amount to thousands) are customarily fitted to reproduce thermodynamical measurements or taken from quantum mechanical calculations.

While it is true that these empirical potentials may be detailed enough to deal with simple conformational transitions in already folded proteins (see the talk by J. Luque) or with collective motions of systems of many proteins (see M. Karplus' talk), and that they may also be used as scoring functions for protein design (as in the talk by A. Jaramillo), all these applications require only that the energetics of the native structure and its surroundings be correctly described. As A. Tramontano told us in her talk, the usefulness of these simple potentials for *de novo* structural prediction (assessed via the CASP contest[4]) remains much limited.

[2] This additional assumption that the electrons are in the fundamental state prevents us from describing the catalytic behaviour of most enzymes, however, the only interest here is to describe the folding process.

[3] For the sake of simplicity, no harmonic terms have been assumed for out-of-plane angles or for hard dihedrals, such as the peptide bond angle ω

[4] See `http://predictioncenter.org`

We believe that one of the reasons of this failure is the lack of accuracy of the potential energy functions used, since, even if the parameters fit is properly carried out, the choice of the very particular dependencies, for example those in (2), constitutes a heavy restriction in the space of functions. Accordingly, one of our aims is the design of classical potentials which are as similar as possible to the effective Born-Oppenheimer one in (1). To do this, one must calculate the electronic energy $E_{\mathrm{e}}^{0}(R)$ using the powerful tools of Quantum Chemistry (see the talks by A. Perczel, J. J. Dannenberg and M. Amzel) and devise numerically efficient approximations to it.

In any case, in order to walk the long path connecting Quantum Mechanics and a classical description amenable to nowadays computers, one must have a meaningful way of comparing different approximations of the potential energy of a system. Much in the spirit of the talk by M. Wall, and using the fact that the complex nature of biological molecules suggests the convenience of statistical analyses, we have designed in [14] a *distance*, denoted by d_{12}, between any two different potential energy functions, V_1 and V_2, that, from a working set of conformations, measures the typical error that one makes in the *energy differences* if V_2 is used instead of the more accurate V_1, admitting a linear rescaling and a shift in the energy reference.

This distance, which has energy units, presents better properties than other quantities customarily used to perform these comparisons, such as the energy RMSD, the average energy error, etc. It may be related to the Pearson's correlation coefficient by

$$d_{12} = \sqrt{2}\,\sigma_2 (1 - r_{12}^2)^{1/2} \ . \tag{3}$$

Finally, due to its physical meaning, it has been argued in [14] that, if the distance between two different approximations of the energy of the same system is less than RT, one may safely substitute one by the other without altering the relevant dynamical or thermodynamical behaviour.

EFFECTS OF CONSTRAINTS

Another reason underlying the difficulties faced in the computational study of the protein folding problem is that the large number of degrees of freedom brings up the necessity to sample an astronomically large conformational space [15]. In addition, the typical timescales of the different movements are in a wide range and, therefore, demandingly small timesteps must be used in Molecular Dynamics simulations in order to properly account for the fastest modes [16], which lie in the femtosecond range; whereas the folding of a large protein may take seconds. In order to deal with these problems, one may naturally consider the reduction of the number of degrees of freedom describing macromolecules via the imposition of constraints.

To manage this situation, we have made progresses in two directions. First, we have devised [17] a set of internal coordinates called *SASMIC* (standing for *Systematic and Approximately Separable and Modular Internal Coordinates*), which are much convenient to describe branched molecules and, specially, polypeptides, without having to rewrite the whole Z-matrix upon addition of new residues to the chain, and also allow to maximally separate the soft and hard movements[5].

Second, we have used these coordinates, the distance discussed before and the factorization of the external variables in the mass-metric determinants that we describe in [18], to study the possibility of neglecting the conformational dependence of the correcting terms that appear in the equilibrium distribution of organic molecules [19].

Constraining the hard coordinates q^I to be specific functions $f^I(q^i)$ of the soft ones (which defines a hypersurface Σ in the whole conformational space) produces two *classical* constrained models which are known to be conceptually [20, 21] and practically [22, 23] inequivalent: they are called *stiff* and *rigid*. In the classical rigid model, the constraints are assumed to be *exact* and all the velocities that are orthogonal to the hypersurface defined by them vanish. In the classical stiff model, on the other hand, the constraints are assumed to be *approximate* and they are implemented by a steep potential that drives the system to the constrained hypersurface. In this case, the orthogonal velocities are activated and may act as ‘heat containers’.

The conformational equilibrium of the system, according to these models, is described by the following probability densities [19]:

$$
\begin{array}{ll}
\text{Classical Stiff Model} & \text{Classical Rigid Model} \\
\hline
P_{\mathrm{s}}(q^u) = \dfrac{\exp\left[-\beta F_{\mathrm{s}}(q^u)\right]}{Z'_{\mathrm{s}}} & P_{\mathrm{r}}(q^u) = \dfrac{\exp\left[-\beta F_{\mathrm{r}}(q^u)\right]}{Z'_{\mathrm{r}}} \\
F_{\mathrm{s}}(q^u) := V_\Sigma(q^i) - T\left(S^c_{\mathrm{s}}(q^i) + S^{\mathrm{k}}_{\mathrm{s}}(q^u)\right) & F_{\mathrm{r}}(q^u) := V_\Sigma(q^i) - T S^{\mathrm{k}}_{\mathrm{r}}(q^u) \\
S^{\mathrm{k}}_{\mathrm{s}}(q^u) := \dfrac{R}{2}\ln\left[\det G\left(q^u, f^I(q^i)\right)\right] & S^{\mathrm{k}}_{\mathrm{r}}(q^u) := \dfrac{R}{2}\ln\left[\det g(q^u)\right] \\
S^c_{\mathrm{s}}(q^i) := -\dfrac{R}{2}\ln\left[\det \mathscr{H}(q^i)\right] &
\end{array}
\tag{4}
$$

where $\beta := 1/RT$, V_Σ is the potential energy in Σ (the Potential Energy Surface (PES) frequently used in Quantum Chemistry), and G, g and $\mathscr{H}$ denote, respectively, the whole-space mass-metric tensor, the reduced mass-metric tensor in Σ and the Hessian of the constraining part of the potential.

The different terms that correct the PES V_Σ in (4) are regarded (and denoted) as entropies because they are linear in the temperature T and come from the averaging out of certain degrees of freedom (sometimes coordinates, sometimes momenta). Accordingly, the effective potentials occurring in the exponent of the equilibrium probabilities are regarded (and denoted) as free energies.

Now, if Monte Carlo simulations in the coordinate space are to be performed [24, 25] and the probability densities that correspond to any of these two models sampled, the correcting entropies in (4) should be included or, otherwise, showed to be negligible.

On the other hand, if rigid Molecular Dynamics simulations are performed with the intention of sampling from the *stiff* equilibrium probability P_{s} [26–28], then, the so-

[5] An automatic Perl script that generates the SASMIC Z-matrix, in the format of typical Quantum Chemistry packages, such as GAMESS or Gaussian03, from the sequence of amino acids, may be found at `http://neptuno.unizar.es/files/public/gen_sasmic/`

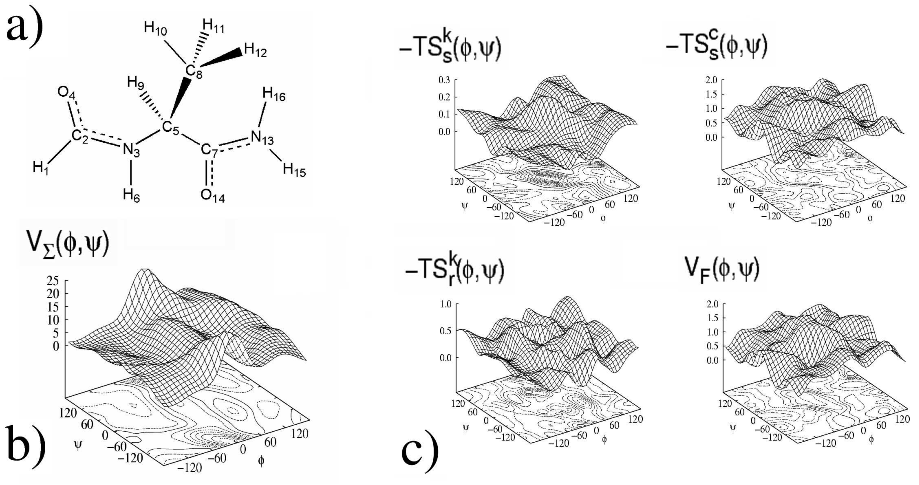

FIGURE 1. **a)** Model dipeptide HCO-L-Ala-NH$_2$ numbered according to the SASMIC [17] scheme. **b)** Potential Energy Surface. **c)** Conformational dependence of the correcting terms. All energies are given in kcal/mol.

called *Fixman's compensating potential* [29],

$$V_{\mathrm{F}}(q^u) := TS^{\mathrm{k}}_{\mathrm{r}}(q^u) - TS^{c}_{\mathrm{s}}(q^i) - TS^{\mathrm{k}}_{\mathrm{s}}(q^u) = \frac{RT}{2} \ln \left[\frac{\det G(q^u)}{\det \mathscr{H}(q^i) \det g(q^u)} \right] , \qquad (5)$$

must be added to the PES V_Σ.

The conformational dependence of most of the determinants appearing in (4) and (5) is frequently assumed to be negligible in the literature and they are consequently dropped from the calculations [30–33]. Also, subtly entangled to the assumptions underlying these simplifications, a second type of approximation is made that consists of assuming that the equilibrium values of the hard coordinates do not depend on the soft coordinates [31–34]. This has been argued to be only approximate even in the case of classical force fields [35–37].

In [19], we have eliminated all simplifying assumptions and measured the conformational dependence on the Ramachandran angles ϕ and ψ (the soft coordinates) of *all correcting terms* and of the Fixman's compensating potential in the model dipeptide HCO-L-Ala-NH$_2$. The potential energy function used was the effective Born-Oppenheimer potential for the nuclei (see (1)) derived from ab initio quantum mechanical calculations including electron correlation at the MP2/6-31++G(d,p) level of the theory.

In table 1, the main results of our work are presented. The importance of all the correcting terms is assessed by comparing (with the statistical distance d_{12} described in the previous section) the effective potential V_1, containing the term, with the approximate one V_2, lacking it. Moreover, if one assumes that the effective energies compared will be used to construct a polypeptide potential, the number N_{res} of residues up to which one may go keeping the distance between the two approximations of the the N-residue potential below RT is (see eq. (23) in [14]):

TABLE 1. Quantitative assessment of the importance of the different correcting terms involved in the study of the constrained equilibrium of the protected dipeptide HCO-L-Ala-NH$_2$ (see [19]).

Corr.[*]	V_1[†]	V_2[**]	d_{12}[‡]	N_{res}[§]	b_{12}[¶]	r_{12}[‖]
$-TS_{\mathrm{s}}^{\mathrm{k}} - TS_{\mathrm{s}}^{\mathrm{c}}$	F_{s}	V_{Σ}	0.74 RT	1.82	0.98	0.9967
$-TS_{\mathrm{s}}^{\mathrm{c}}$	F_{s}	$V_{\Sigma} - TS_{\mathrm{s}}^{\mathrm{k}}$	0.74 RT	1.83	0.98	0.9967
$-TS_{\mathrm{s}}^{\mathrm{k}}$	F_{s}	$V_{\Sigma} - TS_{\mathrm{s}}^{\mathrm{c}}$	0.11 RT	80.45	1.00	0.9999
$-TS_{\mathrm{r}}^{\mathrm{k}}$	F_{r}	V_{Σ}	0.29 RT	11.62	1.01	0.9995
V_{F}	F_{s}	F_{r}	0.67 RT	2.24	0.97	0.9972

[*] Correcting term whose importance is measured in the corresponding row

[†] ‘Correct’ potential energy; the one containing the correcting term

[**] ‘Approximate’ potential energy; the one lacking the correcting term

[‡] Statistical distance between V_1 and V_2 (see [14])

[§] Number of residues in a polypeptide potential up to which the correcting term may be omitted

[¶] Slope of the linear rescaling between V_1 and V_2

[‖] Pearson’s correlation coefficient

$$N_{\mathrm{res}} = \left(\frac{RT}{d_{12}} \right)^2 . \tag{6}$$

In the table, one can see that, in the stiff model, the Hessian-related correcting term should be included in Monte Carlo simulations for peptides as short as two residues, while the one that depends on G may be neglected up to chains which are $\sim$ 80 residues long. The only correcting term occurring in the rigid model, in turn, may be dropped up to $\sim$ 12 residues. Finally, the Fixman potential, containing all determinants, should be included in MD rigid simulations of peptides with more than two residues[6].

These results are related to a working set of conformations consisting of 144 points regularly distributed in the whole Ramanchandran space. In a second part of the work, we have repeated all the comparisons for a working set consisting of six secondary structure elements. The results suggest that, if one is interested only in this energetically lower region, the distances d_{12} are roughly divided by two and, accordingly, the values of N_{res} are four times larger.

We have also repeated the calculations, with the same basis set (6-31++G(d,p)) and at the Hartree-Fock level of the theory in order to investigate if this less demanding method without electron correlation may be used in further studies. We have found that, indeed, this can be done, obtaining very similar results at a tenth of the computational effort.

As far as we are aware, this is *the first time* that this type of study is performed in a relevant biomolecule with a realistic potential energy function.

[6] One should note that the distance between the PES V_{Σ} at MP2/6-31++G(d,p) and the one computed at HF/6-31++G(d,p) is $d_{12} \simeq 1.2\,RT$. A value slightly larger but of the order of the ones obtained when the most important correcting terms are dropped.

ACKNOWLEDGMENTS

We thank F. Falceto, V. Laliena, F. Plo and D. Zueco for illuminating discussions. The numerical calculations have been performed at the BIFI computing center. We thank I. Campos, for the invaluable CPU time and the efficiency at solving problems.

This work has been supported by the Aragón Government ("Biocomputación y Física de Sistemas Complejos" group) and by the research grants MEC (Spain) FIS2004-05073 and FPA2003-02948, and MCYT (Spain) BFM2003-08532. P. Echenique and I. Calvo are supported by MEC (Spain) FPU grants.

REFERENCES

1. C. M. Dobson, *Nature* **729**, 729 (2002).
2. J. W. Kelly, *Nat. Struct. Biol.* **9**, 323 (2002).
3. M. A. Basharov, *J. Cell. Mol. Med.* **7**, 223–237 (2003).
4. C. B. Anfinsen, *Science* **181**, 223–230 (1973).
5. Many authors, *Science* **309**, 78–102 (2005), `http://www.sciencemag.com`.
6. V. Daggett, and A. Fersht, *Nat. Rev. Mol. Cell Biol.* **4**, 497 (2003).
7. J. Moult, *Curr. Opin. Struct. Biol.* **15**, 285–289 (2005).
8. C. A. Rohl, C. E. Strauss, K. M. Misura, and D. Baker, *Methods Enzymol.* **383**, 66–93 (2004).
9. M. Born, and J. R. Oppenheimer, *Ann. Phys. Leipzig* **84**, 457–484 (1927).
10. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.* **4**, 187–217 (1983).
11. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, J. Merz, K. M., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
12. W. L. Jorgensen, and J. Tirado-Rives, *J. Am. Chem. Soc.* **110**, 1657–1666 (1988).
13. T. A. Halgren, *J. Comp. Chem.* **17**, 490–519 (1996).
14. J. L. Alonso, and P. Echenique, *J. Comp. Chem.* **27**, 238–252 (2006).
15. K. A. Dill, *Prot. Sci.* **8**, 1166–1180 (1999).
16. T. Schlick, E. Barth, and M. Mandziuk, *Annu. Rev. Biophys. Biomol. Struct.* **26**, 181–222 (1997).
17. P. Echenique, and J. L. Alonso, *To appear in J. Comp. Chem.* (2006), (arXiv:q-bio.BM/ 0511004).
18. P. Echenique, and I. Calvo, *Submitted to J. Comp. Chem.* (2006), (arXiv:q-bio.QM/ 0512033).
19. P. Echenique, I. Calvo, and J. L. Alonso, *Submitted to J. Comp. Chem* (2006), (arXiv:q-bio.QM/ 0601042).
20. J. M. Rallison, *J. Fluid Mech.* **93**, 251–279 (1979).
21. E. Helfand, *J. Chem. Phys.* **71**, 5000 (1979).
22. D. Chandler, and B. J. Berne, *J. Chem. Phys.* **71**, 5386–5387 (1979).
23. M. Gottlieb, and R. B. Bird, *J. Chem. Phys.* **65**, 2467 (1976).
24. R. A. Abagyan, M. M. Totrov, and D. A. Kuznetsov, *J. Comp. Chem.* **15**, 488–506 (1994).
25. E. W. Knapp, and A. Irgens-Defregger, *J. Fluid Mech.* **14**, 19–29 (1993).
26. M. Pasquali, and D. C. Morse, *J. Chem. Phys.* **116**, 1834 (2002).
27. S. He, and H. A. Scheraga, *J. Chem. Phys.* **108**, 287 (1998).
28. M. Fixman, *J. Chem. Phys.* **69**, 1527 (1978).
29. M. Fixman, *Proc. Natl. Acad. Sci. USA* **71**, 3050–3053 (1974).
30. N. Gō, and H. A. Scheraga, *Macromolecules* **9**, 535 (1976).
31. A. Patriciu, G. S. Chirikjian, and R. V. Pappu, *J. Chem. Phys.* **121**, 12708–12720 (2004).
32. M. P. Allen, and D. J. Tildesley, *Computer simulation of liquids*, Clarendon Press, Oxford, 2005.
33. D. Frenkel, and S. B., *Understanding molecular simulations: From algorithms to applications*, Academic Press, Orlando FL, 2002, 2nd edn.
34. D. C. Morse, *Adv. Chem. Phys.* **128**, 65–189 (2004).
35. J. Zhou, S. Reich, and B. R. Brooks, *J. Chem. Phys.* **111**, 7919 (2000).
36. B. Hess, H. Saint-Martin, and H. J. C. Berendsen, *J. Chem. Phys.* **116**, 9602 (2002).
37. J. Chen, W. Im, and C. L. Brooks III, *J. Comp. Chem.* **26**, 1565–1578 (2005).