Quantum mechanical calculation of the effects of stiff and rigid constraints in the conformational equilibrium of the Alanine dipeptide

Pablo Echenique[1,2*], Iván Calvo[1,2] and J. L. Alonso[1,2]

[1] Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, Pedro Cerbuna 12, 50009, Zaragoza, Spain.

[2] Instituto de Biocomputación y Física de los Sistemas Complejos (BIFI), Edificio Cervantes, Corona de Aragón 42, 50009, Zaragoza, Spain.

January 25, 2006

Abstract

If constraints are imposed on a macromolecule, two inequivalent classical models may be used: the stiff and the rigid one. The equilibrium probability densities that must be sampled in Monte Carlo simulations include the determinants of different mass-metric tensors and also of the Hessian matrix of the constraining part of the potential in the stiff case. These determinants, which also occur in the Fixman's compensating potential, produce correcting terms to the potential energy that are customarily assumed to be independent of the internal conformation of the macromolecule and thus dropped from the expressions. In addition, the equilibrium values of the hard coordinates are typically assumed to be independent from the soft ones. In this work, we exhaustively review the use of these approximations in the literature and analyze their validity in the model dipeptide HCO-L-Ala-NH_2 with ab initio Quantum Mechanics calculations including electron correlation at the MP2 level. The conformational dependence of *all correcting terms* and the Fixman's compensating potential is measured without any simplifying assumption and showed to be non-negligible in some cases if one is interested in the whole Ramachandran space. 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.

PACS: 87.14.Ee, 87.15.-v, 87.15.Aa, 87.15.Cc, 89.75.-k

*Corresponding author. E-mail address: pnique@unizar.es

1 Introduction

In computer simulations of large complex systems, such as macromolecules and, specially, proteins [1–6], one of the main bottlenecks to design efficient algorithms is the necessity to sample an astronomically large conformational space [3, 7]. In addition, being the typical timescales of the different movements in a wide range, demandingly small timesteps must be used in Molecular Dynamics simulations in order to properly account for the fastest modes, which lie in the femtosecond range. However, most of the biological interesting behaviour (allosteric transitions, protein folding, enzymatic catalysis) is related to the slowest conformational changes, which occur in the timescale of milliseconds or even seconds [4, 8–11]. Fortunately, the fastest modes are also the most energetic ones and are rarely activated at room temperature. Therefore, in order to alleviate the computational problems and also simplify the images used to think about these elusive systems, one may naturally consider the reduction of the number of degrees of freedom describing macromolecules via the imposition of constraints [12].

How to study the conformational equilibrium of these constrained systems has been an object of much debate [13–17]. Two different classical models exist in the literature which are conceptually [13–16, 18, 19] and practically [6, 13, 20–24] inequivalent. 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* [1] 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".

In this work, we do not address the question of which model is a better approximation of physical reality. Although, in the literature, it is commonly assumed (often implicitly) that the classical stiff model should be taken as a reference [6, 9, 16, 19, 20, 22, 26], we believe that this opinion is much influenced by the use of popular classical force fields [6, 27–30] (which are stiff by construction) and by the goal of reproducing their results at a lower computational cost, i.e., using rigid Molecular Dynamics simulations [4, 5, 8, 9, 14, 19, 21–23, 25, 26, 31–35]. In our opinion, the question whether the rigid or the stiff model should be used to approximate the real quantum mechanical statistics of an arbitrary organic molecule has not been satisfactorily answered yet. For discussions about the topic, see references [13–15, 17, 18, 36–38]. In this work, we adopt the cautious position that any of the two models may be useful in certain cases or for certain purposes and we study them both on equal footing. Our concern is, then, to

[1] Some authors use the word *flexible* to refer to this model [15, 21, 22, 25]. We, however, prefer to term it *stiff* [18] and keep the name *flexible* to refer to the case in which no constraints are imposed.

study the effects that either way of imposing constraints causes in the conformational equilibrium of macromolecules.

In the Born-Oppenheimer approximation [39] customarily used in Quantum Mechanics and in the majority of the classical force fields, the relevant degrees of freedom are the Euclidean (also called *Cartesian* by some authors) $3n$ coordinates of the n nuclei. However, it is frequent to define a different set of coordinates in which the overall translation and rotation of the system are distinguished and the remaining $3n-6$ degrees of freedom are chosen (according to different prescriptions as *internal coordinates*, which are simple geometrical parameters (typically consisting of bond lengths, bond angles and dihedral angles) that describe the internal structure of the system [40].

Monte Carlo and Molecular Dynamics simulations may be performed in both sets of coordinates with different pros and cons. In internal coordinates (or in simple linear combinations of them), geometry optimizations typically converge faster [41–45] and efficient Monte Carlo movements are more easily designed [5, 46–49]. In Euclidean coordinates, on the other hand, the equations of motion and, specially, the Statistical Mechanics formulae are simpler and their coding in computer applications is more straightforward. But where these two sets of coordinates really differ is in the implementation of constraints [5,50–52]. In macromolecules, the natural constraints are those derived from the relative rigidity of the internal covalent structure of groups of atoms that share a common center (and also from the rigidity of rotation around double or triple bonds) compared to the energetically "cheaper" rotation around single bonds. In internal coordinates, these chemical constraints may be directly implemented by asking that some conveniently selected *hard* coordinates (normally, bond lengths, bond angles and some dihedrals) have constant values or values that depend on the remaining *soft* coordinates (see ref. [15] for a definition). In Euclidean coordinates, on the other hand, the expression of the constraints is more cumbersome and complicated procedures [25, 26, 33, 53–55] must be used at each timestep to implement them in Molecular Dynamics simulations. This is why, in the classical stiff model, as well as in the rigid one, it is common to use internal coordinates and they are also the choice throughout this work.

In the equilibrium Statistical Mechanics of both the stiff and rigid models, the marginal probability density in the coordinate part of the phase space in these internal coordinates is not proportional to the naive $\exp[-\beta V_\Sigma(q^i)]$, where $V_\Sigma(q^i)$ denotes the potential energy on the constrained hypersurface[2]. Instead, some correcting terms that come from different sources must be added to the potential energy $V_\Sigma(q^i)$ [13, 15, 18, 19, 32, 56, 57]. On one side, the integration over the hard coordinates in the stiff model produces an entropic term related to the determinant of the Hessian matrix of the constraining part of the potential and very similar to the *conformational entropies* appearing in quasiharmonic analysis [6, 58, 59]. On the other side, the necessary use of non-Euclidean coordinates produces the square root of the mass-metric tensor determinant, in the

[2]By q^i, we denote the soft internal coordinates of the system. See secs. 2.1 and 2.5 for a precise definition.

stiff case, and the square root of the determinant of the reduced mass-metric tensor on the constrained hypersurface, in the rigid case (the corresponding correcting terms may be called *kinetic entropies* [17], since, as will be discussed later, they come from averaging out the momenta). If Monte Carlo simulations in the coordinate space are to be performed [5, 46–49, 60] and the probability densities that correspond to any of these two models sampled, the corrections should be included or, otherwise, showed to be negligible.

Additionally, the three different correcting terms are involved in the definition of the so-called Fixman's compensating potential [16], which is frequently used to reproduce the stiff equilibrium distribution using rigid Molecular Dynamics simulations [9, 14, 19, 21–23, 31, 32, 35, 56].

Customarily in the literature, some of these corrections to the potential energy are assumed to be independent of the conformation and thus dropped from the basic expressions [15, 16, 19, 20, 22, 26, 34, 48, 56, 61, 62]. For certain simple systems subject to ad hoc designed potentials, the analytic expression of some of the correcting terms has been computed and, in some cases, shown to be non-negligible [13, 14, 21–23, 31, 60, 63]. For a serial polymer with fixed bond lengths and bond angles, Gō and Scheraga [15] showed that the determinant of the mass-metric tensor in the stiff case is independent of the soft coordinates. In the same system, Patriciu, Chirikjian and Pappu [20], very recently, measured the conformational dependence of the determinant of the reduced mass-metric tensor (actually, with the aim of studying the Fixman's compensating potential) and showed it to be non-negligible.

Also, subtly entangled to the assumptions underlying many classical results as the ones mentioned above, 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 [6, 13, 15, 16, 18–20, 26, 31–35, 48, 56, 57, 61–65]. This is not even so [15, 25, 38, 66] in the simple force fields [6, 27–30] customarily used to calculate the energy of macromolecules and, therefore, it is interesting to eliminate this assumption when studying the conformational dependence of the correcting terms.

In this work, we measure the conformational dependence of *all correcting terms* and of the Fixman's compensating potential in the model dipeptide HCO-L-Ala-NH_2 without any simplifying assumption. The potential energy function is considered to be the effective Born-Oppenheimer potential for the nuclei derived from ab initio quantum mechanical calculations including electron correlation at the MP2 level. We also repeat 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. It is *the first time*, as far as we are aware, that this type of study is performed in a relevant biomolecule with a realistic potential energy function.

In sec. 2, we derive the Statistical Mechanics formulae of the rigid and stiff models in the general case and we summarize the factorization of the external coordinates presented in ref. [67]. In sec. 2.5, we exhaustively review the use of the different approximations in the literature and we give a precise definition of

exactly and *approximately separable hard and soft coordinates* which will shed some light on the relation between the different types of simplifications aforementioned. In sec. 3, we describe the methods used to assess the approximation that consists of neglecting the different corrections to the potential energy in the model dipeptide HCO-L-Ala-NH_2, without any simplifying assumption, which is the central aim of this work. Sec. 4 is devoted to the presentation and discussion of the results obtained and the conclusions are summarized in sec. 5.

2 Theory

2.1 General conventions and definitions

First of all, it is convenient to introduce certain notational conventions that will be used extensively in the rest of the work:

- The superindex T indicates matrix transposition. By $\vec{a}^{\,T}$ we shall understand the row vector (a^1, a^2, a^3).
- The Einstein's sum convention is assumed on repeated indices.
- The time derivative of A will be denoted by an overdot: as in $\dot{A}$.
- The system under scrutiny will be a set of n mass points termed *atoms*. The Euclidean coordinates of the atom α in a set of axes fixed in space are denoted by $\vec{x}_\alpha$. The subscript α runs from 1 to n.
- To define the set of axes *fixed in the system*, we select three atoms (denoted by 1, 2 and 3) in such a way that $\vec{X}$ is the position of atom 1 (i.e., $\vec{x}_1 = \vec{X}$). The orientation of the fixed axes (x', y', z') is chosen such that atom 2 lies in the positive half of the z'-axis and atom 3 is contained in the (x', z')-plane, in the positive half of the x'-axis (see fig. 1). The position of atom α in these axes is denoted by $\vec{x}'_\alpha$.
- The components of trivectors, such as $\vec{x}_\alpha$ or $\vec{x}'_\alpha$, are denoted by x^p_α or x'^p_α, with $p = 1, 2, 3$.
- The curvilinear coordinates suitable to describe the system will be denoted by $q^\mu,\ \mu = 1, \ldots, 3n$ and the set of Euclidean coordinates by x^μ when no explicit reference to the atoms index needs to be made. We shall often use $N := 3n$ for the total number of degrees of freedom.
- We choose the coordinates q^μ so that the first six are *external coordinates*. They are denoted by q^A and their ordering is $q^A \equiv (X, Y, Z, \phi, \theta, \psi)$. The first three ones, $\vec{X}^T := (X, Y, Z)$, describe the overall position of the system. The three angles (ϕ, θ, ψ) are related to its overall orientation. More concretely, they give the orientation of the frame fixed in the system in fig. 1 with respect to the frame fixed in space.

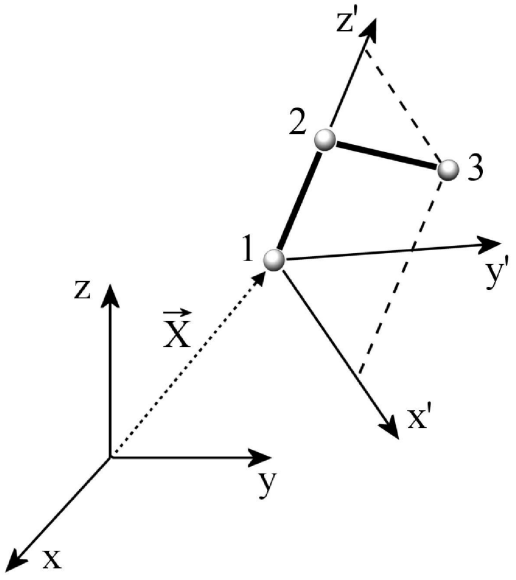

Figure 1: Definition of the axes fixed in the system.

- The coordinates q^μ are split into $(q^A, q^a),\ a = 7, \ldots, N$. The coordinates q^a are said *internal coordinates* and determine the positions of the atoms in the frame fixed in the system. The coordinates q^a parameterize what we shall call the *internal subspace* or *conformational space*, denoted by $\mathcal{I}$ and the coordinates q^A parameterize the *external subspace*, denoted by $\mathcal{E}$.

- The *general set-up* of the problem may be described as follows: Instead of us being interested on the conformational equilibrium of the system in the external subspace $\mathcal{E}$ plus the whole internal subspace $\mathcal{I}$ (i.e., the *whole space*, denoted by $\mathcal{E} \times \mathcal{I}$), we wish to find the probability density on a hypersurface $\Sigma \subset \mathcal{I}$ of dimension M (plus the external subspace $\mathcal{E}$), i.e., on $\mathcal{E} \times \Sigma$.

- In typical internal coordinates q^a, normally consisting of bond lengths, bond angles and dihedral angles (see ref. 50 and references therein), the hypersurface Σ is described via $L := N - M - 6$ constraints:

$$q^I = f^I(q^i) \qquad I = M + 7, \ldots, N \ , \tag{2.1}$$

  where the q^a are split into $q^a \equiv (q^i, q^I)$, and the q^i, $i = 7, \ldots, M + 6$, which parameterize Σ, are called *internal soft coordinates*, whereas the q^I are termed *hard coordinates*. The external coordinates q^A, together with the q^i, form the whole set of *soft coordinates*, denoted by $q^u \equiv (q^A, q^i)$, $u = 1, \ldots, M + 6$.

Finally, in table 1, a summary of the indices used is given.

Indices	Range	Number	Description
$\alpha, \beta, \gamma, \ldots$	$1, \ldots, n$	n	Atoms
$p, q, r, s, \ldots$	$1, 2, 3$	3	Components of trivectors
$\mu, \nu, \rho, \ldots$	$1, \ldots, N$	$N = 3n$	All coordinates
$A, B, C, \ldots$	$1, \ldots, 6$	6	External coordinates
$a, b, c, \ldots$	$7, \ldots, N$	$N - 6$	Internal coordinates
$i, j, k, \ldots$	$7, \ldots, M + 6$	M	Soft internal coordinates
$I, J, K, \ldots$	$M + 7, \ldots, N$	$L = N - M - 6$	Hard internal coordinates
$u, v, w, \ldots$	$1, \ldots, M + 6$	$M + 6$	All soft coordinates

Table 1: Definition of the indices used.

2.2 Classical stiff model

In the classical stiff model, the constraints in eq. (2.1) are implemented by imposing an strong energy penalization when the internal conformation of the system, described by q^a, departs from the constrained hypersurface Σ. To ensure this, we must have that the potential energy function in $\mathcal{I}$ satisfies certain conditions. First, we write the potential $V(q^a)$ as follows[3]:

$$V(q^i, q^I) = \underbrace{V\big(q^i, f^I(q^i)\big)}_{V_\Sigma(q^i)} + \underbrace{\Big[V(q^i, q^I) - V\big(q^i, f^I(q^i)\big)\Big]}_{V_{\mathrm{c}}(q^i, q^I)} \,. \tag{2.2}$$

Next, we impose the following conditions on the *constraining potential* $V_{\mathrm{c}}(q^i, q^I)$ defined above:

(i) That $V_{\mathrm{c}}\big(q^i, f^I(q^i)\big) \leq V_{\mathrm{c}}(q^i, q^I) \quad \forall q^i, q^I$, i.e., that Σ be the global minimum of V_{c} (and, henceforth, a local one too) with respect to variations of the hard coordinates.

(ii) That, for small variations Δq^I on the hard coordinates (i.e., for changes Δq^I considered as physically irrelevant), the associated changes in $V_{\mathrm{c}}(q^i, q^I)$ are much larger than the thermal energy RT.

The advantages of this formulation, much similar to that on [15], are many. First, it sets a convenient framework for the derivation of the Statistical Mechanics formulae of the classical stiff model relating it to the fully flexible model in the whole space $\mathcal{E} \times \mathcal{I}$. Second, it clearly separates the potential energy on Σ from the part that is responsible of implementing the constraints. Third, contrarily to the formulation based on delta functions [56], it allows to clearly

[3] Note that we have simply added and subtracted from the total potential energy $V(q^i, q^I) \equiv V(q^a)$ of the system the same quantity, $V\big(q^i, f^I(q^i)\big)$.

understand the necessity of including the correcting term associated to the determinant of the Hessian of $V_{\rm c}$ (see the derivation that follows). Finally, and more importantly for us, it provides a direct prescription for calculating $V_\Sigma(q^i)$ and Σ (the Potential Energy Surface (PES), frequently used in Quantum Chemistry calculations [68–72]) via geometry optimization at fixed values of the soft coordinates.

We also remark that, in order to satisfy point (ii) above and to allow the derivation of the different correcting terms that follows and the validity of the final expressions, the hard coordinates q^I must be indeed hard, however, the *soft coordinates* q^i do not have to be soft (in the sense that they produce energetic changes much smaller than RT when varied). They may be interesting for some other reason and hence voluntarily picked to describe the system studied, *without altering the formulae presented in this section.* Despite this qualifications, the terms *soft* and *hard* will be kept in this work for consistence with most of the existing literature [15, 18, 57, 61, 62], although, in some cases, the labels *important* and *unimportant* (for q^i and q^I respectively), proposed by Karplus and Kushick [59], may be more appropriate.

In the case of the model dipeptide HCO-L-Ala-NH_2 investigated in this work, for example, the barriers in the Ramachandran angles ϕ and ψ may be as large as $\sim 40\ RT$, however, the study of small dipeptides is normally aimed to the design of effective potentials for polypeptides [73–75], where long-range interactions in the sequence may compensate these local energy penalizations. This and the fact that the Ramachandran angles are the relevant degrees of freedom to describe the conformation of the backbone of these systems, make it convenient to choose them as *soft coordinates* q^i despite the fact that they may be energetically hard in the case of the dipeptide HCO-L-Ala-NH_2. As remarked above, this does not affect the calculations.

Now, due to condition (ii) above, the statistical weights of the conformations which lie far away from the constrained hypersurface Σ are negligible and, therefore, it suffices to describe the system in the vicinity of the equilibrium values of the q^I. In this region, for each value of the internal soft coordinates q^i, we may expand $V_{\rm c}(q^i, q^I)$ in eq. (2.2) up to second order in the hard coordinates around Σ (i.e., around $q^I = f^I(q^i)$) and drop the higher order terms:

$$\begin{aligned} V_{\rm c}(q^i, q^I) \;\simeq\; & V_{\rm c}\big(q^i, f^I(q^i)\big) + \left[\frac{\partial V_{\rm c}}{\partial q^J}\right]_\Sigma \big(q^J - f^J(q^i)\big) + \\ & + \frac{1}{2} \underbrace{\left[\frac{\partial^2 V_{\rm c}}{\partial q^J \partial q^K}\right]_\Sigma}_{\mathcal{H}_{JK}(q^i)} \big(q^J - f^J(q^i)\big)\big(q^K - f^K(q^i)\big) \;, \end{aligned} \tag{2.3}$$

where the subindex Σ indicates evaluation on the constrained hypersurface and a more compact notation, $\mathcal{H}(q^i)$, has been introduced for the Hessian matrix of $V_{\rm c}$ with respect to the hard variables evaluated on Σ.

In this expression, the zeroth order term $V_{\rm c}\big(q^i, f^I(q^i)\big)$ is zero by definition of $V_{\rm c}$ (see eq. (2.2)) and the linear term is also zero, because of the condition

(i) above. Hence, the first non-zero term of the expansion in eq. (2.3) is the second order one. Using this, together with eq. (2.2), we may write the *stiff Hamiltonian*

$$\begin{aligned} H_{\rm s}(q^\mu, p_\mu) \;\; &:= \;\; \frac{1}{2} p_\nu G^{\nu\rho}(q^u, q^I) p_\rho + V_\Sigma(q^i) + \\ &+ \frac{k}{2} \mathcal{H}_{JK}(q^i)\big(q^J - f^J(q^i)\big)\big(q^K - f^K(q^i)\big) \;, \end{aligned} \tag{2.4}$$

the *mass-metric tensor* $G_{\nu\rho}$ being

$$G_{\nu\rho}(q^u, q^I) := \sum_{\sigma=1}^{N} \frac{\partial x^\sigma(q^\mu)}{\partial q^\nu} m_\sigma \frac{\partial x^\sigma(q^\mu)}{\partial q^\rho} \tag{2.5}$$

and $G^{\nu\rho}$ its inverse, defined by

$$G^{\nu\sigma}(q^u, q^I)\, G_{\sigma\rho}(q^u, q^I) = \delta^\nu_\rho \;, \tag{2.6}$$

where δ^ν_ρ denotes the Kronecker's delta.
Therefore, the *stiff partition function* of the system is[4]

$$Z_{\rm s} = \frac{\alpha_{QM}}{h^N} \int {\rm d}q^\mu {\rm d}p_\mu \exp\big[-\beta H_{\rm s}(q^\mu, p_\mu)\big] \;, \tag{2.7}$$

where h is Planck's constant, we denote $\beta := 1/RT$ (per mole energy units are used throughout the article, so RT is preferred over $k_B T$) and α_{QM} is a combinatorial number that accounts for quantum indistinguishability and that must be specified in each particular case (e.g., for a gas of N indistinguishable particles, $\alpha_{QM} = 1/N!$).

Now, using the condition (ii) again, the q^I appearing in the mass-metric tensor G in $H_{\rm s}$ (in eq. (2.7)) can be approximately evaluated at their equilibrium values $f^I(q^i)$, yielding, for the stiff partition function,

$$\begin{aligned} Z_{\rm s} = \frac{\alpha_{QM}}{h^N} \int {\rm d}q^u {\rm d}q^I {\rm d}p_\mu \exp\Bigg[-\beta\Bigg(&\frac{1}{2} p_\nu G^{\nu\rho}\big(q^u, f^I(q^i)\big) p_\rho + \\ &+ V_\Sigma(q^i) + \frac{1}{2}\mathcal{H}_{JK}(q^i)\big(q^J - f^J(q^i)\big)\big(q^K - f^K(q^i)\big)\Bigg)\Bigg] \;. \end{aligned} \tag{2.8}$$

If we now integrate over the hard coordinates q^I, we have

$$\begin{aligned} Z_{\rm s} = \left(\frac{2\pi}{\beta}\right)^{\frac{L}{2}} \frac{\alpha_{QM}}{h^N} \int {\rm d}q^u {\rm d}p_\mu \exp\Bigg[-\beta\Bigg(&\frac{1}{2} p_\nu G^{\nu\rho}\big(q^u, f^I(q^i)\big) p_\rho + \\ &+ V_\Sigma(q^i) + T\frac{R}{2} \ln\Big[\det \mathcal{H}(q^i)\Big]\Bigg)\Bigg] \;. \end{aligned} \tag{2.9}$$

[4]No Jacobian appears in the integral measure because q^μ and p_μ are obtained from the Euclidean coordinates via a canonical transformation [76].

where the part of the result of the Gaussian integral consisting of $\det^{-1/2}\mathcal{H}$ has been taken to the exponent.

It is also frequent to integrate over the momenta in the partition function. Doing this in eq. (2.9) and taking the determinant of the mass-metric tensor that shows up[5] to the exponent, we may write the partition function as an integral only on the coordinates:

$$Z_{\mathrm{s}} = \chi_{\mathrm{s}}(T) \int \mathrm{d}q^u \exp\Bigg[-\beta\Big(V_\Sigma(q^i) + \; T\frac{R}{2} \ln\Big[\det \mathcal{H}(q^i)\Big] - \\ -T\frac{R}{2} \ln\Big[\det G\big(q^u, f^I(q^i)\big)\Big]\Big)\Bigg] \;, \tag{2.10}$$

where the multiplicative factor that depends on T has been defined as follows:

$$\chi_{\mathrm{s}}(T) := \left(\frac{2\pi}{\beta}\right)^{\frac{N+L}{2}} \frac{\alpha_{QM}}{h^N} \;. \tag{2.11}$$

If the exponent in eq. (2.10) is seen as a free energy, then, $V_\Sigma(q^i)$ may be regarded as the internal energy and the two conformation-dependent correcting terms that are added to it as effective entropies (which is compatible with their being linear in RT). The second one comes only from the desire to write the marginal probabilities in the coordinate space (i.e., averaging the momenta) and may be called a *kinetic entropy* [17], the first term, on the other hand, is truly an entropic term that comes from the averaging out of certain degrees of freedom and it is reminiscent of the *conformational* or *configurational entropies* appearing in quasiharmonic analysis [6, 58, 59].

In this spirit, we define

$$F_{\mathrm{s}}(q^u) := V_\Sigma(q^i) - T\big(S^c_{\mathrm{s}}(q^i) + S^{\mathrm{k}}_{\mathrm{s}}(q^u)\big) \;, \tag{2.12a}$$

$$S^c_{\mathrm{s}}(q^i) := -\frac{R}{2} \ln\Big[\det \mathcal{H}(q^i)\Big] \;, \tag{2.12b}$$

$$S^{\mathrm{k}}_{\mathrm{s}}(q^u) := \frac{R}{2} \ln\Big[\det G\big(q^u, f^I(q^i)\big)\Big] \;. \tag{2.12c}$$

In such a way that the *stiff equilibrium probability* in the soft subspace $\mathcal{E} \times \Sigma$ is given by

$$P_{\mathrm{s}}(q^u) = \frac{\exp\big[-\beta F_{\mathrm{s}}(q^u)\big]}{Z'_{\mathrm{s}}} \;, \quad \text{with} \quad Z'_{\mathrm{s}} := \int \mathrm{d}q^u \exp\big[-\beta F_{\mathrm{s}}(q^u)\big] \;. \tag{2.13}$$

Now, it is worth remarking that, although the kinetic entropy $S^{\mathrm{k}}_{\mathrm{s}}$ depends on the external coordinates q^A, we have recently shown [67] that the determinant

[5] Note that, by G, we denote the matrix that corresponds to the mass-metric tensor with two covariant indices $G_{\mu\nu}$. The same convention has been followed for the Hessian matrix $\mathcal{H}$ in eq. 2.9 and for the reduced mass-metric tensor g in eq. 2.21.

of the mass-metric tensor G may be written, for any molecule, general internal coordinates and arbitrary constraints, as a product of two functions; one depending only on the external coordinates, and the other only on the internal ones q^a. Hence the externals-dependent factor in eq. (2.12c) may be integrated out independently to yield an effective free energy and a probability density $P_{\rm s}$ that depend only on the soft internals q^i (see sec. 2.4).

2.3 Classical rigid model

If the relations in eq. (2.1) are considered to hold *exactly* and are treated as holonomic constraints, the Hamiltonian function that describes the Classical Mechanics in the subspace $(\mathcal{E} \times \Sigma) \subset (\mathcal{E} \times \mathcal{I})$, spanned by the coordinates q^u, may be written as follows:

$$H_{\rm r}(q^u, \eta_u) := \frac{1}{2}\eta_v g^{vw}(q^u)\eta_w + V_\Sigma(q^i) \ , \tag{2.14}$$

where the *reduced mass-metric tensor* (also called *induced* in the mathematical literature [77]) $g_{vw}(q^u)$ in $\mathcal{E} \times \Sigma$, that appears in the kinetic energy, is (see what follows)

$$\begin{aligned} g_{vw}(q^u) &= G_{vw}\big(q^u, f^I(q^i)\big) + \frac{\partial f^J(q^i)}{\partial q^v} G_{JK}\big(q^u, f^I(q^i)\big) \frac{\partial f^K(q^i)}{\partial q^w} \ + \\ &+ G_{vK}\big(q^u, f^I(q^i)\big) \frac{\partial f^K(q^i)}{\partial q^w} + \frac{\partial f^J(q^i)}{\partial q^v} G_{Jw}\big(q^u, f^I(q^i)\big) := \\ &= \frac{\partial \tilde{f}^\mu}{\partial q^v} G_{\mu\nu}\big(q^u, f^I(q^i)\big) \frac{\partial \tilde{f}^\nu}{\partial q^w} \ , \end{aligned} \tag{2.15}$$

and $g^{vw}(q^u)$ is defined to be its inverse in the sense of eq. (2.6). Also, the notation

$$\tilde{f}^\mu := \begin{cases} q^u & \text{if} \quad u := \mu = 1, \dots, M+6 \\ f^I(q^i) & \text{if} \quad I := \mu = M+7, \dots, N \end{cases} \tag{2.16}$$

has been introduced for convenience.

Note that eq. (2.15) may derived from the unconstrained Hamiltonian in $(\mathcal{E} \times \mathcal{I})$,

$$H(q^\mu, p_\mu) := \frac{1}{2} p_\nu G^{\nu\rho}(q^\mu) p_\rho + V(q^a) \ , \tag{2.17}$$

using the constraints in eq. (2.1), together with its time derivatives:

$$\dot{q}^I := \frac{\partial f^I(q^i)}{\partial q^j} \dot{q}^j \tag{2.18}$$

and defining the momenta η_v as

$$\eta_v := g_{vw}(q^u)\,\dot{q}^w = g_{vw}(q^u)\,G^{w\mu}\big(q^u, f^I(q^i)\big)\,p_\mu \ . \tag{2.19}$$

Hence, the *rigid partition function* is

$$Z_{\mathrm{r}} = \frac{\alpha_{QM}}{h^{M+6}} \int \mathrm{d}q^u \mathrm{d}\eta_u \exp\left[-\beta\left(\frac{1}{2}\eta_v g^{vw}(q^u)\eta_v + V_\Sigma(q^i)\right)\right] . \tag{2.20}$$

Integrating over the momenta, we obtain the marginal probability density in the coordinate space analogous to eq. (2.10):

$$Z_{\mathrm{r}} = \chi_{\mathrm{r}}(T) \int \mathrm{d}q^u \exp\Big[-\beta\Big(V_\Sigma(q^i) - T\frac{R}{2}\ln\Big[\det g(q^u)\Big]\Big)\Big] \ , \tag{2.21}$$

where

$$\chi_{\mathrm{r}}(T) := \left(\frac{2\pi}{\beta}\right)^{\frac{M+6}{2}} \frac{\alpha_{QM}}{h^{\frac{M+6}{2}}} \ . \tag{2.22}$$

Repeating the analogy with free energies and entropies in the last paragraphs of the previous subsection, we define

$$F_{\mathrm{r}}(q^u) := V_\Sigma(q^i) - T S^{\mathrm{k}}_{\mathrm{r}}(q^u) \ , \tag{2.23a}$$

$$S^{\mathrm{k}}_{\mathrm{r}}(q^u) := \frac{R}{2}\ln\Big[\det g(q^u)\Big] \ , \tag{2.23b}$$

being the *rigid equilibrium probability* in the soft subspace $\mathcal{E}\times\Sigma$

$$P_{\mathrm{r}}(q^u) = \frac{\exp\big[-\beta F_{\mathrm{r}}(q^u)\big]}{Z'_{\mathrm{r}}} \ , \quad \text{with} \quad Z'_{\mathrm{r}} := \int \mathrm{d}q^u \exp\big[-\beta F_{\mathrm{r}}(q^u)\big] \ . \tag{2.24}$$

As in the case of G, we have shown in ref. [67] that the determinant of the reduced mass-metric tensor g may be written, for any molecule, general internal coordinates and arbitrary constraints, as a product of two functions; one depending only on the external coordinates, and the other only on the internal ones q^i. Hence the externals-dependent factor in $\det g(q^u)$ may be integrated out independently to yield a free energy and a probability density P_{r} that depend only on the soft internals q^i (see the following subsection).

To end this subsection, we remark that it is frequent in the literature [9,18,19, 21–23,31,35,56,60] to define the so-called *Fixman's compensating potential* [16] as the difference between $F_{\mathrm{s}}(q^u)$, in eq. (2.12), and $F_{\mathrm{r}}(q^u)$, defined above, i.e.,

$$\begin{aligned} V_{\mathrm{F}}(q^u) \ &:= \ T S^{\mathrm{k}}_{\mathrm{r}}(q^u) - T S^{c}_{\mathrm{s}}(q^i) - T S^{\mathrm{k}}_{\mathrm{s}}(q^u) = \\ &= \ \frac{RT}{2}\ln\left[\frac{\det G(q^u)}{\det \mathcal{H}(q^i)\det g(q^u)}\right] . \end{aligned} \tag{2.25}$$

Hence, performing rigid Molecular Dynamics simulations, which would yield an equilibrium distribution proportional to $\exp[-\beta F_{\mathrm{r}}(q^u)]$, and adding $V_{\mathrm{F}}(q^u)$ to the potential energy $V_\Sigma(q^i)$ one can reproduce instead the stiff probability density $P_{\mathrm{s}} \propto \exp[-\beta F_{\mathrm{s}}(q^u)]$ [14, 18, 19, 21–23, 31, 32, 35, 56]. This allows to obtain at a lower computational cost (due to the timescale problems discussed in the introduction) equilibrium averages that otherwise must be extracted from expensive fully flexible whole-space simulations. In fact, it seems that this particular application of the theoretical tools herein described, and not the search for the correct probability density to sample in Monte Carlo simulations, was what prompted the interest in the study of mass-metric tensors effects.

2.4 Factorization of the external coordinates

In the recent work [67], we have shown that the determinant of the mass-metric tensor G in eq. (2.12c) can be written as follows if the SASMIC [50] coordinates for general branched molecules are used:

$$\det G = \left(\prod_{\alpha=1}^{n} m_\alpha^3 \right) \sin^2\theta \left(\prod_{\alpha=2}^{n} r_\alpha^4 \right) \left(\prod_{\alpha=3}^{n} \sin^2\theta_\alpha \right) \ , \tag{2.26}$$

where the r_α are bond lengths and the θ_α bond angles.

Note that this expression, whose validity was proved for the more particular case of serial polymers by Gō and Scheraga [15] and, before, by Volkenstein [78], does not explicitly depend on the dihedral angles. However, it may depend on them via the hard coordinates if the constraints in the form presented in eq. (2.1) are used.

The term depending on the masses of the atoms in the expression above may be dropped from eq. (2.12c), because it does not depend on the conformation, and the only part of $\det G$ that depend on the external coordinates, $\sin^2\theta$, may be integrated out in eq. (2.10). Hence, the kinetic entropy due to the mass-metric tensor G in the stiff case, may be written, up to additive constants, as

$$S_{\mathrm{s}}^{\mathrm{k}}(q^i) = \frac{R}{2} \left[\sum_{\alpha=2}^{n} \ln\left(r_\alpha^4\right) + \sum_{\alpha=3}^{n} \ln\left(\sin^2\theta_\alpha\right) \right] \ , \tag{2.27}$$

where the individual contributions of each degree of freedom have been factorized.

Also in reference [67], we have shown that the determinant of the reduced mass-metric tensor g in eq. (2.23b) can be written as follows:

$$\det g = \sin^2\theta \ \det g_2(q^i) \ , \tag{2.28}$$

being the matrix g_2

$$g_2 = \left(\begin{array}{cc|c} m_{tot} I^{(3)} & m_{tot}\, v(\vec{R}) & \cdots \quad m_{tot}\dfrac{\partial \vec{R}}{\partial q^j} \quad \cdots \\ m_{tot}\, v^T(\vec{R}) & \mathcal{J} & \cdots \sum\limits_\alpha m_\alpha \dfrac{\partial \vec{x}'_\alpha}{\partial q^j} \times \vec{x}'_\alpha \cdots \\ \hline \vdots & \vdots & \vdots \\ m_{tot}\dfrac{\partial \vec{R}}{\partial q^i} & \sum\limits_\alpha m_\alpha \left(\dfrac{\partial \vec{x}'_\alpha}{\partial q^i} \times \vec{x}'_\alpha\right)^T & \cdots \sum\limits_\alpha m_\alpha \dfrac{\partial \vec{x}'^{\,T}_\alpha}{\partial q^i}\dfrac{\partial \vec{x}'_\alpha}{\partial q^j} \cdots \\ \vdots & \vdots & \vdots \end{array}\right) \tag{2.29}$$

and denoting the *total mass* of the system by $m_{tot} := \sum_\alpha m_\alpha$, the position of the *center of mass* of the system in the primed reference frame (see sec. 2.1) by $\vec{R} := m_{tot}^{-1} \sum_\alpha m_\alpha \vec{x}'_\alpha$ and the *inertia tensor* of the system, also in the primed reference frame, by

$$\mathcal{J} := \begin{pmatrix} \sum_\alpha m_\alpha ((x'^2_\alpha)^2 + (x'^3_\alpha)^2) & -\sum_\alpha m_\alpha x'^1_\alpha x'^2_\alpha & -\sum_\alpha m_\alpha x'^1_\alpha x'^3_\alpha \\ -\sum_\alpha m_\alpha x'^1_\alpha x'^2_\alpha & \sum_\alpha m_\alpha ((x'^1_\alpha)^2 + (x'^3_\alpha)^2) & -\sum_\alpha m_\alpha x'^2_\alpha x'^3_\alpha \\ -\sum_\alpha m_\alpha x'^1_\alpha x'^3_\alpha & -\sum_\alpha m_\alpha x'^2_\alpha x'^3_\alpha & \sum_\alpha m_\alpha ((x'^1_\alpha)^2 + (x'^2_\alpha)^2) \end{pmatrix}. \tag{2.30}$$

The matrix $v(\vec{R})$ is defined as:

$$v(\vec{R}) := \begin{pmatrix} 0 & -R^3 & R^2 \\ R^3 & 0 & -R^1 \\ -R^2 & R^1 & 0 \end{pmatrix} \tag{2.31}$$

and $\times$ denotes the usual vector cross product.

Then, since $\sin^2\theta$ may be integrated out in eq. (2.21), we can write, omitting additive constants, the kinetic entropy associated to the reduced mass-metric tensor g depending only on the soft internals q^i:

$$S^{\mathrm{k}}_{\mathrm{r}}(q^i) = \frac{R}{2} \ln \left[\det g_2(q^i)\right] . \tag{2.32}$$

Finally, one may note that, since $\sin^2\theta$ divides out in the second line of eq. (2.25) or, otherwise stated, eqs. (2.27) and (2.32) may be introduced in the first line, then the Fixman's potential is independent of the external coordinates as well.

2.5 Usual approximations

Many approximations may be done to simplify the calculation of the different correcting terms introduced in the previous subsections. The most frequently found in the literature are the following three:

(i) To neglect the conformational dependence of $\det G$.

(ii) To neglect the conformational dependence of $\det \mathcal{H}$.

(iii) To assume that the hard coordinates are constant, i.e, that the $f^I(q^i)$ in eq. (2.1) do not depend on the soft coordinates q^i.

The conformational dependence of $\det g$ is customarily regarded as important since it was shown to be non-negligible even for simple systems some decades ago [13, 21–23] (normally in an indirect way, while studying the influence of the Fixman's compensating potential in eq. (2.25); see discussion below). With this same aim, Patriciu et al. [20] have very recently measured the conformational dependence of $\det g$ for a serial polymer with fixed bond lengths and bond angles (in the approximation (iii)), showing that it is non-negligible and suggesting that it may be so also for more general systems.

Note that, if approximations (i) and (ii) are assumed, then the Fixman's potential depends only on $\det g$. In fact, whereas in the general case the Fixman's compensating potential cannot be simplified beyond the expression in eq. (2.25), if one assumes approximation (iii), then the reduced mass-metric tensor g turns out to be the subblock of G with soft indices and, in this case, the quotient $\det G / \det g$ has been shown to be equal to $1/\det h$ by Fixman [16], where h denotes the subblock of G^{-1} with hard indices, i.e.,

$$h^{IJ}(q^\mu) := \sum_{\sigma=1}^{N} \frac{\partial q^I}{\partial x^\sigma} \frac{1}{m_\sigma} \frac{\partial q^J}{\partial x^\sigma} \ . \tag{2.33}$$

This result has been extensively used in the literature [21–23,31,35,60], since each of the internal coordinates q^a typically used in macromolecular simulations only involves a small number of atoms, thus rendering the matrix h above sparse and allowing for efficient algorithms to be used in order to find its determinant.

Now, although $\det g$ is customarily regarded as important, the conformational variations of $\det G$ are almost unanimously neglected (approximation (i)) in the literature [15, 48] and may only be said to be indirectly included in h by the authors that use the expression above [20–23, 31, 60]. This is mainly due to the fact, reported by Gō and Scheraga [15] and, before, by Volkenstein [78], that $\det G$ in a serial polymer may be expressed as in eq. (2.26), being independent of the dihedral angles (which are customarily taken as the soft coordinates). If one also assumes approximation (iii), which, as will be discussed later, is very common, then $\det G$ is a constant for every conformation of the molecule.

Probably due to computational considerations, but also sometimes to the use of a formulation of the stiff case based on delta functions [56], the conformational dependence of $\det \mathcal{H}$ is almost unanimously neglected (approximation (ii)) in the literature [15, 16, 19, 20, 34, 48, 61, 62]. Only a few authors include this term in different stages of the reasoning [13–15,18,19,26,32], most of them only to argue later that it is negligible.

Although for some simple ad hoc designed potentials that lack long-range terms [21,22,60], the aforementioned simplifying assumptions and the ones that will be discussed in the following paragraphs may be exactly fulfilled, in the

case of the potential energies used in force fields for macromolecular simulation [6, 27–30], they are not. The typical energy function in this case, has the form

$$\begin{aligned} V_{\rm ff}(q^a) &:= \frac{1}{2}\sum_{\alpha=1}^{N_r} K_{r_\alpha}(r_\alpha - r_\alpha^0)^2 + \frac{1}{2}\sum_{\alpha=1}^{N_\theta} K_{\theta_\alpha}(\theta_\alpha - \theta_\alpha^0)^2 + \\ &+ V_{\rm ff}^{\rm tors}(\phi_\alpha) + V_{\rm ff}^{\rm long-range}(q^a) \ , \end{aligned} \tag{2.34}$$

where r_α are bond lengths, θ_α are bond angles, ϕ_α are dihedral angles and, 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 ω). N_r is the number of bond lengths, N_θ the number of bond angles and the quantities K_{r_α}, K_{θ_α}, r_α^0 and θ_α^0 are constants. The term denoted by $V_{\rm ff}^{\rm tors}(\phi_\alpha)$ is a commonly included torsional potential that depends only on the dihedral angles ϕ_α and $V_{\rm ff}^{\rm long-range}(q^a)$ normally comprises long-range interactions such as Coulomb or van der Waals; hence, it depends on the atomic positions $\vec{x}'_\alpha$ which, in turn, depend on all the internal coordinates q^a.

One of the reasons given for neglecting $\det \mathcal{H}$, when classical force fields are used with potential energy functions such as the one in eq. (2.34), is that the harmonic constraining terms dominate over the rest of interactions and, since the constants appearing on these terms (the K_{r_α}, K_{θ_α} in eq. (2.34)) are independent of the conformation by construction, so is $\det \mathcal{H}$ [15, 19, 26]. Here, we analyze a more realistic quantum-mechanical potential and these considerations are not applicable, however, *they also should be checked in the case of classical force fields*, since, for a potential energy such as the one in eq. (2.34)), the quantities K_{r_α} and K_{θ_α} are finite and the long-range terms will also affect the Hessian at each point of the constrained hypersurface Σ, rendering its determinant *conformation-dependent*.

For the same reason, *even in classical force fields, the equilibrium values of the hard coordinates are not the constant quantities* r_α^0 *and* θ_α^0 in eq. (2.34) but some functions $f^I(q^i)$ of the soft coordinates (see eq. (2.1)). This fact, recognized by some authors [15, 25, 38, 66], provokes that, if one chooses to assume approximation (iii) and the constants r_α^0 and θ_α^0 appearing in eq. (2.34) are designated as the equilibrium values, the potential energy in Σ may be heavily distorted, the cause being simply that the long-range interactions between atoms separated by three covalent bonds are not fully relaxed [66]. This effect is probably larger if bond angles, and not only bond lengths, are also constrained, which may partially explain the different dynamical behaviour found in ref. 6 when comparing these types of constraints in Molecular Dynamics simulations. In quantum mechanical calculations of small dipeptides, on the other hand, the fact that the bond lengths and bond angles depend on the Ramachandran angles (ϕ, ψ) has been pointed out by Schäffer et al. [79]. Therefore, approximation (iii), which is very common in the literature [6, 13, 15, 16, 18–20, 26, 31–35, 48, 56, 57, 61, 62, 64, 65], should be critically analyzed in each particular case.

Apart from the typical internal coordinates q^a used until now, in terms of which the constrained hypersurface Σ is described by the relations $q^I = f^I(q^i)$ in eq. (2.1), with $I = M+7, \ldots, N$, one may define a different set Q^a such that, on Σ, the corresponding hard coordinates are arbitrary constants $Q^I = C^I$ (the external coordinates q^A and Q^A are irrelevant for this part of the discussion). To do this, for example, let

$$\begin{array}{lll} Q^i := q^i & i = 7, \ldots, M+6 & \text{and} \\ Q^I := q^I - f^I(q^i) + C^I & I = M+7, \ldots, N\ . & \end{array} \tag{2.35}$$

Well then, while the relation between bond lengths, bond angles and dihedral angles (the typical q^a [50]) and the Euclidean coordinates is straightforward and simple, the expression of the transformation functions $Q^a(x^\mu)$ needs the knowledge of the f^I, which must be calculated numerically in most real cases. This drastically reduce the practical use of the Q^a, however, it is also true that they are conceptually appealing, since they have a property that closely match our intuition about what the soft and hard coordinates should be (namely, that the hard coordinates Q^I are constant on the relevant hypersurface Σ); and this is why we term them *exactly separable hard and soft coordinates*. Now, we must also point out that, although the real internal coordinates q^a do not have this property, they are usually close to it. The customary labeling of soft and hard coordinates in the literature is based on this circumstance. Somehow, the dihedral angles are the "softest" of the internal coordinates, i.e., the ones that "vary the most" when the system visits different regions of the hypersurface Σ; and this is why we term the real q^a *approximately separable hard and soft coordinates*, considering approximation (iii) as a useful reference case.

To sum up, the three simplifying assumptions (i), (ii) and (iii) in the beginning of this section should be regarded as approximations in the case of classical force fields, as well as in the case of the more realistic quantum-mechanical potential investigated in this work, and they should be critically assessed in the systems of interest. In the following sections, while studying the model dipeptide HCO-L-Ala-NH_2 (see fig. 2), no simplifying assumptions of this type are made.

3 Methods

In the particular molecule treated in this work (the model dipeptide HCO-L-Ala-NH_2 in fig. 2), the formulae in the preceding sections must be used with $M = 2$, being the internal soft coordinates $q^i \equiv (\phi, \psi)$ the typical Ramachandran angles [80] (see table 2), the total number of coordinates $N = 48$ and the number of hard internals $L = 40$.

Regarding the side chain angle χ, it has been argued elsewhere [50] that it is soft with the same right as the angles ϕ and ψ, i.e., the barriers that hinder the rotation on this dihedral are comparable to the ones existing in the Ramachandran surface. However, the height of these barriers is sufficient ($\sim$ 6-12 RT, see ref. [50]) for the condition (ii) in sec. 2.2 to hold and, therefore, its inclusion in

the set of hard coordinates is convenient due to its *unimportant* character (see discussion in sec. 2.2). Moreover, to describe the behaviour associated to χ with a probability density different from a Gaussian distribution (i.e., its potential energy different from an harmonic oscillator), for example with the tools used in the field of circular statistics [81–83], would severely complicate the derivation of the classical stiff model without adding any conceptual insight to the problem. In addition, although χ is a periodic coordinate with threefold symmetry, the considerable height of the barriers between consecutive minima allows to make the quadratic assumption in eq. (2.3) at each equivalent valley and permits the approximation of the integral on χ by three times a Gaussian integral. The multiplicative factor 3 simply adds a temperature- and conformation-independent reference to the configurational entropy S_s^c in eq. (2.12b).

The same considerations are applied to the dihedral angles, ω_0 and ω_1 (see table 2), that describe the rotation around the peptide bond, and the quadratic approximation described above can also be used, since the heights of the rotation barriers around these degrees of freedom are even larger than the ones in the case of χ.

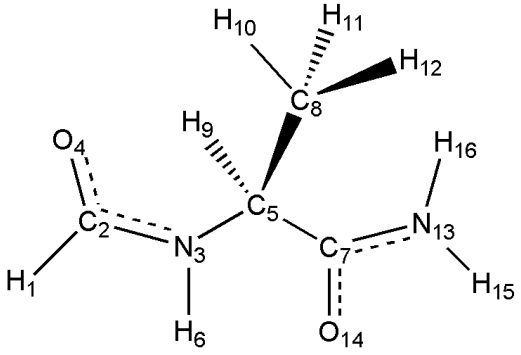

Figure 2: Atom numeration of the protected dipeptide HCO-L-Ala-NH_2.

The ab initio quantum mechanical calculations have been done with the package GAMESS [84] under Linux and in 3.20 GHz PIV machines. The coordinates used for the HCO-L-Ala-NH_2 dipeptide in the GAMESS input files and the ones used to generate them with automatic Perl scripts are the SASMIC coordinates introduced in ref. 50. They are presented in table 2 indicating the name of the conventional dihedral angles (see also fig. 2 for reference). To perform the energy optimizations, however, they have been converted to Delocalized Coordinates [44] in order to accelerate convergence.

First, we have calculated the typical Potential Energy Surface (PES) in a regular 12x12 grid of the bidimensional space spanned by the Ramachandran angles ϕ and ψ, with both angles ranging from -165^o to 165^o in steps of 30^o. This has been done by running constrained energy optimizations at the MP2/6-31++G(d,p) level of the theory, freezing the two Ramachandran

Atom name	Bond length	Bond angle	Dihedral angle
H_1			
C_2	(2,1)		
N_3	(3,2)	(3,2,1)	
O_4	(4,2)	(4,2,1)	(4,2,1,3)
C_5	(5,3)	(5,3,2)	ω_0 :=**(5,3,2,1)**
H_6	(6,3)	(6,3,2)	(6,3,2,5)
C_7	(7,5)	(7,5,3)	ϕ :=**(7,5,3,2)**
C_8	(8,5)	(8,5,3)	(8,5,3,7)
H_9	(9,5)	(9,5,3)	(9,5,3,7)
H_{10}	(10,8)	(10,8,5)	χ :=**(10,8,5,3)**
H_{11}	(11,8)	(11,8,5)	(11,8,5,10)
H_{12}	(12,8)	(12,8,5)	(12,8,5,10)
N_{13}	(13,7)	(13,7,5)	ψ :=**(13,7,5,3)**
O_{14}	(14,7)	(14,7,5)	(14,7,5,13)
H_{15}	(15,13)	(15,13,7)	ω_1 :=**(15,13,7,5)**
H_{16}	(16,13)	(16,13,7)	(16,13,7,15)

Table 2: SASMIC internal coordinates (Echenique P. and Alonso J. L., *To be published in J. Comp. Chem.*, `arXiv:q-bio.BM/0511004`) in Z-matrix form of the protected dipeptide HCO-L-Ala-NH_2. Principal dihedrals are indicated in bold face and their typical biochemical name is given.

angles at each value of the grid, starting from geometries previously optimized at a lower level of the theory and setting the gradient convergence criterium to `OPTTOL`=10^{-5} and the self-consistent Hartree-Fock convergence criterium to `CONV`=10^{-6}.

The results of these calculations (which took $\sim$ 100 days of CPU time) are 144 conformations that define Σ and the values of $V_\Sigma(\phi,\psi)$ at these points (the PES itself).

Then, at each optimized point of Σ, we have calculated the Hessian matrix in the coordinates of table 2 removing the rows and columns corresponding to the soft angles ϕ and ψ, the result being the matrix $\mathcal{H}(\phi,\psi)$ in eq. (2.12b). This has been done, again, at the MP2/6-31++G(d,p) level of the theory, taking $\sim$ 140 days of CPU time.

Eqs. (2.27) and (2.32) in sec. 2.4 have been used to calculate the kinetic entropy terms associated to the determinants of the mass-metric tensors G and g, respectively. The quantities in eq. (2.27), being simply internal coordinates, have been directly extracted from the GAMESS output files via automated Perl scripts. On the other hand, in order to calculate the matrix g_2 in eq. (2.29) that appears in the kinetic entropy of the classical rigid model, the Euclidean coordinates $\vec{x}'_\alpha$ of the 16 atoms in the primed reference frame defined in sec. 2.1, as well as their derivatives with respect to $q^i \equiv (\phi,\psi)$, must be computed. For this, two additional 12x12 grids as the one described above have been computed; one

of them displaced 2^o in the positive ϕ-direction and the other one displaced 2^o in the positive ψ-direction. This has been done, again, at the MP2/6-31++G(d,p) level of the theory, starting from the optimized structures found in the computation of the PES described above and taking ~ 75 days of CPU time each grid. Using the values of the positions $\vec{x}'_\alpha$ in these two new grids and also in the original one, the derivatives of these quantities with respect to the angles ϕ and ψ, appearing in g_2, have been numerically obtained as finite differences.

The three calculations have been repeated for six special points in the Ramachandran space that correspond to important elements of secondary structure (see sec. 4), the total CPU time needed for computing all correcting terms at these points has been ~ 16 days. A total of ~ 406 days of CPU time has been needed to perform the whole study at the MP2/6-31++G(d,p) level of the theory.

Finally, we have repeated all the calculations at the HF/6-31++G(d,p) level of the theory in order to investigate if this less demanding method (~ 10 days for the PES, ~ 8 days for the Hessians, ~ 10 days for each displaced grid, ~ 2 days for the special secondary structure points, being a total of ~ 40 days of CPU time) may be used instead of MP2 in further studies.

4 Results

In table 3, the maximum variation, the average and the standard deviation in the 12x12 grid defined in the Ramachandran space of the protected dipeptide HCO-L-Ala-NH_2 are shown for the three energy surfaces, V_Σ, F_s and F_r (see eqs. (2.12) and (2.23)), for the three correcting terms, $-TS^\mathrm{k}_\mathrm{s}$, $-TS^\mathrm{c}_\mathrm{s}$, and $-TS^\mathrm{k}_\mathrm{r}$ and for the Fixman's compensating potential V_F (see eq. (2.25)). All the functions have been referenced to zero in the grid.

In fig. 3, the Potential Energy Surface V_Σ, at the MP2/6-31++G(d,p) level of the theory, is depicted with the reference set to zero for visual convenience[6]. Neither the surfaces defined by F_s and F_r at the MP2/6-31++G(d,p) level of the theory nor the three energy surfaces V_Σ, F_s and F_r at HF/6-31++G(d,p) are shown graphically since they are visually very similar to the surface in fig. 3.

In fig. 4, the three correcting terms, $-TS^\mathrm{k}_\mathrm{s}$, $-TS^\mathrm{c}_\mathrm{s}$ and $-TS^\mathrm{k}_\mathrm{r}$ and the Fixman's compensating potential V_F, at the MP2/6-31++G(d,p) level of the theory, are depicted with the reference set to zero. The analogous surfaces at the HF/6-31++G(d,p) level of the theory are visually very similar to the ones in fig. 4 and have been therefore omitted.

From the results presented, one may conclude that, although the conformational dependence of the correcting terms $-TS^\mathrm{k}_\mathrm{s}$, $-TS^\mathrm{c}_\mathrm{s}$ and $-TS^\mathrm{k}_\mathrm{r}$ is more than an order of magnitude smaller than the conformational dependence of the Potential Energy Surface V_Σ in the worst case, if *chemical accuracy* (typically defined in the field of ab initio quantum chemistry as 1 kcal/mol [85]) is sought, they may be relevant. In fact, they are of the order of magnitude of the differences

[6] At the level of the theory used in the calculations, the minimum of $V_\Sigma(\phi, \psi)$ in the grid is -416.0733418995 hartree.

	MP2/6-31++G(d,p)			HF/6-31++G(d,p)		
	Max.[a]	Ave.[b]	Std.[c]	Max.[a]	Ave.[b]	Std.[c]
V_Σ	21.64	6.76	3.88	23.62	6.92	4.35
F_s	21.43	6.47	3.93	23.78	7.17	4.38
F_r	21.09	6.46	3.82	23.09	6.76	4.31
$-TS_\mathrm{s}^\mathrm{k}$	0.24	0.09	0.05	0.23	0.09	0.04
$-TS_\mathrm{s}^\mathrm{c}$	1.67	0.98	0.32	1.34	0.63	0.30
$-TS_\mathrm{r}^\mathrm{k}$	0.81	0.37	0.12	0.75	0.38	0.12
V_F	1.68	0.89	0.30	1.35	0.55	0.27

Table 3: [a]Maximum variation, [b]average and [c]standard deviation in the 12x12 grid defined in the Ramachandran space of the protected dipeptide HCO-L-Ala-NH_2 for the three energy surfaces, V_Σ, F_s and F_r, the three correcting terms, $-TS_\mathrm{s}^\mathrm{k}$, $-TS_\mathrm{s}^\mathrm{c}$, and $-TS_\mathrm{r}^\mathrm{k}$ and the Fixman's compensating potential V_F. The results at both MP2/6-31++G(d,p) and HF/6-31++G(d,p) levels of the theory are presented and all the functions have been referenced to zero in the grid. The units used are kcal/mol.

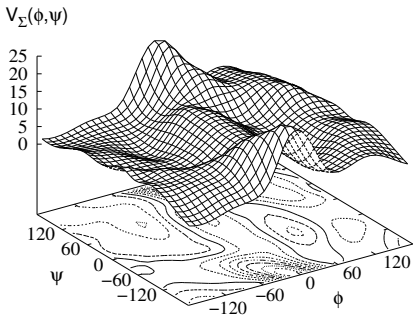

Figure 3: Potential Energy Surface (PES) of the model dipeptide HCO-L-Ala-NH_2, computed at the MP2/6-31++G(d,p) level of the theory. The surface has been referenced to zero and smoothed with bicubic splines for visual convenience. The units in the z-axis are kcal/mol.

between the energy surfaces V_Σ, F_s and F_r calculated at MP2/6-31++G(d,p) and the ones calculated at HF/6-31++G(d,p).

For the same reasons, we may conclude that, if ab initio derived potentials are used to carry out Molecular Dynamics simulations of peptides, the Fixman's compensating potential V_F should be included. Finally, regarding the relative importance of the different correcting terms $-TS_\mathrm{s}^\mathrm{k}$, $-TS_\mathrm{s}^\mathrm{c}$ and $-TS_\mathrm{r}^\mathrm{k}$, the results

in table 3 suggest that the less important one is the kinetic entropy $-TS_{\rm s}^{\rm k}$ of the stiff case (related to the determinant of the mass-metric tensor G) and that the most important one is the one related to the determinant of the Hessian matrix $\mathcal{H}$ of the constraining part of the potential, i.e., the conformational entropy $-TS_{\rm s}^{\rm k}$. The first conclusion is in agreement with the approximations typically made in the literature, the second one, however, is not (see sec. 2.5).

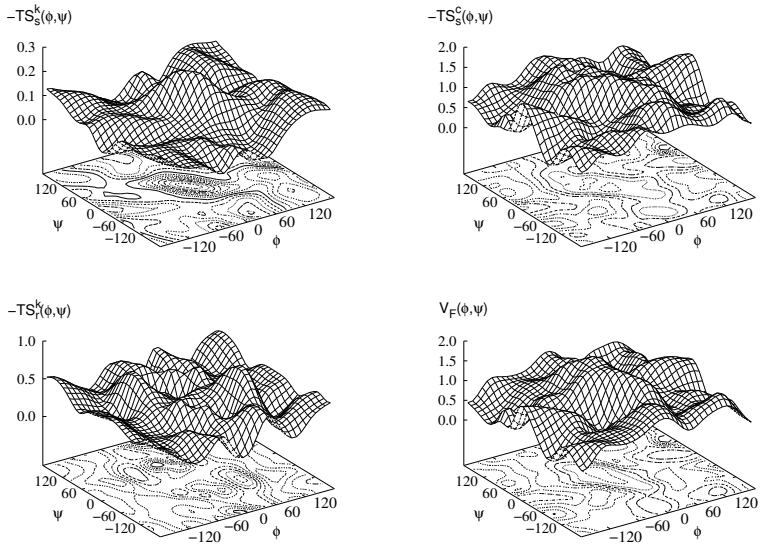

Figure 4: Ramachandran plots of the correcting terms appearing in eqs. (2.12) and (2.23), together with the Fixman's compensating potential defined in eq. (2.25), computed at the MP2/6-31++G(d,p) level of the theory in the model dipeptide HCO-L-Ala-NH_2. The surfaces have been referenced to zero and smoothed with bicubic splines for visual convenience. The units in the z-axes are kcal/mol.

Now, although the relative sizes of the conformational dependence of the different terms may be indicative of their importance, the degree of correlation among the surfaces is also relevant (see table 5). Hence, in order to arrive to more precise conclusions, we reexamine here the results using a physically meaningful criterium to compare potential energy functions that has been introduced in ref. 86. The *distance*, denoted by d_{12}, between any two different potential energy functions, V_1 and V_2, is an statistical quantity that, from a working set of conformations (in this case, the 144 points of the grid), measures the typical error that one makes in the *energy differences* if V_2 is used instead of V_1, admitting a linear rescaling.

In table 4, which contains the central results of this work, the distances between some of the energy surfaces that play a role in the problem are shown. We present the result in units of RT (at 300^o K, where $RT \simeq 0.6$ kcal/mol)

Corr.[a]	V_1[b]	V_2[c]	d_{12}[d]	$N_{\rm res}$[e]	b_{12}[f]	r_{12}[g]
		MP2/6-31++G(d,p)				
$-TS_{\rm s}^{\rm k}-TS_{\rm s}^{\rm c}$	$F_{\rm s}$	V_Σ	0.74 RT	1.82	0.98	0.9967
$-TS_{\rm s}^{\rm c}$	$F_{\rm s}$	$V_\Sigma-TS_{\rm s}^{\rm k}$	0.74 RT	1.83	0.98	0.9967
$-TS_{\rm s}^{\rm k}$	$F_{\rm s}$	$V_\Sigma-TS_{\rm s}^{\rm c}$	0.11 RT	80.45	1.00	0.9999
$-TS_{\rm r}^{\rm k}$	$F_{\rm r}$	V_Σ	0.29 RT	11.62	1.01	0.9995
$V_{\rm F}$	$F_{\rm s}$	$F_{\rm r}$	0.67 RT	2.24	0.97	0.9972
		HF/6-31++G(d,p)				
$-TS_{\rm s}^{\rm k}-TS_{\rm s}^{\rm c}$	$F_{\rm s}$	V_Σ	0.73 RT	1.90	0.99	0.9975
$-TS_{\rm s}^{\rm c}$	$F_{\rm s}$	$V_\Sigma-TS_{\rm s}^{\rm k}$	0.71 RT	2.00	0.99	0.9976
$-TS_{\rm s}^{\rm k}$	$F_{\rm s}$	$V_\Sigma-TS_{\rm s}^{\rm c}$	0.10 RT	90.99	1.00	0.9999
$-TS_{\rm r}^{\rm k}$	$F_{\rm r}$	V_Σ	0.26 RT	14.83	1.01	0.9997
$V_{\rm F}$	$F_{\rm s}$	$F_{\rm r}$	0.61 RT	2.69	0.98	0.9982
		MP2/6-31++G(d,p) vs. HF/6-31++G(d,p)				
	V_Σ	V_Σ	1.25 RT	0.64	1.12	0.9925
	$F_{\rm s}$	$F_{\rm s}$	1.18 RT	0.72	1.11	0.9934
	$F_{\rm r}$	$F_{\rm r}$	1.18 RT	0.72	1.12	0.9932

Table 4: Comparison of different energy surfaces involved in the study of the constrained equilibrium of the protected dipeptide HCO-L-Ala-NH_2. [a]Correcting term whose importance is measured in the corresponding row, [b]reference potential energy V_1 (the "correct" one, the one containing the correcting term), [c]approximated potential energy V_2 (i.e, V_1 minus the correcting term in column a), [d]statistical distance between V_1 and V_2 (see Alonso J. L. and Echenique P., *J. Comp. Chem.* **27** (2006) 238–252), [e]maximum number of residues in a polypeptide potential up to which the correcting term in column a may be omitted, [f]slope of the linear rescaling between V_1 and V_2 and [g]Pearson's correlation coefficient. All quantities are dimensionless, except for d_{12} which is given in units of the thermal energy RT at 300^o K.

because it has been argued in ref. 86 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 physical properties. Moreover, if one assumes that the effective energies compared will be used to construct a polypeptide potential and that it will be designed as simply the sum of mono-residue ones (making each term suitably depend on different pairs of Ramachandran angles), then, the number $N_{\rm res}$ 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 ref. 86):

$$N_{\mathrm{res}} = \left(\frac{RT}{d_{12}}\right)^2 . \tag{4.1}$$

This number is also shown in table 4, together with the slope b_{12} of the linear rescaling between V_1 and V_2 and the Pearson's correlation coefficient [87], denoted by r_{12}.

$V_1{}^a$		$V_2{}^b$	$r_{12}{}^c$
MP2/6-31++G(d,p)			
V_Σ	vs.	$-TS_{\mathrm{s}}^{\mathrm{c}}$	0.1572
V_Σ	vs.	$-TS_{\mathrm{s}}^{\mathrm{k}}$	-0.0008
V_Σ	vs.	$-TS_{\mathrm{r}}^{\mathrm{k}}$	-0.3831
V_Σ	vs.	V_{F}	0.3334
HF/6-31++G(d,p)			
V_Σ	vs.	$-TS_{\mathrm{s}}^{\mathrm{c}}$	0.0682
V_Σ	vs.	$-TS_{\mathrm{s}}^{\mathrm{k}}$	0.0897
V_Σ	vs.	$-TS_{\mathrm{r}}^{\mathrm{k}}$	-0.3544
V_Σ	vs.	V_{F}	0.2404
MP2/6-31++G(d,p) vs. HF/6-31++G(d,p)			
$-TS_{\mathrm{s}}^{\mathrm{c}}$	vs.	$-TS_{\mathrm{s}}^{\mathrm{c}}$	0.9136
$-TS_{\mathrm{s}}^{\mathrm{k}}$	vs.	$-TS_{\mathrm{s}}^{\mathrm{k}}$	0.9808
$-TS_{\mathrm{r}}^{\mathrm{k}}$	vs.	$-TS_{\mathrm{r}}^{\mathrm{k}}$	0.9316
V_{F}	vs.	V_{F}	0.9217

Table 5: Correlation between the different correcting terms involved in the study of the constrained equilibrium of the protected dipeptide HCO-L-Ala-NH_2. [a]Reference potential energy, [b]approximated potential energy, [c]Pearson's correlation coefficient.

The results at both MP2/6-31++G(d,p) and HF/6-31++G(d,p) levels of the theory are presented. The first three rows in each of the first two blocks are related to the classical stiff model, the next row to the classical rigid model and the last one in each block to the comparison between the two models. The third block in the table is associated to the comparison between the two different levels of the theory used.

The F_{s} vs. V_Σ row (in the first two blocks) assess the importance of the two correcting terms, $-TS_{\mathrm{s}}^{\mathrm{k}}$ and $-TS_{\mathrm{s}}^{\mathrm{c}}$, in the stiff case. The result $d_{12} = 0.74RT$ indicates that, for the alanine dipeptide, V_Σ may be used as an approximation of F_{s} with caution if accurate results are sought. In fact, the low value of $N_{\mathrm{res}} = 1.82 < 2$ shows that, *if we wanted to describe a 2-residue peptide omitting the stiff correcting terms, we would typically make an error greater than the thermal noise in the energy differences.* The next two rows investigate the

effect of each one of the individual correcting terms. The conclusion that can be extracted from them (as the relative sizes in table 3 already suggested) is that the conformational entropy associated to the determinant of the Hessian matrix $\mathcal{H}$ is much more relevant than the correcting term $-TS_{\mathrm{s}}^{\mathrm{k}}$, related to the mass-metric tensor G, *allowing to drop the latter up to* ~ 80 *residues* (according to MP2/6-31++G(d,p) calculations). As has been already remarked, this second conclusion is in agreement with the approximations frequently done in the literature; however, it turns out that the importance of the Hessian-related term has been persistently underestimated (see sec. 2.5 for a discussion).

The F_{r} vs. V_Σ row, in turn, shows the data associated to the kinetic entropy term $-TS_{\mathrm{r}}^{\mathrm{k}}$, which is related to the determinant of the reduced mass-metric tensor g in the classical rigid model. From the results there ($d_{12} = 0.29RT$ and $N_{\mathrm{res}} = 11.62$ at the MP2/6-31++G(d,p) level), we can conclude that the only correction term in the rigid case is less important than the ones in the stiff case and that V_Σ *may be used as an approximation of* F_{r} *for oligopeptides of up to* ~ 12 *residues.*

The last row in each of the first two blocks in table 4 is related to the interesting question in Molecular Dynamics of whether or not one should include the Fixman's compensating potential V_F (see eq. (2.25)) in rigid simulations in order to obtain the stiff equilibrium distribution, $\exp(-\beta F_{\mathrm{s}})$, instead of the rigid one, $\exp(-\beta F_{\mathrm{r}})$. This question is equivalent to asking whether or not F_{r} is a good approximation of F_{s}. From the results in the table, we can conclude that *the Fixman's potential is relevant for peptides of more than 2 residues and its omission may cause an error greater than the thermal noise in the energy differences.*

The appreciable sizes of the different correcting terms, shown in table 3, together with their low correlation with the Potential Energy Surface V_Σ, presented in the first two blocks of table 5, explain their considerable relevance discussed in the preceding paragraphs.

Moreover, from the comparison of the MP2/6-31++G(d,p) and the HF/6-31++G(d,p) blocks, one can tell that *the study herein performed may well have been done at the lower level of the theory* (if we had known) with a tenth of the computational effort (see sec. 3). This fact, explained by the high correlation, presented in the third block of table 5, between the correcting terms calculated at the two levels, is *very relevant for further studies* on more complicated dipeptides or longer chains and it indicates that the differences in size between the different correcting terms at MP2/6-31++G(d,p) and HF/6-31++G(d,p), which are presented in table 3, are mostly due to a harmless linear scaling effect similar to the well-known empirical scale factor frequently used in ab initio vibrational analysis [88–90]. This view is supported by the data in the third block of table 4, related to the comparison between the energy surfaces calculated at MP2/6-31++G(d,p) and HF/6-31++G(d,p), where the slopes b_{12} are consistently larger than unity.

A last conclusion that may be extracted from the block labeled "MP2/6-31++G(d,p) vs. HF/6-31++G(d,p)" in table 4 is that the typical error in the energy differences (given by the distances d_{12}) produced when one reduces the

level of the theory from MP2/6-31++G(d,p) to HF/6-31++G(d,p) *is comparable* (less than twice) to the error made if the most important correcting terms of the classical constrained models studied in this work are dropped. This is a useful hint for researchers interested in the conformational analysis of peptides with quantum chemistry methods [68–72, 75, 91] and also to those whose aim is the design and parametrization of classical force fields from ab initio quantum mechanical calculations [73–75].

	ϕ	ψ
α-helix	-57	-47
3_{10}-helix	-49	-26
π-helix	-57	-70
polyproline II	-79	149
parallel β-sheet	-119	113
antiparallel β-sheet	-139	135

Table 6: Ramachandran angles (in degrees) of some important secondary structure elements in polypeptides. Data taken from Lesk A. M., *Introduction to Protein Architecture*, Oxford University Press, Oxford, 2001.

Finally, in order to enrich and qualify the analysis, a new *working set* of conformations, different from the 144 points of the grid in the Ramachandran space, have been selected and the whole study has been repeated on them. These new conformations are six important secondary structure elements which form repetitive patterns stabilized by hydrogen bonds in polypeptides. Their conventional names and the corresponding values of the ϕ and ψ angles have been taken from ref. 92 and are shown in table 6.

In fig. 5, the relative energies of these conformations are shown for the three relevant potentials, V_Σ, F_s and F_r, at both MP2/6-31++G(d,p) and HF/6-31++G(d,p) levels of the theory. Since the antiparallel β-sheet is the structure with the minimum energy in all the cases, it has been set as the reference and the rest of energies in the figure should be regarded as relative to it.

The meaningful assessment, using the statistical distance described above, of the typical error made in the energy differences has been also performed on this new working set of conformations. The results are presented in table 7.

The distances between the free energies, F_s and F_r, and their corresponding approximations obtained dropping the correcting entropies, $-TS_\mathrm{s}^\mathrm{k}$, $-TS_\mathrm{s}^\mathrm{c}$ and $-TS_\mathrm{r}^\mathrm{k}$, or the Fixman's compensating potential V_F, in the first two blocks of the table, are *consistently smaller than the ones found in the study of the grid defined in the whole Ramachandran space* (cf. table 4). And so are the distances between the three relevant potentials, V_Σ, F_s and F_r, calculated at the MP2/6-31++G(d,p) and HF/6-31++G(d,p) levels of the theory.

Although the distance d_{12} used is a statistical quantity and, therefore, one must be cautious when working with such a small set of conformations (of size

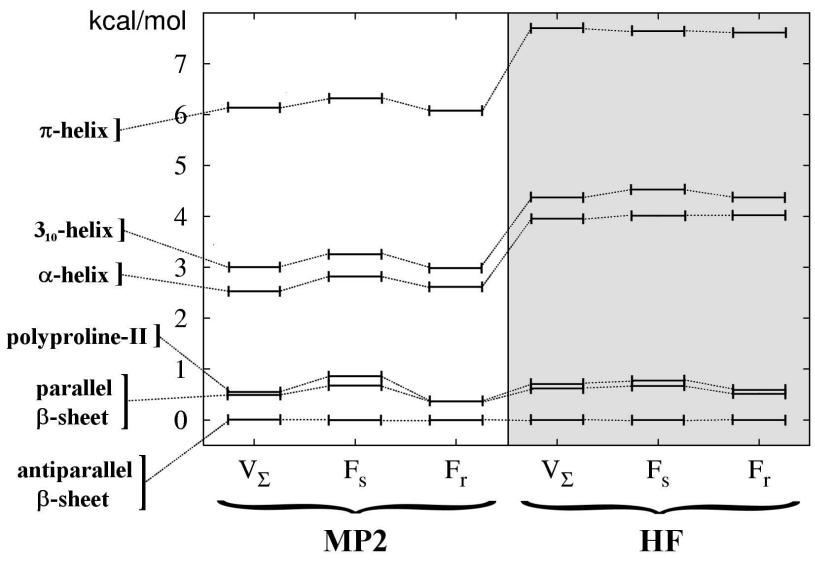

Figure 5: Relative energies of some important elements of secondary structure for the three potentials V_Σ, F_s and F_r, in the model dipeptide HCO-L-Ala-NH_2 and at both MP2/6-31++G(d,p) and HF/6-31++G(d,p) levels of the theory. The energy of the antiparallel β-sheet has been taken as reference. The units are kcal/mol.

six, in this case), the conclusion drawn from this second part of the study is that, if one is interested only in the "lower region" of the Ramachandran surface, where the typical secondary structure elements lie, then, *one may safely neglect the conformational dependence of the different correcting terms appearing in the study of the constrained equilibrium of peptides*. At least, up to oligopeptides (poly-alanines) of ~ 10 residues in the worst case (the neglect of the Fixman's compensating potential V_F in the F_s vs. F_r comparison at MP2/6-31++G(d,p)).

This difference between the two working set of conformations may be explained looking at one of the ways of expressing the statistical distance used (see eq. (12a) in ref. 86):

$$d_{12} = \sqrt{2}\,\sigma_2(1 - r_{12}^2)^{1/2}\ , \tag{4.2}$$

where r_{12} is the Pearson's correlation coefficient between the potential energies denoted by V_1 and V_2 and σ_2 is the standard deviation in the values of V_2 on the relevant working set of conformations.

This last quantity, σ_2, is the responsible of the differences between tables 4 and 7, since the set of conformations comprised by the six secondary structure elements in table 6 spans a smaller energy range than the whole Potential Energy Surface in fig. 3 (or F_s, or F_r, which have very similar variations). Accordingly, the dispersion in the energy values is smaller: $\sigma_2 \simeq 2$ kcal/mol in the case of

Corr.[a]	V_1[b]	V_2[c]	d_{12}[d]	$N_{\rm res}$[e]	b_{12}[f]	r_{12}[g]
		MP2/6-31++G(d,p)				
$-TS_{\rm s}^{\rm k}-TS_{\rm s}^{\rm c}$	$F_{\rm s}$	V_Σ	0.22 RT	19.72	0.99	0.9990
$-TS_{\rm s}^{\rm c}$	$F_{\rm s}$	$V_\Sigma-TS_{\rm s}^{\rm k}$	0.26 RT	14.07	0.98	0.9985
$-TS_{\rm s}^{\rm k}$	$F_{\rm s}$	$V_\Sigma-TS_{\rm s}^{\rm c}$	0.06 RT	298.13	1.01	0.9999
$-TS_{\rm r}^{\rm k}$	$F_{\rm r}$	V_Σ	0.20 RT	25.64	0.99	0.9992
$V_{\rm F}$	$F_{\rm s}$	$F_{\rm r}$	0.34 RT	8.73	0.99	0.9977
		HF/6-31++G(d,p)				
$-TS_{\rm s}^{\rm k}-TS_{\rm s}^{\rm c}$	$F_{\rm s}$	V_Σ	0.14 RT	47.94	1.00	0.9997
$-TS_{\rm s}^{\rm c}$	$F_{\rm s}$	$V_\Sigma-TS_{\rm s}^{\rm k}$	0.15 RT	46.12	1.00	0.9997
$-TS_{\rm s}^{\rm k}$	$F_{\rm s}$	$V_\Sigma-TS_{\rm s}^{\rm c}$	0.05 RT	380.30	1.00	0.9999
$-TS_{\rm r}^{\rm k}$	$F_{\rm r}$	V_Σ	0.15 RT	41.85	0.99	0.9997
$V_{\rm F}$	$F_{\rm s}$	$F_{\rm r}$	0.18 RT	30.12	1.01	0.9996
		MP2/6-31++G(d,p) vs. HF/6-31++G(d,p)				
	V_Σ	V_Σ	0.77 RT	1.68	1.28	0.9929
	$F_{\rm s}$	$F_{\rm s}$	0.77 RT	1.69	1.26	0.9928
	$F_{\rm r}$	$F_{\rm r}$	0.71 RT	1.96	1.28	0.9939

Table 7: Comparison of different approximations to the energies of some important elements of secondary structure (see table 6) in the study of the constrained equilibrium of the protected dipeptide HCO-L-Ala-NH_2. See the caption of table 4 for an explanation of the keys in the different columns.

the secondary structure elements and $\sigma_2 \simeq 4$ kcal/mol for the grid in the whole Ramachandran space (see table 3). Since the correlation coefficient in both cases are of similar magnitude, the differences in σ_2 produce a smaller distance d_{12} for the second set of conformations studied, i.e., a smaller typical error made in the energy differences when omitting the correcting terms derived from the consideration of constraints.

To end this section, we remark that, although this "lower region" of the Ramachandran space contains the most relevant secondary structure elements (which are also the most commonly found in experimentally resolved native structures of proteins [93–96]) and may be the only region explored in the dynamical or thermodynamical study of small peptides, if the aim is the design of effective potentials for computer simulation of polypeptides [73–75], then, some caution is recommended, since long-range interactions in the sequence may temporarily compensate local energy penalizations and the higher regions of the energy surfaces studied could be important in transition states or in some

relevant dynamical paths of the system.

In the following section, the many results discussed in the preceding paragraphs are summarized.

5 Conclusions

In this work, the theory of classical constrained equilibrium has been collected for the stiff and rigid models. The pertinent correcting terms, which may be regarded as effective entropies, as well as the Fixman's compensating potential, have been derived and theoretically discussed (see eqs. (2.12), (2.23) and (2.25), together with the formulae in sec. 2.4). Their inclusion in the literature has been thoroughly reviewed in sec. 2.5. In addition, the common approximation of considering that, for typical internals, the equilibrium values of the hard coordinates do not depend on the soft ones, has also been discussed and related to the rest of simplifications.

In the central part of the work (sec. 4), the relevance of the different correcting terms has been assessed in the case of the model dipeptide HCO-L-Ala-NH_2, with quantum mechanical calculations including electron correlation. Also, the possibility of performing analogous studies at the less demanding Hartree-Fock level of the the theory has been investigated. The results found are summarized in the following points:

- *In Monte Carlo simulations of the classical stiff model* at room temperature, the effective entropy $-TS_{\mathrm{s}}^{\mathrm{k}}$, associated to the determinant of the mass-metric tensor G, may be neglected for peptides of up to ~ 80 residues. Its maximum variation in the Ramachandran space is 0.24 kcal/mol.

- *In Monte Carlo simulations of the classical stiff model* at room temperature, the effective entropy $-TS_{\mathrm{s}}^{\mathrm{c}}$, associated to the determinant of the Hessian $\mathcal{H}$ of the constraining part of the potential, should be included for peptides of more than 2 residues. Its maximum variation in the Ramachandran space is 1.67 kcal/mol.

- *In Monte Carlo simulations of the classical rigid model* at room temperature, the effective entropy $-TS_{\mathrm{r}}^{\mathrm{k}}$, associated to the determinant of the reduced mass-metric tensor g, may be neglected for peptides of up to ~ 12 residues. Its maximum variation in the Ramachandran space is 0.81 kcal/mol.

- *In rigid Molecular Dynamics simulations intended to yield the stiff equilibrium distribution* at room temperature, the Fixman's compensating potential V_{F} should be included for peptides of more than 2 residues. Its maximum variation in the Ramachandran space is 1.68 kcal/mol.

- If the assumption that only the more stable region of the Ramachandran space, where the principal elements of secondary structure lie, is relevant,

then, *the importance of the correcting terms decreases and the limiting number of residues in a polypeptide potential up to which they may be omitted is approximately four times larger in each of the previous points.*

- In both cases (i.e., either if the whole Ramachandran space is considered relevant, or only the lower region), *the errors made if the most important correcting terms are neglected are of the same order of magnitude as the errors due to a decrease in the level of theory from* MP2/6-31++G(d,p) *to* HF/6-31++G(d,p).

- *The whole study of the relevance of the different correcting terms (or future analogous investigations) may be performed at the* HF/6-31++G(d,p) *level of the theory*, yielding very similar results to the ones obtained at MP2/6-31++G(d,p) and using a tenth of the computational effort.

To end this discussion, some qualifications should be made. On one hand, the conclusions above refer to the case in which a classical potential *directly extracted* from the quantum mechanical (Born-Oppenheimer) one is used; for the considerably simpler force fields typically used for macromolecular simulations, the study should be repeated and different results may be obtained. On the other hand, the investigation performed in this work has been done in one of the simplest dipeptides; both its isolated character and the relatively small size of its side chain play a role in the results obtained. Hence, for bulkier residues included in polypeptides, these conclusions should be approached with caution and much interesting work remains to be done.

Acknowledgments

We would like to thank F. Falceto and V. Laliena for illuminating discussions. The numerical calculations have been performed at the BIFI computing facilities. We thank I. Campos, for the invaluable CPU time and the efficiency at solving the problems encountered.

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] J. L. Alonso, G. A. Chass, I. G. Csizmadia, P. Echenique, and A. Tarancón, Do theoretical physicists care about the protein folding problem?, in *Meeting on Fundamental Physics 'Alberto Galindo'*, edited by R. F. Álvarez-Estrada et al., Aula Documental, Madrid, 2004, (arXiv:q-bio.BM/0407024).

[2] C. M. Dobson, Protein folding and misfolding, *Nature* **426**, 884 (2003).

[3] K. A. Dill, Polymer principles and protein folding, *Prot. Sci.* **8**, 1166 (1999).

[4] S. He and H. A. Scheraga, Brownian dynamics simulations of protein folding, *J. Chem. Phys.* **108**, 287 (1998).

[5] R. A. Abagyan, M. M. Totrov, and D. A. Kuznetsov, ICM: A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation, *J. Comp. Chem.* **15**, 488 (1994).

[6] W. F. Van Gunsteren and M. Karplus, Effects of constraints on the dynamics of macromolecules, *Macromolecules* **15**, 1528 (1982).

[7] C. Levinthal, How to fold graciously, in *Mossbauer Spectroscopy in Biological Systems*, edited by J. T. P. DeBrunner and E. Munck, pp. 22–24, Allerton House, Monticello, Illinois, 1969, University of Illinois Press.

[8] H. M. Chun, C. E. Padilla, D. N. Chin, M. Watanabe, V. I. Karlov, H. E. Alper, K. Soosaar, K. B. Blair, O. M. Becker, L. S. D. Caves, R. Nagle, D. N. Haney, and B. L. Farmer, MBO(N)D: A multibody method for long-time Molecular Dynamics simulations, *J. Comp. Chem.* **21**, 159 (2000).

[9] S. Reich, Smoothed Langevin dynamics of highly oscillatory systems, *Physica D* **118**, 210 (2000).

[10] S. Reich, Multiple time scales in classical and quantum-classical molecular dynamics, *J. Comput. Phys.* **151**, 49 (1999).

[11] T. Schlick, E. Barth, and M. Mandziuk, Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation, *Annu. Rev. Biophys. Biomol. Struct.* **26**, 181 (1997).

[12] N. G. Van Kampen and J. J. Lodder, Constraints, *Am. J. Phys.* **52**, 419 (1984).

[13] J. M. Rallison, The role of rigidity constraints in the rheology of dilute polymer solutions, *J. Fluid Mech.* **93**, 251 (1979).

[14] E. Helfand, Flexible vs. rigid constraints in Statistical Mechanics, *J. Chem. Phys.* **71**, 5000 (1979).

[15] N. Gō and H. A. Scheraga, On the use of classical statistical mechanics in the treatment of polymer chain conformation, *Macromolecules* **9**, 535 (1976).

[16] M. Fixman, Classical Statistical Mechanics of constraints: A theorem and application to polymers, *Proc. Natl. Acad. Sci. USA* **71**, 3050 (1974).

[17] N. Gō and H. A. Scheraga, Analysis of the contributions of internal vibrations to the statistical weights of equilibrium conformations of macromolecules, *J. Chem. Phys.* **51**, 4751 (1969).

[18] D. C. Morse, Theory of constrained Brownian motion, *Adv. Chem. Phys.* **128**, 65 (2004).

[19] W. K. Den Otter and W. J. Briels, Free energy from molecular dynamics with multiple constraints, *Mol. Phys.* **98**, 773 (2000).

[20] A. Patriciu, G. S. Chirikjian, and R. V. Pappu, Analysis of the conformational dependence of mass-metric tensor determinants in serial polymers with constraints, *J. Chem. Phys.* **121**, 12708 (2004).

[21] D. Perchak, J. Skolnick, and R. Yaris, Dynamics of rigid and flexible constraints for polymers. Effect of the Fixman potential, *Macromolecules* **18**, 519 (1985).

[22] M. R. Pear and J. H. Weiner, Brownian dynamics study of a polymer chain of linked rigid bodies, *J. Chem. Phys.* **71**, 212 (1979).

[23] D. Chandler and B. J. Berne, Comment on the role of constraints on the conformational structure of n-butane in liquid solvent, *J. Chem. Phys.* **71**, 5386 (1979).

[24] M. Gottlieb and R. B. Bird, A Molecular Dynamics calculation to confirm the incorrectness of the random-walk distribution for describing the Kramers freely jointed bead-rod chain, *J. Chem. Phys.* **65**, 2467 (1976).

[25] J. Zhou, S. Reich, and B. R. Brooks, Elastic molecular dynamics with self-consistent flexible constraints, *J. Chem. Phys.* **111**, 7919 (2000).

[26] H. J. C. Berendsen and W. F. Van Gunsteren, Molecular Dynamics with constraints, in *The Physics of Superionic Conductors and Electrode Materials*, edited by J. W. Perram, volume NATO ASI Series B92, pp. 221–240, Plenum Press, 1983.

[27] A. D. MacKerell Jr., B. Brooks, C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus, CHARMM: The energy function and its parameterization with an overview of the program, in *The Encyclopedia of Computational Chemistry*, edited by P. v. R. Schleyer et al., pp. 217–277, John Wiley & Sons, Chichester, 1998; B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comp. Chem.* **4**, 187 (1983).

[28] 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, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem.*

Soc. **117**, 5179 (1995); D. A. Pearlman, D. A. Case, J. W. Caldwell, W. R. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules, *Comp. Phys. Commun.* **91**, 1 (1995).

[29] W. L. Jorgensen and J. Tirado-Rives, The OPLS potential functions for proteins. Energy minimization for crystals of cyclic peptides and Crambin, *J. Am. Chem. Soc.* **110**, 1657 (1988); W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.* **118**, 11225 (1996).

[30] T. A. Halgren, Merck molecular force field. I. Basis, form, scope, parametrization, and performance of MMFF94, *J. Comp. Chem.* **17**, 490 (1996); T. A. Halgren, Merck molecular force field. II. MMFF94 van der Waals and electrostatica parameters for intermolecular interactions, *J. Comp. Chem.* **17**, 520 (1996); T. A. Halgren, Merck molecular force field. III. Molecular geometrics and vibrational frequencies for MMFF94, *J. Comp. Chem.* **17**, 553 (1996); T. A. Halgren, Merck molecular force field. IV. Conformational energies and geometries for MMFF94, *J. Comp. Chem.* **17**, 587 (1996); T. A. Halgren, Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules, *J. Comp. Chem.* **17**, 616 (1996).

[31] M. Pasquali and D. C. Morse, An efficient algorithm for metric correction forces in simulations of linear polymers with constrained bond lengths, *J. Chem. Phys.* **116**, 1834 (2002).

[32] W. K. Den Otter and W. J. Briels, The calculation of free-energy differences by constrained molecular-dynamics simulations, *J. Chem. Phys.* **109**, 4139 (1998).

[33] G. Ciccotti and J. P. Ryckaert, Molecular dynamics simulation of rigid molecules, *Comput. Phys. Rep.* **4**, 345 (1986).

[34] H. J. C. Berendsen and W. F. Van Gunsteren, Molecular Dynamics simulations: Techniques and approaches, in *Molecular Liquids-Dynamics and Interactions*, edited by A. J. e. a. Barnes, pp. 475–500, Reidel Publishing Company, 1984.

[35] M. Fixman, Simulation of polymer dynamics. I. General theory, *J. Chem. Phys.* **69**, 1527 (1978).

[36] R. F. Álvarez-Estrada and G. F. Calvo, Models for biopolymers based on quantum mechanics, *Mol. Phys.* **100**, 2957 (2002).

[37] R. F. Álvarez-Estrada, Models of macromolecular chains based on Classical and Quantum Mechanics: comparison with Gaussian models, *Macromol. Theory Simul.* **9**, 83 (2000).

[38] B. Hess, H. Saint-Martin, and H. J. C. Berendsen, Flexible constraints: An adiabatic treatment of quantum degrees of freedom, with application to the flexible and polarizable mobile charge densities in harmonic oscillators model for water, *J. Chem. Phys.* **116**, 9602 (2002).

[39] M. Born and J. R. Oppenheimer, Zur Quantentheorie der Molekeln, *Ann. Phys. Leipzig* **84**, 457 (1927).

[40] E. B. Wilson Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, Dover Publications, New York, 1980.

[41] K. Németh and M. Challacombe, The quasi-independent curvilinear coordinate approximation for geometry optimization, *J. Chem. Phys.* **121**, 2877 (2004).

[42] B. Paizs, J. Baker, S. Suhai, and P. Pulay, Geometry optimization of large biomolecules in redundant internal coordinates, *J. Chem. Phys.* **113**, 6566 (2000).

[43] M. von Arnim and R. Ahlrichs, Geometry optimization in generalized natural internal coordinates, *J. Chem. Phys.* **111**, 9183 (1999).

[44] J. Baker, A. Kessi, and B. Delley, The generation and use of delocalized internal coordinates in geometry optimization, *J. Chem. Phys.* **105**, 192 (1996).

[45] G. Fogarasi, X. Zhou, P. W. Taylor, and P. Pulay, The calculation of ab initio molecular geometries: Efficient natural internal coordinates and empirical correction by offset forces, *J. Am. Chem. Soc.* **114**, 8191 (1992); P. Pulay and G. Fogarasi, Geometry optimization in redundant internal coordinates, *J. Chem. Phys.* **96**, 2856 (1992); P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs, Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives, *J. Am. Chem. Soc.* **101**, 2550 (1979).

[46] A. R. Dinner, Local deformations of polymers with nonplanar rigid main-chain coordinates, *J. Comp. Chem.* **21**, 1132 (2000).

[47] J. Schofield and M. A. Ratner, Monte Carlo methods for short polypeptides, *J. Chem. Phys.* **109**, 9177 (1998).

[48] A. J. Pertsin, J. Hahn, and H. P. Grossmann, Incorporation of bond-lengths constraints in Monte Carlo simulations of cyclic and linear molecules: conformational sampling for cyclic alkanes as test systems, *J. Comp. Chem.* **15**, 1121 (1994).

[49] E. W. Knapp and A. Irgens-Defregger, Off-lattice Monte Carlo method with constraints: Long-time dynamics of a protein model without nonbonded interactions, *J. Fluid Mech.* **14**, 19 (1993).

[50] P. Echenique and J. L. Alonso, Definition of Systematic, Approximately Separable and Modular Internal Coordinates (SASMIC) for macromolecular simulation, *To be published in J. Comp. Chem.*, 2006, (arXiv:q-bio.BM/0511004).

[51] R. A. Abagyan and A. K. Mazur, New methodology for computer-aided modelling of biomolecular structure and dynamics. 2. Local deformations and cycles, *J. Biomol. Struct. Dyn.* **6**, 833 (1989).

[52] A. K. Mazur and R. A. Abagyan, New methodology for computer-aided modelling of biomolecular structure and dynamics. 1. Non-cyclic structures, *J. Biomol. Struct. Dyn.* **6**, 815 (1989).

[53] E. Barth, K. Kuczera, B. Leimkuhler, and R. D. Skeel, Algorithms for constrained Molecular Dynamics, *J. Comp. Chem.* **16**, 1192 (1995).

[54] H. C. Andersen, Rattle: A "velocity" version of the Shake algorithm for molecular dynamics calculations, *J. Comput. Phys.* **52**, 24 (1983).

[55] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes, *J. Comput. Phys.* **23**, 327 (1977).

[56] J. Schlitter and M. Klän, The free energy of a reaction coordinate at multiple constraints: a concise formulation, *Mol. Phys.* **101**, 3439 (2003).

[57] W. F. Van Gunsteren, Methods for calculation of free energies and binding constants: Successes and problems, in *Computer Simulations of Biomolecular Systems*, edited by W. F. Van Gunsteren and P. K. Weiner, pp. 27–59, Escom science publishers, Netherlands, 1989.

[58] I. Andricioaei and M. Karplus, On the calculation of entropy from covariance matrices of the atomic fluctuations, *J. Chem. Phys.* **115**, 6289 (2001).

[59] M. Karplus and J. N. Kushick, Method for estimating the configurational entropy of macromolecules, *Macromolecules* **14**, 325 (1981).

[60] N. G. Almarza, E. Enciso, J. Alonso, F. J. Bermejo, and M. Álvarez, Monte Carlo simulations of liquid n-butane, *Mol. Phys.* **70**, 485 (1990).

[61] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Clarendon Press, Oxford, 2005.

[62] D. Frenkel and S. B., *Understanding molecular simulations: From algorithms to applications*, Academic Press, Orlando FL, 2nd edition, 2002.

[63] D. A. Svetogorsky, A freely jointed polymer chain with bond vectors of fixed length, *J. Phys. A: Math. Gen.* **11**, 2349 (1978).

[64] M. Mazars, Canonical partition function of freely jointed chains, *J. Phys. A: Math. Gen.* **31**, 1949 (1998).

[65] M. Mazars, Statistical physics of the freely jointed chain, *Phys. Rev. E* **53**, 6297 (1996).

[66] J. Chen, W. Im, and C. L. Brooks III, Application of torsion angle molecular dynamics for efficient sampling of protein conformations, *J. Comp. Chem.* **26**, 1565 (2005).

[67] P. Echenique and I. Calvo, Explicit factorization of external coordinates in constrained Statistical Mechanics models, *Submitted to J. Chem. Phys.*, 2006, (arXiv:q-bio.QM/0512033).

[68] A. Láng, I. G. Csizmadia, and A. Perczel, Peptide models. XLV: Conformational properties of N-formyl-L-methioninamide ant its relevance to methionine in proteins, *PROTEINS: Struct. Funct. Bioinf.* **58**, 571 (2005).

[69] A. Perczel, O. Farkas, I. Jakli, I. A. Topol, and I. G. Csizmadia, Peptide models. XXXIII. Extrapolation of low-level Hartree-Fock data of peptide conformation to large basis set SCF, MP2, DFT and CCSD(T) results. The Ramachandran surface of alanine dipeptide computed at various levels of theory, *J. Comp. Chem.* **24**, 1026 (2003).

[70] R. Vargas, J. Garza, B. P. Hay, and D. A. Dixon, Conformational study of the alanine dipeptide at the MP2 and DFT levels, *J. Phys. Chem. A* **106**, 3213 (2002).

[71] C.-H. Yu, M. A. Norman, L. Schäfer, M. Ramek, A. Peeters, and C. van Alsenoy, Ab initio conformational analysis of N-formyl L-alanine amide including electron correlation, *J. Mol. Struct.* **567–568**, 361 (2001).

[72] A. G. Császár and A. Perczel, Ab initio characterization of building units in peptides and proteins, *Prog. Biophys. Mol. Biol.* **71**, 243 (1999).

[73] A. R. MacKerell Jr., M. Feig, and C. L. Brooks III, Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *J. Comp. Chem.* **25**, 1400 (2004).

[74] A. J. Bordner, C. N. Cavasotto, and R. A. Abagyan, Direct derivation of van der Waals force fields parameters from quantum mechanical interaction energies, *J. Phys. Chem. B* **107**, 9601 (2003).

[75] M. Beachy, D. Chasman, R. Murphy, T. Halgren, and R. Friesner, Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields, *J. Am. Chem. Soc.* **119**, 5908 (1997).

[76] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Graduate Texts in Mathematics, Springer, New York, 2nd edition, 1989.

[77] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern Geometry — Methods and Applications*, volume I. The Geometry of Surfaces, Transformation Groups and Fields, Springer, New York, 1984.

[78] M. V. Volkenstein, *Configurational Statistical of Polymeric Chains*, Interscience, New York, 1959.

[79] L. Schäfer and C. Ming, Predictions of protein backbone bond distances and angles from first principles, *J. Mol. Struct.* **333**, 201 (1995).

[80] G. N. Ramachandran and C. Ramakrishnan, Stereochemistry of polypeptide chain configurations, *J. Mol. Biol.* **7**, 95 (1963).

[81] V. Hnizdo, A. Fedorowicz, H. Singh, and E. Demchuk, Statistical thermodynamics of internal rotation in a hindering potential of mean force obtained from computer simulations, *J. Comp. Chem.* **24**, 1172 (2003).

[82] E. Demchuk and H. Singh, Statistical thermodynamics of hindered rotation from computer simulations, *Mol. Phys.* **99**, 627 (2001).

[83] K. V. Mardia and P. E. Jupp, *Directional Statistics*, John Wiley & Sons, Chichester, 2000.

[84] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, H. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, General Atomic and Molecular Electronic Structure System, *J. Comp. Chem.* **14**, 1347 (1993).

[85] I. P. Daykov, T. A. Arias, and T. D. Engeness, Robust ab initio calculation of condensed matter: Transparent convergence through semicardinal multiresolution analysis, *Phys. Rev. Lett.* **90**, 216402 (2003).

[86] J. L. Alonso and P. Echenique, A physically meaningful method for the comparison of potential energy functions, *J. Comp. Chem.* **27**, 238 (2006).

[87] J. D. Dobson, *Applied multivariate data analysis*, volume I, Springer-Verlag, New York, 1991.

[88] I. N. Levine, *Quantum Chemistry*, Prentice Hall, Upper Saddle River, 5th edition, 1999.

[89] M. D. Halls, J. Velkovski, and H. B. Schlegel, Harmonic Frequency Scaling Factors for Hartree-Fock, S-VWN, B-LYP, B3-LYP, B3-PW91 and MP2 and the Sadlej pVTZ Electric Property Basis Set, *Theo. Chem. Acc.* **105**, 413 (2001).

[90] A. P. Scott and L. Radom, Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and semiempirical scale factors, *J. Phys. Chem.* **100**, 16502 (1996).

[91] M. Elstner, K. J. Jalkanen, M. Knapp-Mohammady, and S. Suhai, Energetics and structure of glycine and alanine based model peptides: Approximate SCC-DFTB, AM1 and PM3 methods in comparison with DFT, HF and MP2 calculations, *Chem. Phys.* **263**, 203 (2001).

[92] A. M. Lesk, *Introduction to Protein Architecture*, Oxford University Press, Oxford, 2001.

[93] P. Chakrabarti and D. Pal, The interrelationships of side-chain and main-chain conformations in proteins, *Prog. Biophys. Mol. Biol.* **76**, 1 (2001).

[94] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, The protein data bank, *Nucleic Acids Research* **28**, 235 (2000).

[95] K. Gunasekaran, C. Ramakrishnan, and P. Balaram, Disallowed Ramachandran conformations of amino acid residues in protein structures, *J. Mol. Biol.* **264**, 191 (1996).

[96] T. E. Creighton, *Proteins: Structures and Molecular Properties*, Freeman, W. H., New York, 2nd edition, 1992.