

Ball-and-Stick Local Elevation Umbrella Sampling: Molecular Simulations Involving Enhanced Sampling within Conformational or Alchemical Subspaces of Low Internal Dimensionalities, Minimal Irrelevant Volumes, and Problem-Adapted Geometries

Halvor S. Hansen and Philippe H. Hünenberger*

Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland

Received June 5, 2010

Abstract: A new method, ball-and-stick local elevation umbrella sampling (B&S-LEUS), is proposed to enhance the sampling in computer simulations of (bio)molecular systems. It enables the calculation of conformational free-energy differences between states (or alchemical free-energy differences between molecules), even in situations where the definition of these states relies on a conformational subspace involving more than a few degrees of freedom. The B&S-LEUS method consists of the following steps: (A) choice of a reduced conformational subspace; (B) representation of the relevant states by means of spheres (“balls”), each associated with a biasing potential involving a one-dimensional radial memory-based term and a radial confinement term; (C) definition of a set of lines (“sticks”) connecting these spheres, each associated with a biasing potential involving a one-dimensional longitudinal memory-based term and a transverse confinement term; (D) unification of the biasing potentials corresponding to the union of all of the spheres and lines (active subspace) into a single biasing potential according to the enveloping distribution sampling (EDS) scheme; (E) build-up of the memory using the local elevation (LE) procedure, leading to a biasing potential enabling a nearly uniform sampling (radially within the spheres, longitudinally within the lines) of the active subspace; (F) generation of a biased ensemble of configurations using this preoptimized biasing potential, following an umbrella sampling (US) approach; and (G) calculation of the relative free energies of the states via reweighting and state assignment. The main characteristics of this approach are: (i) a low internal dimensionality, that is, the memory only involves one-dimensional grids (acceptable memory requirements); (ii) a minimal irrelevant volume, that is, the conformational volume opened to sampling includes a minimal fraction of irrelevant regions in terms of the free energy of the physical system or of user-specified metastable states (acceptable build-up duration requirements, high statistical efficiency); and (iii) a problem-adapted geometry (a priori specification of the conformational regions considered as relevant or irrelevant). In particular, the use of lines to connect the spheres ensures both a minimal irrelevant volume and a sufficient number of transitions between the states. As an illustration, the B&S-LEUS method is applied here to three test systems: (i) a solvated (blocked) alanine monopeptide (two-dimensional conformational subspace), used as a toy system to illustrate the versatility of the method in promoting the sampling of arbitrary regions of the Ramachandran map; (ii) a solvated polyalanine decapeptide (nine-dimensional conformational subspace), to evaluate the relative free energies of three different types of helices (π , α , and β_{10}) based on a single simulation; and (iii) a solvated artificial hexopyranose, termed the “mother” of all D-hexopyranoses and constructed as a hybrid of all D-hexopyranose stereoisomers, where the method is applied (seven-dimensional mixed alchemical and conformational subspace) to calculate the relative free energies of the corresponding 32 isomers, anomers, and chair conformers, based on a single simulation.

1. Introduction

Classical atomistic simulations, and in particular molecular dynamics (MD), represent nowadays a powerful tool complementary to experiment for investigating the properties of molecular systems relevant in physics, chemistry, and biology.^{1–4} Their success in the context of condensed-phase systems results in particular from a favorable trade-off between the spatial and temporal resolutions of these models (on the order 0.1 nm and 1 fs) and their computational costs, permitting to reach system sizes and time scales relevant for many (bio)molecular applications (on the order of 10 nm and 100 ns). These scales are sufficient to enable in many cases: (i) an appropriate description of bulk-like solvation using discrete solvent molecules; (ii) a reliable calculation of thermodynamic properties via statistical mechanics; and (iii) a direct comparison of simulated properties with experimental data.

In practice, however, the results of atomistic simulations are still affected by four main sources of error, originating from: (i) the classical atomistic approximation;^{5–8} (ii) the approximate force-field representation of interatomic interactions;^{2–4,9} (iii) the presence of finite-size and surface effects;^{10,11} and (iv) the insufficient conformational (and, possibly, alchemical; see below) sampling.^{9,12–15} The reduction of the last type of errors can be viewed as a first-priority target in the improvement of simulation methodologies, because these errors are predominantly nonsystematic, while the three other types of errors are systematic. The present project is concerned with the design of a new sampling-enhancement scheme with the goal of reducing the corresponding errors in the context of free-energy calculations.

Free-energy calculations based on classical atomistic simulations^{2,3,16–23} can be classified into two main categories. On the one hand, the calculation of conformational free energies aims at evaluating the relative free energies of relevant conformational states of a given molecular system (as well as corresponding free-energy profiles or maps). Typical examples include the evaluation of the relative free energies of the bound and unbound states of a molecular complex (e.g., protein–ligand complex^{24,25}), or of conformational states of a macromolecule presenting different spectroscopic or functional properties (e.g., folded and unfolded states of a protein,²⁶ α -helical or β -sheet conformation of a peptide,²⁷ different double-helical forms of an oligonucleotide²⁸). On the other hand, the calculation of alchemical free energies aims at evaluating the relative free energies of different molecules (or, more precisely, molecular topologies) in a given environment (the corresponding profiles or maps are then unphysical and irrelevant). In general, the target quantity is actually in this case a difference between the relative free energies calculated considering two different environments, so as to characterize the environmental effect via a thermodynamic cycle. Typical examples include the evaluation of solvation free energies^{29–31} (difference between the free-energy change upon “creating” a molecule in solution and in the gas phase) or of relative

binding free energies^{32,33} (difference between the free-energy change upon “mutating” a molecule into another one in solution and within a molecular complex). Note that the “creation” or “mutation” must be performed at constant number of atoms, possibly requiring the introduction of “dummy” atoms (covalently linked mass sites free of nonbonded interactions). Finally, for completeness, one might mention a third category of free-energy calculations, involving thermodynamic free-energy changes (i.e., free-energy changes upon variation of a thermodynamic parameter such as temperature or pressure).

In the early days of free-energy calculations, the evaluation of relative free energies was typically restricted to two-states or two-molecules problems, that is, to the evaluation of a single free-energy difference.

The established methods for the calculation of a single conformational free-energy difference are direct counting³⁴ (DC) and umbrella sampling^{35,36} (US). The DC approach is only applicable in the (uncommonly) favorable situation where the dynamics spontaneously samples the two conformational states with a sufficient number of interconversion transitions. In this case, the free-energy difference can be calculated directly from the ratio of the numbers of sampled conformations assigned to either of the two states. The US approach is more generally applicable and relies on the use of a time-independent biasing potential that forces the sampling of the two states with a sufficient number of interconversion transitions. In this case, the free-energy difference can be calculated from the ratio of the reweighted numbers of conformations assigned to either of the two states, the reweighting acting as a correction for the effect of the biasing. In practice, the direct design of a biasing potential satisfying the required properties for two significantly differing conformational states is difficult, and one has to resort to³⁷ multiple-windows,^{38–41} adaptive,^{42–52} or memory-based^{15,53–56} approaches.

The established methods for the calculation of a single alchemical free-energy difference are thermodynamic integration^{57,58} (TI) and free-energy perturbation^{59,60} (FP). In both cases, a hybrid Hamiltonian is constructed by introduction of a coupling parameter λ , in such a way that $\lambda = 0$ corresponds to the initial molecule and $\lambda = 1$ to the final molecule. In the TI approach, the free-energy difference is obtained from the ensemble average of the Hamiltonian λ -derivative evaluated at a discrete set of successive λ -points, via numerical integration. In the FP approach, the free-energy difference is obtained on the basis of an ensemble average performed at a single λ -point, involving the relative Boltzmann weights of the Hamiltonians associated with the two molecules. In practice, the latter scheme becomes inaccurate in the case of two significantly differing molecules, and a multiple-windows¹⁸ approach must be used instead.

The above methods for the calculation of single free-energy differences are still widely used nowadays and, after the resolution of some important methodological issues in the 1990s (Hamiltonian lag,⁶¹ singularity upon atom creation and deletion,^{62,63} metric-tensor effects,^{64–66} contribution of constraints,^{66–68} contribution of restraints,^{24,25,69} standard-state corrections^{24,25}), do not present major difficulties. Their

* Corresponding author phone: +41 44 632 5503; fax: +41 44 632 1039; e-mail: phil@igc.phys.chem.ethz.ch.

main drawback is that they are restricted to two-state or two-molecule (i.e., pairwise) problems (low throughput) and typically require a significant amount of human time (multiple simulations, case-to-case adjustment of the staging, equilibration and sampling protocols, analysis of the data). Recently, however, the field of free-energy calculation has witnessed three important evolutions, which are described in turn below.

As a first recent development, research has turned to the new challenge of calculating multiple free-energy differences from a single simulation (or, possibly, from a set of simulations generated in an “automated” way). This challenge is particularly relevant in the context of alchemical changes, for example, for the calculation of the relative binding free energies of a large collection of possible ligands to the same receptor simultaneously (a “holy grail” for drug design^{70,71}). Initial attempts, relying on the application of the FP formula in an extrapolative way (one-step perturbation) based on a real molecule as reference state, were not very successful.^{72–74} Subsequent attempts, relying on the design of an unphysical reference state instead (e.g., molecule with soft-sites⁶³ aiming at encompassing configurations representative of all possible final states within the reference ensemble), were in specific cases more successful,^{32,75–77} but remained generally speaking moderately reliable. Arguably, the first practically useful approach of this kind is enveloping distribution sampling^{33,78–80} (EDS), where an unphysical reference-state Hamiltonian is constructed automatically, which presents an optimal overlap with (and sufficient interconversion transitions between) the different target molecules. Note that the EDS method has only been applied until now to alchemical free-energy calculations and remains to be generalized to conformational problems. One general lesson from this research is that viable approaches for the calculation of multiple free-energy differences based on a single reference simulation require the targets (states or molecules) to be known *a priori*, that is, purely extrapolative approaches are not very reliable in practice.

As a second recent development, efficient MD-based methods have been developed to address the conformational-searching problem,^{14,15} that is, the problem of scanning a potential energy hypersurface for low-energy configurations over the widest possible volume. Probably the most efficient types of MD-based searching methods available nowadays are those that rely on the progressive build-up of a memory-based penalty potential, preventing the continuous revisiting of previously discovered configurations. Many closely related variants of this approach can be found in the literature, including (chronologically) the deflation,⁸¹ tunneling,⁸² tabu search,⁸³ local elevation,⁸⁴ conformational flooding,⁸⁵ Engkvist–Karlström,⁸⁶ adaptive reaction coordinate force,⁵³ adaptive biasing force,⁵⁶ metadynamics,^{54,55} and filling potential⁸⁷ methods. The first practically useful implementation of this approach in the context of (bio)molecular systems with explicit solvation is probably the local elevation (LE) method of Huber, Torda, and van Gunsteren,⁸⁴ as implemented in the GROMOS96 program.^{88,89} In this method, the searching enhancement is applied along a subset of degrees of freedom of the system (LE-subspace), typically a limited

set of conformationally relevant dihedral angles, by means of a penalty potential defined as a sum of local (grid-based) repulsive functions, the magnitudes of which are made proportional to the number of previous visits to the corresponding conformation (grid cell). Because memory-based searching methods (such as the LE method) have a time-dependent Hamiltonian, they sample in principle no well-defined configurational probability distribution, that is, the resulting trajectories cannot be used for the evaluation of thermodynamic properties (including free energies) via statistical mechanics. However, in view of their very high searching power, there has been a long-standing interest in using their basic principle to design efficient conformational-sampling methods, that is, leading to trajectories suited for the evaluation of thermodynamic properties. This can be done by observing that at the end of a memory-based search, the penalty potential has approximately “flattened” the free-energy hypersurface in the considered subspace up to a certain threshold value above the lowest minimum discovered.⁸⁶ As a result, this final penalty potential represents an optimal biasing potential for a subsequent US simulation. Such a combination is at the heart of the local elevation umbrella sampling^{19,50} (LEUS) method (see below). Note that the LEUS scheme has only been applied until now to conformational free-energy calculations and remains to be generalized to alchemical problems.

Finally, as a third recent development, the so-called λ -dynamics approach^{91–96} has been proposed, in which the λ -variable of an alchemical change (possibly generalized to a λ -vector in the context of multiple changes) evolves dynamically in time along with the physical (atomic) degrees of freedom of the system (extended Lagrangian approach). Taken alone, λ -dynamics does not work very well in practice, because the sampling of the λ -space: (i) is heavily biased toward the regions corresponding to the most stable (lowest free energy) molecules; (ii) is generally hindered by the presence of free-energy barriers and local free-energy minima; and (iii) opens up a large volume of unphysical (thus irrelevant) alchemical space (in the context of multiple free-energy changes), thereby reducing the statistics relevant for the physical molecules. However, this approach has an important merit. It shows that any alchemical free-energy calculation can be reformulated as a pseudo-conformational calculation in an extended space including the λ -variables. In other words, methods that have been developed for the evaluation of conformational free-energy differences (e.g., DC, US and LEUS) can as well be applied to alchemical changes in the context of an extended-system dynamics. Note that a similar principle underlies the combination of LEUS with EDS for alchemical changes, as used previously⁹⁷ and in the present work (see below). The inverse mapping, that is, the application of methods that have been developed for the evaluation of alchemical free-energy changes (e.g., TI, FE, and EDS) to conformational changes, is also in principle possible, requiring the introduction of constraints along the relevant physical degrees of freedom of the system.^{14,28,68,98–100} However, because the constrained degrees of freedom are generally non-Cartesian coordinates, the resulting (projected) conformational subspace is typically non-Euclidean, which

raises a number of issues related to the Jacobian of the transformation and the possible occurrence of metric-tensor effects in MD simulations.^{64–66,98,101,102} These problems do not arise when extending a physical system to include an alchemical subspace, as long as the resulting extended space is chosen to be Euclidean. This is always possible by construction (provided that the physical space itself is Euclidean, i.e., in the absence of constrained internal coordinates) given the unphysical (thus irrelevant) nature of the geometry selected for this alchemical space.

Considering the above discussion, the LEUS method (including its generalization by combination with λ -dynamics or EDS) appears to represent a powerful scheme for the evaluation of both conformational and alchemical free-energy changes, including the determination of multiple changes from a single simulation.

As detailed in the original article,¹⁵ the LEUS scheme consists of two steps: (i) a LE build-up (searching) phase, that is used to progressively construct an optimized memory-based biasing potential within a LE-subspace of N_{LE} conformationally relevant degrees of freedom; and (ii) an US sampling phase, where this potential, now frozen, is used to generate a biased ensemble with extensive coverage of the US-subspace defined by the same $N_{\text{US}} = N_{\text{LE}}$ degrees of freedom. A successful build-up phase will produce a biasing potential that is approximately equal to the negative of the free-energy hypersurface within the considered subspace up to a certain free-energy level, so that a sufficiently long sampling phase will result in a nearly homogeneous coverage of the corresponding region. In addition, because the biasing potential in this second phase is time-independent, thermodynamic information relevant for the physical (unbiased) ensemble can be recovered from the simulated data by means of a simple reweighting procedure.^{15,35,36}

The LEUS scheme is a powerful sampling-enhancement technique in cases where the relevant conformational subspace is of low dimensionality.^{15,37,90} However, this scheme becomes inapplicable in its original form for systems where the dimension of this subspace exceeds a few degrees of freedom. If the relevant subspace is defined by $N_{\text{LE}} = N_{\text{US}} = N$ degrees of freedom, each degree of freedom being discretized by means of N_g grid points, and the biasing potential is expected to map out a fraction f of this subspace, the number of local functions required is $f(N_g)^N$. This number increases exponentially with N , so that the original LEUS approach rapidly becomes intractable, in terms of both memory and build-up duration requirements. A tentative solution to this problem,³⁷ fragment-based LEUS (FB-LEUS), relies on the preoptimization of fragment-based biasing potentials of low dimensionalities, followed by their simultaneous application to each of the corresponding fragments in a molecule, based on a similar principle as suggested in refs 103 and 104. This corresponds to a situation where $N_{\text{US}} = N_F N_{\text{LE}}$, where N_F is the number of fragments in the considered molecule. In principle, the resulting biasing potential should remain appropriate in situations where the free-energy function in the N_{US} -dimensional relevant subspace of the molecule can be approximated as a sum of N_{LE} -dimensional fragment-based contributions, thereby neglecting

the corresponding correlations. In this case, the fragment-based biasing potential will still lead to a nearly homogeneous coverage of this subspace, while the neglected correlations are reintroduced during the reweighting procedure. Application of the FB-LEUS scheme to solvated polyalanine and polyvaline oligopeptides using biasing potentials designed for the corresponding (blocked) monopeptide fragments confirmed the above suggestion in the context of these systems.³⁷ However, it revealed another problem, namely that the enhancement of the searching power (volume of conformational space visited in a given amount of simulation time) is largely offset by a deterioration of the statistical efficiency (representativeness of the biased ensemble in terms of the conformational distribution appropriate for the physical ensemble). If the “flattening” of the free-energy hypersurface largely increases the rate at which new conformations are generated, it also includes a very large number of high free-energy conformations into the sampling. This does not represent a problem for low-dimensionality systems, where the simulation time scale is sufficient to afford the sampling of these extra configurations while maintaining reasonable statistics concerning the low free-energy regions. However, this is no longer the case for high-dimensionality problems, where the combinatorial crowding of the biased ensemble with irrelevant configurations leaves virtually no room for statistics concerning the relevant ones. This situation was tentatively remedied³⁷ by the introduction of alternative biasing potentials inducing the “digging” of valleys between the relevant conformational states at the fragment level, rather than the “flattening” of the corresponding conformational basins. While possibly interesting in the context of other polymers (e.g., oligosaccharides), these potentials did not result in a significant sampling enhancement (as compared to plain MD) for the considered oligopeptides, probably due to the already low conformational transition barriers at the level of the corresponding peptide linkages.

Despite being moderately successful per se, this study³⁷ suggested that an efficient memory-based biasing potential for high-dimensional problems should possess the following three characteristics: (i) it should be of low internal dimensionality (acceptable memory requirements); (ii) it should map out a minimal irrelevant volume (acceptable build-up duration requirements, high statistical efficiency); and (iii) it should possess a problem-adapted geometry (involving a priori specification of the conformational regions considered as relevant or irrelevant). The internal dimensionality refers to the dimensionality of the involved memory map, possibly being inscribed within a relevant conformational subspace of much higher dimensionality (e.g., one-dimensional curve within a multidimensional subspace). The irrelevant volume refers to the volume that is neither useful to the sampling of the relevant states nor strictly necessary to ensure a sufficient number of transitions between them. The above requirements can only be satisfied if the states considered to be relevant are defined prior to the design of the biasing potential, that is, this potential must possess a problem-adapted geometry. The above conditions form the basis of the new scheme proposed in the present work. This scheme is termed ball-

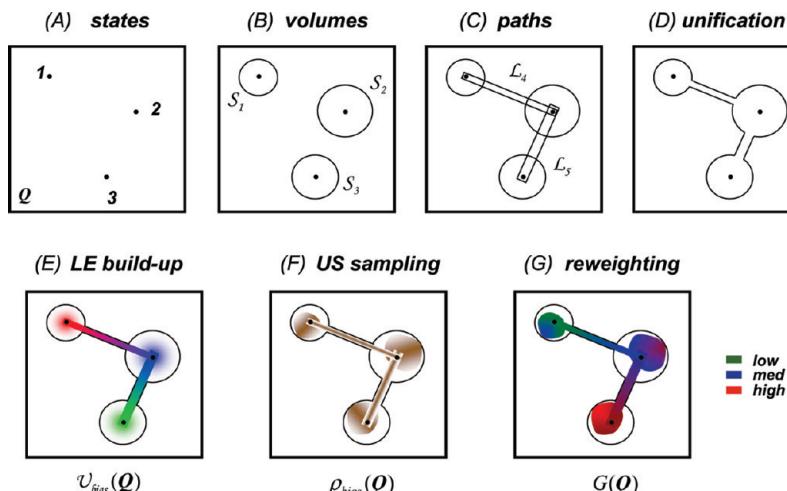


Figure 1. Schematic illustration of the successive steps involved in the B&S-LEUS approach. The reduced conformational space \mathbf{Q} is shown as a two-dimensional plane, and the calculation involves three relevant states (1, 2, and 3). The corresponding $K = 3$ conformational spheres are labeled S_1 , S_2 , and S_3 . They are connected by $L = 2$ conformational lines, labeled L_4 and L_5 . In this illustration, the ranking of the states in terms of increasing free energy is $1 < 2 < 3$.

and-stick LEUS (B&S-LEUS), by reference to the assembly of sampling “balls” (encompassing the relevant states) and “sticks” (used to promote transitions between them) typically involved in the definition of the corresponding biasing potential.

More specifically, the B&S-LEUS approach aims at: (i) the extension of the LEUS scheme to the evaluation of the relative free energies of conformational states in high-dimensionality problems, that is, in cases where the plain LEUS¹⁵ or FB-LEUS³⁷ methods fail; (ii) its extension to the simultaneous determination of multiple conformational free-energy differences from the sampling phase of a single simulation; and (iii) taking a first step toward the generalization of the LEUS scheme to alchemical free-energy calculations, here in conjunction with the EDS scheme. This is achieved by simultaneously taking advantage of the explorative power of the LE searching procedure,⁸⁴ the focusing properties of restraining potentials, the Hamiltonian combination capability of the EDS method,⁷⁸ and the statistical correctness (time-independent Hamiltonian) of the US method.³⁵

In practice, the B&S-LEUS procedure relies on the following steps, illustrated schematically in Figure 1: (A) choice of a reduced conformational subspace permitting the definition of the relevant conformational states; (B) representation of the relevant conformational states by means of centered volumes within this subspace (e.g., conformational spheres or “balls”), each associated with a biasing potential involving a one-dimensional radial memory-based term and a radial confinement (half-harmonic restraining) term; (C) definition of a set of conformational paths connecting the centers of these volumes (e.g., conformational lines or “sticks”), each associated with a biasing potential involving a one-dimensional longitudinal memory-based term and a transverse confinement (flat-bottom half-harmonic restraining) term; (D) unification of the biasing potentials associated with all of the centered volumes and paths into a single biasing potential according to the EDS procedure; (E) optimization of the memory, leading to a biasing potential

enabling a nearly homogeneous sampling (radially within the centered volume, longitudinally within the paths) of the subvolume of the conformational subspace defined by the union of all centered volumes and paths (LE build-up phase); (F) generation of a biased ensemble of configurations using this preoptimized biasing potential (US sampling phase); and (G) calculation of the relative free energies of the states (reweighting and state-assignment procedure).

This Article is organized as follows. Section 2 describes in detail the successive steps (A–G) of the B&S-LEUS scheme. Section 3 introduces the three systems considered here to assess the performance of the scheme and provides the corresponding computational details. These three test systems are: (i) a solvated (blocked) alanine monopeptide (two-dimensional conformational subspace), used as a toy system to illustrate the versatility of the method in promoting the sampling of arbitrary regions of the Ramachandran map; (ii) a solvated (unblocked) polyalanine decapeptide (nine-dimensional conformational subspace), where the method is applied to evaluate the relative free energies of three different types of helices (π , α , and 3_{10}); (iii) a solvated artificial hexopyranose, termed the “mother” of all D-hexopyranoses and constructed as a hybrid of all D-hexopyranose stereoisomers, where the method is applied (seven-dimensional mixed alchemical and conformational subspace) to calculate the relative free energies of the corresponding 32 isomers, anomers, and chair conformers from a single simulation (8 stereoisomers, α - or β -anomer, $^4\text{C}_1$ or $^1\text{C}_4$ chair conformer). Section 4 reports the results of the B&S-LEUS simulations considering these three test systems. Finally, section 5 provides concluding remarks.

2. Method

This section describes the successive steps of the B&S-LEUS scheme, with reference to section 1 (points A–G) and Figure 1.

2.1. Choice of a Reduced Conformational Subspace.

The first step (A) in the B&S-LEUS approach relies on the choice of a reduced conformational subspace permitting the definition of the relevant conformational states. This is done by selecting a subset of N internal coordinates of the physical system, collectively noted by the vector $\mathbf{Q} = \{Q_n, n = 1, \dots, N\}$. These coordinates are assumed to be well-defined and differentiable functions of the vector \mathbf{r} encompassing the Cartesian coordinates of all particles in the system, that is, the vector function $\mathbf{Q} = \mathbf{Q}(\mathbf{r})$ must be defined for any \mathbf{r} and its derivative must be nonsingular.

Examples of possible internal coordinates include, for example, distances between atom pairs, angles between atom triples, dihedral angles between atom quadruples, root-mean-square atomic positional deviations from given reference structures, extended-system variables (e.g., λ -variables in λ -dynamics^{91,92,105}), or any (differentiable) mathematical combination of these. Note that periodic internal coordinates (e.g., angles) should not be “refolded” to a reference period, that is, their time evolution must be continuous. Furthermore, it is assumed that the definition of any internal coordinate (with a specified unit) is associated with the selection of a corresponding reference value σ_n (with the same unit), and that Q_n is defined by the unitless ratio of the two quantities. It is important to stress that the results of a B&S-LEUS simulation depend on a given choice of the σ_n factors, so that these factors must be clearly specified as an integral part of the definition of the conformational subspace.

2.2. Representation of the Relevant Conformational States.

The second step (B) in the B&S-LEUS approach relies on the representation of the relevant conformational states by means of K centered volumes within the reduced conformational subspace. Only the simplest possible type of centered volume will be considered here, namely the sphere.

The biasing potential $\mathcal{B}_k(\mathbf{Q})$ corresponding to a sphere \mathcal{S}_k associated with a state k is defined by the following parameters: a sphere center \mathbf{Q}_k , a radius R_k , a restraining force constant c_k , a number of radial grid points $\Gamma_k + 1$, and a memory force-constant vector $\mathbf{M}_k = \{M_{k,i}, i = 0, \dots, \Gamma_k\}$. The corresponding expression is (for $k = 1, \dots, K$)

$$\mathcal{B}_k(\mathbf{Q}) = \begin{cases} M_{k,\Gamma_k} + \frac{1}{2}c_k(r_k - R_k)^2 & \text{if } r_k \geq R_k \\ \sum_{i=0}^{\Gamma_k} M_{k,i}\gamma(d_{k,i}) & \text{if } r_k < R_k \end{cases} \quad (1)$$

where the function γ is defined as³⁷

$$\gamma(x) = H(1 - |x|)(1 - 3|x|^2 + 2|x|^3) \quad (2)$$

H being the Heaviside step function, and the quantities r_k and $d_{k,i}$ depend on \mathbf{Q} as

$$r_k = \|\mathbf{Q} - \mathbf{Q}_k\| \quad (3)$$

and

$$d_{k,i} = \Gamma_k R_k^{-1} r_k - i \quad (4)$$

Within the sphere ($r_k < R_k$), the memory vector \mathbf{M}_k permits to enforce a radially dependent potential of arbitrary form

(with an approximate resolution $\Gamma_k^{-1}R_k$), expressed as a weighted sum of $\Gamma_k + 1$ repulsive local functions γ . Outside the sphere ($r_k \geq R_k$), the potential is changed to an attractive half-harmonic restraint. It is easily verified that the biasing potential \mathcal{B}_k defined by eq 1 is continuous and differentiable.³⁷

The extension to nonspherical centered volumes (e.g., ellipsoids or polyhedra) is in principle straightforward and requires the generalization of the radius R_k to a function $R_k(\mathbf{Q})$ accounting for the center–surface distance in the different directions. The “radial” grid points will then map to scaled versions of this surface rather than to spheres. In any case, the internal dimensionality of the biasing potential is one, which implies a limited memory cost. Extending this dimensionality so as to include some memory-based directional dependence (still of limited dimensionality) into the biasing potential is certainly feasible, but should not be required if the reduced conformational subspace has been chosen appropriately, that is, if it encompasses enough variables to define all states as distinct regions. A carefully adjusted one-dimensional memory may be used to obtain a biasing potential enforcing a homogeneous radial sampling of the centered volume, but this potential will generally not lead to a homogeneous sampling of the multidimensional volume itself, the directional (nonradial) dimensions remaining unbiased. Note also that such a potential will guarantee that the center of the volume is sampled.

2.3. Definition of a Set of Conformational Paths. The third step (C) in the B&S-LEUS approach relies on the definition of a set of L conformational paths connecting the centers of the volumes representing the K states. Only the simplest possible type of path will be considered here, namely the line (or, more precisely, the line segment). The numbering of the corresponding biasing potentials will start at $K + 1$.

The biasing potential $\mathcal{B}_l(\mathbf{Q})$ corresponding to a line \mathcal{L}_l is defined by the following parameters: a starting point \mathbf{Q}_l , an ending point \mathbf{Q}'_l , a (double) width W_l , a restraining force constant c_l , a number of longitudinal grid points $\Gamma_l + 1$, and a memory force-constant vector $\mathbf{M}_l = \{M_{l,i}, i = 0, \dots, \Gamma_l\}$. The corresponding expression is (for $l = K + 1, \dots, K + L$)

$$\mathcal{B}_l(\mathbf{Q}) = \begin{cases} M_{l,0} + \frac{1}{2}c_l H(r_l - W_l)(r_l - W_l)^2 & \text{if } u_l \leq 0 \\ M_{l,\Gamma_l} + \frac{1}{2}c_l H(r'_l - W_l)(r'_l - W_l)^2 & \text{if } u_l \geq U_l \\ \sum_{i=0}^{\Gamma_l} [M_{l,i} + \frac{1}{2}c_l H(p_l - W_l)(p_l - W_l)^2] \gamma(d_{l,i}) & \text{if } 0 < u_l < U_l \end{cases} \quad (5)$$

where the function γ is defined by eq 2, U_l is the line length

$$U_l = \|\mathbf{Q}'_l - \mathbf{Q}_l\| \quad (6)$$

and the quantities u_l , p_l , r_l , r'_l , and $d_{l,i}$ depend on \mathbf{Q} as

$$u_l = U_l^{-1}(\mathbf{Q}'_l - \mathbf{Q}_l)^T(\mathbf{Q} - \mathbf{Q}_l) \quad (7)$$

$$p_l = \|(\mathbf{Q} - \mathbf{Q}_l) - U_l^{-1} u_l (\mathbf{Q}'_l - \mathbf{Q}_l)\| \quad (8)$$

$$r_l = \|\mathbf{Q} - \mathbf{Q}_l\| \quad (9)$$

$$r'_l = \|\mathbf{Q} - \mathbf{Q}'_l\| \quad (10)$$

and

$$d_{l,i} = \Gamma_l U_l^{-1} u_l - i \quad (11)$$

v^T indicating the transpose of a vector v . Note that due to the identity³⁷

$$\sum_{i=0}^{\Gamma_l} \gamma(d_{l,i}) = 1 \quad \text{if } 0 < u_l < U_l \quad (12)$$

the third conditional statement in eq 5 could in principle be simplified to

$$\mathcal{B}_l(\mathbf{Q}) = \frac{1}{2} c_l H(p_l - W_l)(p_l - W_l)^2 + \sum_{i=0}^{\Gamma_l} M_{l,i} \gamma(d_{l,i}) \quad \text{if } 0 < u_l < U_l \quad (13)$$

This simplification was not undertaken, so as to allow for a possible generalization to displaced lines and lines with longitudinally dependent widths or force constants (see below). The quantity u_l represents the longitudinal distance between the starting point of the line and the current point \mathbf{Q} , while the parameter p_l represents the corresponding transverse (perpendicular) distance. Within the line ($0 < u_l < U_l$), the memory vector M_l permits to enforce a longitudinally dependent potential of arbitrary form (with an approximate resolution $\Gamma_l^{-1} U_l$), expressed as a weighted sum of $\Gamma_l + 1$ repulsive local functions γ , and applied together with a transverse attractive flat-bottom (width W_l) half-harmonic restraining potential. Outside the line, that is, when going past its two terminal points in terms of longitudinal distance ($u_l \leq 0$ or $u_l \geq U_l$), the potential is changed to an attractive flat-bottom (width W_l) half-harmonic restraint depending on the distance to the corresponding end point. It is easily verified that the biasing potential \mathcal{B}_l defined by eq 5 is continuous and differentiable.

The extension to nonlinear paths (e.g., arbitrary curves) is in principle possible and requires the generalization of the end points \mathbf{Q}_l and \mathbf{Q}'_l to a parametric path $\mathbf{Q}_l(u)$, where $\mathbf{Q}_l(0)$ is the starting point and $\mathbf{Q}_l(U_l)$ is the ending point, U_l being the path length. A simpler variant of this approach, the displaced line, involves modifying a normal line by allocating to all nonterminal grid points an offset coordinate $\Delta\mathbf{Q}_{l,i}$ perpendicular to the line (i.e., with $\Delta\mathbf{Q}_{l,i}^T (\mathbf{Q}'_l - \mathbf{Q}_l) = 0$ and $\Delta\mathbf{Q}_{l,0} = \Delta\mathbf{Q}_{l,\Gamma_l} = 0$), and replacing p_l in eq 5 by

$$p_{l,i} = \|(\mathbf{Q} - \mathbf{Q}_l) - U_l^{-1} u_l (\mathbf{Q}'_l - \mathbf{Q}_l) - \Delta\mathbf{Q}_{l,i}\| \quad (14)$$

Another possible variant involves the use of longitudinally dependent line widths or/and restraining force constants, that is, the replacement of c_l and W_l in eq 5 by corresponding grid-point dependent quantities $c_{l,i}$ and $W_{l,i}$. In any case, the internal dimensionality of the biasing potential is one, which implies a limited memory cost. Extending this dimensionality

so as to include some memory-based transverse dependence (still of limited dimensionality) into the biasing potential is certainly feasible, but not required in practice, because paths are only meant to promote transitions between states and typically located in irrelevant regions of the reduced conformational subspace. A carefully adjusted one-dimensional memory may be used to obtain a biasing potential enforcing a homogeneous longitudinal sampling of the path, but this potential will generally not lead to a homogeneous transverse sampling. Note also that, in contrast to the centered volume case, such a potential will not automatically guarantee that the two end points of the path are sampled, although this will generally be the case in practice if the line width is sufficiently small (or can be enforced otherwise by decreasing the line width close to the end points).

In principle, the L paths will be chosen to connect pairs among the K centered volumes defining the states. This must be done in such a way that all states are connected to each other via at least one path or succession thereof. The minimum number of paths is thus $L = K - 1$ (maximum-spanning tree), but it may be advantageous in terms of convergence properties to include additional (redundant) paths. Note that the end points of the paths must be identical to the centers of the states (e.g., they should not connect to the periphery of the centered volumes). This is essential because independent biasing potentials leading to a homogeneous radial sampling of the centered volumes and longitudinal sampling of the paths can only guarantee that these specific points are sampled (for lines, assuming sufficiently small line widths, at least at the end points).

2.4. Unification of the Biasing Potentials. The fourth step (*D*) in the B&S-LEUS approach relies on the unification of the biasing potentials associated with the $M = K + L$ centered volumes and paths into a single biasing potential according to the EDS procedure.⁷⁸ For the ease of notation, these objects have been given the generic notation B_m , where the index m ranges from 1 to $M = K + L$ (1, ..., K for the centered volumes, $K + 1$, ..., M for the paths). The various LEUS potentials are combined following the EDS principle as

$$\mathcal{U}_{\text{bias}}(\mathbf{r}; \mathbf{M}) = -\frac{1}{\beta s} \ln \left(\sum_{m=1}^M \exp[-\beta s \mathcal{B}_m(\mathbf{Q}(\mathbf{r}))] \right) \quad (15)$$

where $\beta = (k_B T)^{-1}$, k_B being Boltzmann's constant and T the absolute temperature, \mathbf{r} represents the system configuration (Cartesian coordinates of all particles), $\mathbf{Q}(\mathbf{r})$ the corresponding representative point in the reduced subspace, s a (positive) smoothing parameter, and \mathbf{M} the joint memories of the M objects, that is, a vector containing $N_M = \sum_{m=1}^M (\Gamma_m + 1)$ elements.

Qualitatively speaking, the exponential weighting in eq 15 ensures that the combined biasing potential $\mathcal{U}_{\text{bias}}$ is low in the regions of the conformational subspace where any of the \mathcal{B}_m is low, and high in the regions where all of the \mathcal{B}_m are high. For the ease of reference, the subvolume of the reduced conformational subspace where any of the \mathcal{B}_m is low, that is, the union of all centered volumes and paths, will be referred to as the active subspace. In the absence of

memory ($\mathbf{M} = \mathbf{0}$) and assuming a hypothetical (physical-system) dynamics leading to a homogeneous sampling in terms of \mathbf{Q} , the introduction of $\mathcal{U}_{\text{bias}}$ will effectively restrict the sampling to the active subspace by action of the individual restraining potentials within the M objects, with a slight bias toward regions where multiple objects overlap. In the presence of a memory ($\mathbf{M} \neq \mathbf{0}$) and considering a real molecular system (physical potential energy function affecting the sampling in terms of \mathbf{r}), the regions sampled will be determined by an interplay between five factors: (i) a strong bias toward the active subspace induced by the restraints; (ii) a slight bias toward the regions where multiple objects overlap; (iii) a physical bias in terms of \mathbf{r} ; (iv) a bias related to the Jacobian of the $\mathbf{Q}(\mathbf{r})$ transformation; and (v) a memory-based bias within the objects. The goal of the LE build-up phase will be to adjust the memory so that the four latter types of biases cancel out, leading, in conjunction with the first bias, to a nearly homogeneous (radial within the centered volumes, longitudinal within the paths) sampling, restricted to the active subspace. Note that the requirements of homogeneous sampling within all single objects may in some cases be conflicting at the level of the unified biasing potential, in regions where multiple objects overlap. However, these homogeneity violations are expected to be marginal, especially in high-dimensional cases.

The forces derived from U_{bias} in eq 15 are given by

$$\mathbf{F}_{\text{bias}}(\mathbf{r}) = -\frac{\partial \mathcal{U}_{\text{bias}}(\mathbf{r}; \mathbf{M})}{\partial \mathbf{r}} = -\sum_{m=1}^M w_m(\mathbf{Q}) \frac{d\mathcal{B}_m(\mathbf{Q})}{d\mathbf{Q}} \frac{\partial \mathbf{Q}}{\partial \mathbf{r}} \quad (16)$$

where

$$w_m(\mathbf{Q}) = \frac{\exp[-\beta s \mathcal{B}_m(\mathbf{Q})]}{\sum_{m=1}^M \exp[-\beta s \mathcal{B}_m(\mathbf{Q})]} \quad (17)$$

can be interpreted as measuring the relative influence (weight) of a single-object biasing potential m on the dynamics of the system in a conformation \mathbf{Q} . Note that the forces defined by eq 16 are nonsingular, because $\mathcal{B}_m(\mathbf{Q})$ and $\mathbf{Q}(\mathbf{r})$ are both differentiable functions of their arguments. The parameter s in eqs 15 and 17 will affect the extent to which differences between the M single-object biasing potentials are “smoothed out” in $\mathcal{U}_{\text{bias}}$. Because the paths connect to the states at their centers, this parameter will have little influence on the biasing potential within the active subspace. On the other hand, it will affect the “sharpness” with which the restraining potentials define this subspace, a high value indicating a “sharper” combination (with the risk of occurrence of high restraining forces) and a low value a more “fuzzy” combination (with the risk of unnecessarily increasing the irrelevant volume). Because this parameter is nevertheless expected to have a minor overall influence on the B&S-LEUS scheme when selected within reasonable bounds, it was simply set to $s = 1$ in the present study.

2.5. LE Build-Up Phase. The fifth step (*E*) of the B&S-LEUS procedure (LE build-up phase, duration t_{LE}) relies on the optimization of the memory, leading to a biasing potential

enabling nearly uniform sampling (radially within the centered volumes, longitudinally within the paths) of the active subspace. In this phase, the (time-dependent) potential-energy function used for (thermostatted) MD searching is written

$$\mathcal{U}(\mathbf{r}; \mathbf{M}(t)) = \mathcal{U}_{\text{phys}}(\mathbf{r}) + \mathcal{U}_{\text{bias}}(\mathbf{r}; \mathbf{M}(t)) \quad (18)$$

where $\mathcal{U}_{\text{phys}}$ is the physical potential-energy function (force field) and $\mathcal{U}_{\text{bias}}$ is given by eq 15. The updating scheme for the memory $\mathbf{M}(t)$ relies on the equation

$$\mathbf{M}_{m,i}(t + \Delta t) = \mathbf{M}_{m,i}(t) + k_{\text{LE}} f_{\text{LE}}^{I_{\mathcal{R}}(t; \gamma_{\text{LE}}, n_{\text{LE}})} j_{m,i}(\mathbf{Q}) h_{m,i}(\mathbf{Q}) w_m(\mathbf{Q}) \quad (19)$$

where $\mathbf{M}_{m,i}$ is the memory associated with grid point i of object m , $\mathbf{Q} = \mathbf{Q}(\mathbf{r}(t))$, Δt is the simulation time step, k_{LE} is the basis force-constant increment, f_{LE} is a force-constant reduction factor, $I_{\mathcal{R}}$ is a force-constant reduction counter, associated with a defined conformational region \mathcal{R} , γ_{LE} is a local visiting cutoff (real), n_{LE} is a global visiting cutoff (integer), $j_{m,i}$ is a distribution-alteration function, $h_{m,i}$ is a grid-assignment function, and w_m is the weight defined by eq 17. In the absence of prior knowledge concerning the form on the free-energy hypersurface, the memory will typically be initiated to $\mathbf{M}(0) = \mathbf{0}$. The different factors involved in eq 19 are explained below.

The grid-assignment function evaluates to one for a single grid point in each of the single-object biasing potentials (and to zero for all other grid points), namely the grid point i in object m that is (radially for centered volumes, longitudinally for paths) closest to \mathbf{Q} . For a sphere k , one has ($k = 1 \dots K$)

$$h_{k,i}(\mathbf{Q}) = \begin{cases} \delta_{i,\Gamma_k} & \text{if } r_k \geq R_k \\ \delta_{i,\text{NINT}(\Gamma_k R_k^{-1} r_k)} & \text{if } r_k < R_k \end{cases} \quad (20)$$

where δ is the Kronecker symbol and the function NINT returns the nearest integer to a real number. For a line l , one has ($l = K + 1, \dots, M$)

$$h_{k,i}(\mathbf{Q}) = \begin{cases} \delta_{i,0} & \text{if } u_l \leq 0 \\ \delta_{i,\Gamma_l} & \text{if } u_l \geq U_l \\ \delta_{i,\text{NINT}(\Gamma_l U_l^{-1} r_p)} & \text{if } 0 < u_l < U_l \end{cases} \quad (21)$$

As a result, the build-up always affects one and only one grid point in each of the M single-object memories. However, the presence of the weight factor w_m in eq 19 ensures that the build-up is only significant within the objects encompassing or closest to point \mathbf{Q} (note that the sum of w_m over all objects is one).

The distribution-alteration function is generally set to

$$j_{m,i}(\mathbf{Q}) = 1 \quad (22)$$

leading to a nearly homogeneous sampling (radially within the centered volumes, longitudinally within the paths) of the active subspace. However, this function may be used to enforce deviations from this homogeneous sampling. As a simple example, one may observe that the volume of relevant conformational subspace accounted for by a radial grid point i within a sphere k (distance $\Gamma_k^{-1} i R_k$ from the center) increases

with $(i_k + 1/2)^{N-1}$ (Jacobian factor), where N is the subspace dimensionality. One may then decide to bias the sampling of the sphere toward its periphery, which can be achieved by setting for all spheres k ($k = 1, \dots, M$)

$$j_{k,i}(\mathbf{Q}) = \frac{\left(i + \frac{1}{2}\right)^{1-N}}{\sum_{j=0}^{i_k} \left(j + \frac{1}{2}\right)^{1-N}} \quad (23)$$

Unless otherwise specified, eq 22 (rather than eq 23) was employed in the present study.

The force-constant reduction factor can be used in the context of an iterative procedure to progressively decrease the build-up rate during the searching phase. As noted previously by other authors,^{86,106} a high build-up rate is desired in the early stage of the searching, where the deep free-energy basins have to be “filled up” coarsely (i.e., without wasting computer time), while a low build-up rate (near-equilibrium situation) is preferable in the later stage, where the remaining shallower free-energy wiggles have to be “leveled off” (so as to produce a close-to-optimal biasing potential). This can be achieved by a progressive reduction of the build-up rate, enforced in eq 19 by using $f_{\text{LE}} < 1$ along with a force-constant reduction counter $I_{\mathcal{R}}$ progressively increasing with time (the choice $f_{\text{LE}} = 1$ switches off the force-reduction procedure). In the B&S-LEUS algorithm, the force-reduction procedure is associated with a region \mathcal{R} within the active subspace, defined by a specific collection of grid points. The reduction counter $I_{\mathcal{R}}$ is propagated in time according to the following procedure. $I_{\mathcal{R}}(t)$ as well as an auxiliary counter $N_c(t)$ are set to 0 at $t = 0$. An auxiliary memory $A(t)$ is also set to 0 at $t = 0$ and propagated in time according to the equation

$$A_{m,i}(t + \Delta t) = A_{m,i}(t) + w_m(\mathbf{Q})h_{m,i}(\mathbf{Q}) \quad (24)$$

When $A_{m,i}(t)$ exceeds a specified local visiting cutoff γ_{LE} for all grid points $(m,i) \in \mathcal{R}$, the auxiliary counter is increased by one and the auxiliary memory reset to zero. When the auxiliary counter exceeds a global visiting cutoff n_{LE} , $I_{\mathcal{R}}$ is increased by one and the auxiliary counter reset to zero. Two possible (reasonable) choices for \mathcal{R} are: (i) the $i = 0$ (central) grid points of all centered volumes k ($k = 1, \dots, K$), a choice that will be noted $\mathcal{R} = \mathcal{C}$; and (ii) all grid points i of all objects m ($m = 1, \dots, M$), a choice that will be referred to as $\mathcal{R} = \mathcal{A}$. Possible (reasonable) choices for the parameters γ_{LE} and n_{LE} are 1.0 and 2, respectively. The reasoning behind the present force-constant reduction scheme (assuming $\gamma_{\text{LE}} = 1.0$ and $n_{\text{LE}} = 2$) is that when all grid points of \mathcal{R} have undergone an “effective” number of visits (auxiliary memory, i.e., based on the w_m weights) of one, it is still possible that the “flattened” free-energy hypersurface retains an overall “slope”. However, when all of these points have undergone an “effective” number of visits of one for the second time, even the points that were “uphill” have been revisited. When this condition is met, it becomes advantageous to reduce the build-up rate by incrementing $I_{\mathcal{R}}$, which in effect scales this rate by a factor f_{LE} .

Finally, the constant k_{LE} in eq 19 represents the basic force-constant increment (units of energy) and determines the initial rate of the build-up. Note that the above force-reduction procedure also presents the advantage of permitting a convergence assessment of the build-up phase, by monitoring the time evolution of $I_{\mathcal{R}}$. The build-up phase can, for example, be terminated whenever $I_{\mathcal{R}}$ reaches a threshold value $I_{\mathcal{R}}^{\max}$. In this case, the procedure guarantees that all grid points of $I_{\mathcal{R}}$ have undergone an “effective” number of visits of at least $n_{\text{LE}}\gamma_{\text{LE}}I_{\mathcal{R}}^{\max}$, while the energetic resolution of the biasing potential is of the order of $f_{\text{LE}}^{\max} k_{\text{LE}}$. Alternatively, the termination may be based on the time interval separating successive increments of $I_{\mathcal{R}}$. In the initial stage of the build-up, the diffusion of the system within the active subspace will be accelerated (hill surfing). However, as the free-energy hypersurface becomes increasingly “flat” and the build-up rate is decreased, this diffusion will progressively slow down toward a “natural” regime (as determined by the physical system after removal of the free-energy bias). Thus, the force-reduction procedure could also be terminated when the interval separating successive increments of $I_{\mathcal{R}}$ has increased and leveled off to an approximately constant time.

Two specific features of the B&S-LEUS scheme make it particularly robust with respect to the details of the build-up protocol and well-suited for an automatic force-reduction procedure, as compared to the standard LEUS scheme³⁷ and related approaches.^{52,54} First, in the B&S-LEUS scheme, the grid points are defined a priori, and it is known that all of the corresponding grid cells must have been sufficiently sampled during a successful build-up phase. In contrast, in the standard LEUS scheme, new grid points are steadily added, and because the total number of possible grid points within the reduced subspace is typically extremely large, the build-up is stopped long before all of them have been visited. As a result, it can never be guaranteed that all of the relevant grid points have actually been visited, and an inappropriate build-up procedure (e.g., too high build-up rate or too fast force-reduction procedure) may lead to the omission of important conformational regions. Second, in the B&S-LEUS scheme, the effective volume of the active subspace is unaffected by the build-up. The reason is that the confinement potentials are “attached” to the surface of the different objects (centered volumes or paths, see eqs 1 and 5). As a result, they “rise” simultaneously with the memory-based component of the biasing potential. This can equivalently be seen by observing that the addition of an arbitrary vector with identical components to the memory \mathbf{M} in eq 15 only changes $\mathcal{U}_{\text{bias}}$ by a constant, that is, does not affect the dynamics. In contrast, in the standard LEUS scheme, the “rise” of the memory-based potential slowly expands the accessed conformational volume, predominantly including irrelevant (high free energy) regions.

Similarly to the original LEUS method,¹⁵ a successful LE build-up phase will generate a biasing potential that is (approximately) equal to the negative of the free-energy function $G(\mathbf{Q})$ in the active subspace, that is

$$\mathcal{U}_{\text{bias}}(\mathbf{r}; \mathbf{M}(t_{\text{LE}})) \approx -G(\mathbf{Q}) \quad (25)$$

There is a slight difference in the interpretation of this approximate equality within the plain LEUS and B&S-LEUS approaches. In plain LEUS, this result will only be valid up to a certain free-energy threshold above the lowest free-energy point encountered, the latter threshold depending on the duration and rate of the build-up. However, the presence of regions with a free energy lower than this threshold, but separated by high barriers from the free-energy basin that has been searched, cannot be ruled out. In the B&S-LEUS scheme, the states to be sampled are known in advance, and it can be verified explicitly that all of them have been visited (e.g., definition of \mathcal{R} in the force-reduction procedure). However, it still cannot be guaranteed that eq 25 holds everywhere within the active subspace, because the sampling within the states is only enhanced along one (radial) degree of freedom, leaving room for barriers along the $N - 1$ remaining ones. In other words, $\mathcal{U}_{\text{bias}}$ may be characteristic of a limited subregion of each state, while G characterizes the entire state.

As discussed elsewhere,¹⁵ the approximate equality in eq 25 is in general not sufficient for an accurate evaluation of the relative free energies of the conformational states, which requires a subsequent sampling phase.

2.6. US Sampling Phase. The sixth step (*F*) of the B&S-LEUS procedure (US sampling phase, duration t_{US}) relies on the generation of a biased ensemble of configurations, using the biasing potential preoptimized during the LE build-up phase. In this phase, the (time-independent) potential-energy function used for (thermostatted) MD sampling is written

$$\mathcal{U}(\mathbf{r}; \mathbf{M}) = \mathcal{U}_{\text{phys}}(\mathbf{r}) + \mathcal{U}_{\text{bias}}(\mathbf{r}; \mathbf{M}) \quad (26)$$

where \mathbf{M} is from here on a short notation for $\mathbf{M}(t_{\text{LE}})$. Due to eqs 15 and 25, the biased sampling during this phase should be approximately homogeneous (radially within the centered volumes, longitudinally along the paths) within the active subspace. In particular, all states should be sampled with equal populations (assuming identical radii and numbers of grid points for all spheres). In addition, due to the presence of the connecting paths, diffusive interconversion transitions should occur frequently between these states. As a result, an accurate evaluation of their relative free energies becomes possible.

2.7. Reweighting and State Assignment. The seventh step (*G*) of the B&S-LEUS procedure (reweighting and state assignment) involves the postprocessing of the data accumulated during the sampling phase, so as to calculate the relative free energies of the states in the physical ensemble. For each state k , the free energy can be written (for $k = 1, \dots, K$):

$$G_k = -\beta^{-1} \ln \langle \exp[\beta \mathcal{U}_{\text{bias}}(\mathbf{r}; \mathbf{M})] \rangle_{Q(\mathbf{r}) \in \mathcal{J}'_k} + C_G \quad (27)$$

where C_G is an arbitrary offset constant (typically chosen so that $G_k = 0$ for the lowest free-energy state) and $\langle \dots \rangle_{Q(\mathbf{r}) \in \mathcal{J}'_k}$ denotes ensemble (trajectory) averaging over the biased ensemble (sampling phase), restricted to conformations belonging to state k . The symbol \mathcal{J}'_k has been used rather than \mathcal{J}_k to underline the fact that the regions used to assign

the states need not necessarily be exactly identical to the centered volumes involved in the construction of the biasing potential. Although the calculation of relative free energies is the main concern of the present work, it should be stressed that reweighting formulas analogous to eq 24 can be formulated for any other type of thermodynamic (e.g., enthalpy, entropy, heat capacity, or volume) or structural (ensemble average of a given instantaneous observable) quantity. Note that the reweighting requires the quantity $\mathcal{U}_{\text{bias}}$ to be stored along with each successive frame along the trajectory.

2.8. Additional Remarks. The B&S-LEUS approach involves “physical” parameters, namely the definition of the reduced conformational subspace and the choice of the centered volumes (central conformation and radius) associated with the states. In favorable situations, the free-energy hypersurface will involve well-defined basins around the central conformations selected to represent the different states. In this case, the results should be relatively insensitive to the selected radii, provided that they are chosen large enough to encompass the low free-energy regions of the different states. In practice, however, basins may not be centered at the selected reference conformations, for example, due to an inaccurate choice of these conformations or to the approximate force-field representation of the system. Consequently, the sensitivity of the calculated relative free energies to these parameters should be investigated. On the long run, a method where the centers and radii (and possibly shapes) of the centered volumes defining the states are refined adaptively could be envisioned.

The B&S-LEUS approach also involves “numerical” parameters, namely the choice of a specific set of paths along with the corresponding path widths, the restraining force constants, the numbers of grid points used for the centered volumes and paths, the EDS smoothing parameter, the various parameters of the build-up procedure, the build-up time t_{LE} , and the sampling time t_{US} . The sensitivity of the calculated relative free energies to these parameters should also be investigated.

Although the B&S-LEUS approach has been formulated here in terms of a common reduced space \mathbf{Q} for all objects (e.g., set of conformationally relevant dihedral angles), this requirement is actually not critical to the method. Of particular interest would be the definition on centered volumes based on the root-mean-square atomic positional deviation from corresponding reference structures. The paths would then represent arbitrary configurational pathways interconverting the reference structures of the connected states (these could be constructed, e.g., via reaction-path approaches^{107–110} or targeted molecular dynamics^{28,111,112}). The latter variant would be practically extremely relevant, providing a direct connection (via simulation and statistical mechanics) between three-dimensional structures (as available experimentally from X-ray or NMR determinations) and free-energy differences (between conformational states defined as their configurational neighborhoods).

As a final remark, it should be kept in mind that the B&S-LEUS approach will only be able to handle conformational states of “limited” extents (i.e., amenable to sufficient

Table 1. Characteristics of the Conformational Objects (Spheres or/and Lines) Defining the B&S-LEUS Biasing Potentials in the Different Simulations^a

system	<i>M</i>	object	\mathbf{Q}_m	\mathbf{Q}'_m	R_m	W_m	c_m [kJ mol ⁻¹]	$\Gamma_m + 1$
A ₁	0							
A ₂	1	\mathcal{S}_1	(180,180)		45		0.5	10
A ₃	1	\mathcal{S}_1	(180,180)		45		0.5	10
A ₄	1	\mathcal{L}_1	(60,60)	(300,300)		10	0.1	20
A ₅	1	\mathcal{L}_1	(60,60)	(300,300)		<i>b</i>	0.1	20
A ₆	1	\mathcal{L}_1^c	(60,60)	(300,300)		10	0.1	20
A ₇	2	\mathcal{L}_1	(60,60)	(300,300)		10	0.5	20
		\mathcal{L}_2	(300,60)	(60,300)		10	0.5	20
A ₈	6	\mathcal{L}_1	(60,60)	(60,300)		10	0.5	20
		\mathcal{L}_2	(60,60)	(300,60)		10	0.5	20
		\mathcal{L}_3	(60,60)	(300,300)		10	0.5	20
		\mathcal{L}_4	(60,300)	(300,300)		10	0.5	20
		\mathcal{L}_5	(300,60)	(60,300)		10	0.5	20
		\mathcal{L}_6	(300,60)	(300,300)		10	0.5	20
A ₉	8	\mathcal{L}_1	(50,60)	(70,300)		10	0.5	20
		\mathcal{L}_2	(160,60)	(180,300)		10	0.5	20
		\mathcal{L}_3	(230,60)	(250,300)		10	0.5	20
		\mathcal{L}_4	(290,60)	(310,300)		10	0.5	20
		\mathcal{L}_5	(50,60)	(90,60)		10	0.5	20
		\mathcal{L}_6	(60,180)	(100,180)		10	0.5	20
		\mathcal{L}_7	(70,300)	(250,300)		10	0.5	20
		\mathcal{L}_8	(240,180)	(300,180)		10	0.5	20
A ₁₀	6	\mathcal{S}_1	(60,60)		45		0.5	10
		\mathcal{S}_2	(60,300)		45		0.5	10
		\mathcal{S}_3	(300,60)		45		0.5	10
		\mathcal{S}_4	(300,300)		45		0.5	10
		\mathcal{L}_1	(60,60)	(300,300)		10	0.5	20
		\mathcal{L}_2	(300,60)	(60,300)		10	0.5	20
P ₁	4	\mathcal{S}_1	(-130) ₉		115.5		0.02	15
		\mathcal{S}_2	(-105) ₉		115.5		0.02	15
		\mathcal{S}_3	(-75) ₉		115.5		0.02	15
		\mathcal{L}_1	(-130) ₉	(-75) ₉		8.25	0.02	21
P ₂	1	\mathcal{S}_1	(-150) ₉	(-50) ₉		100	0.02	21
H ₁	63	\mathcal{S}_{1-32}	<i>d</i>		50.0		1.0	6
		\mathcal{A}_{-31}	<i>e</i>	<i>e</i>		10.0	1.0	14
H ₂	112	β_{1-32}	<i>d</i>		50.0		1.0	6
		\mathcal{L}_{1-80}	<i>f</i>	<i>f</i>		10.0	1.0	14

^a The blocked alanine monopeptide (A₁–A₁₀), unblocked polyalanine decapeptide (P₁ and P₂), and artificial hexopyranose (H₁ and H₂) systems, as well as the corresponding relevant conformational subspaces (two, nine, and seven dimensions, respectively) are described in sections 3.1–3.3. For each object *m* (*m* = 1, ..., *M*, with *M* = *K* + *L*, where *K* is the number of spheres and *L* the number of lines), the following quantities are reported: center (sphere) or starting point (line) \mathbf{Q}_m , ending point (line only) \mathbf{Q}'_m , radius (sphere only) *R_m*, width (line only) *W_m*, restraining force constant *c_m*, and number of radial (sphere) or longitudinal (line) grid-points $\Gamma_m + 1$. For the decapeptide system, \mathbf{Q}_m and \mathbf{Q}'_m are nine-dimensional vectors with identical components, indicated as (*Q*)₉, where *Q* is the corresponding common value. The corresponding B&S-LEUS protocol parameters are listed in Table 2. ^b Longitudinally dependent width, as described in section 2.3, using *W₁* = 10 (6, 5, 5, 4, 4, 3, 3, 2, 2, 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6). ^c Displaced line, as described in section 2.3, using $\Delta\mathbf{Q}_1 = 5(2^{1/2})(-1, 1)\mathbf{d}$, where $\mathbf{d} = (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0)$. ^d Centered at ideal values appropriate for the 32 isomers, as described in section 3.3. ^e Maximum-spanning tree as described in section 3.3 and illustrated in Figure 3b. ^f Redundant spanning tree (Manhattan metric) as described in section 3.3.

sampling on the time scale accessible to the simulation). For example, an extension of the scheme to the determination of the folding free energy of peptides (or proteins) appears difficult, because a proper sampling of the unfolded state would require longer time scales than currently feasible, except for the smallest systems.

3. Computational Details

Three test systems were considered to assess the performance of the B&S-LEUS scheme. This section provides the computational details concerning these systems and the corresponding simulations. In the three cases, the reduced conformational subspace is defined in terms of a set of angular coordinates. The corresponding reference values σ_n (section 2.1) are systematically set to $\sigma = 1$ degree, leading to unitless variables. This convention is applied throughout,

but unitless quantities *X* (e.g., coordinates Q_n , sphere radii *R_k*, and line widths *W_k*) are sometimes reported as σX in units of degrees for the clarity of the discussion. Note that, although the variables Q_n are not “refolded” to a reference period during the calculations, this “refolding” is actually performed in the displayed figures.

3.1. Blocked Alanine Monopeptide. A solvated (blocked) alanine monopeptide was used as a toy system to illustrate the versatility of the B&S-LEUS method in promoting the sampling of arbitrary regions of the Ramachandran map. This monopeptide consists of an N-acetylated (ac) and C-methylamidated (am) alanine residue (see Figure 1 in ref 37) and was simulated in the presence of $N_{\text{H}_2\text{O}} = 1300$ water molecules (see section 3.4 for details). The two-dimensional reduced conformational subspace was defined by the variables $Q_1 = \sigma^{-1}\phi$ and $Q_2 = \sigma^{-1}\psi$, where ϕ and ψ are the

Table 2. Parameters of the B&S-LEUS Protocol Used in the Different Simulations^a

system	<i>s</i>	k_{LE} [kJ mol ⁻¹]	\mathcal{R}	f_{LE}	γ_{LE}	n_{LE}	t_{LE} [ns]	$I_{\mathcal{R}}^{\max}$	t_{US} [ns]
A ₁ ^b		2×10^{-3}		1.0			15		50
A ₂		5×10^{-5}		1.0			5		1
A ₃ ^c		5×10^{-5}		1.0			5		1
A ₄		1×10^{-4}		1.0			5		1
A ₅		1×10^{-4}		1.0			5		1
A ₆		1×10^{-4}		1.0			5		1
A ₇	1.0	2×10^{-4}		1.0			10		1
A ₈	1.0	5×10^{-4}		1.0			20		5
A ₉	1.0	5×10^{-4}		1.0			20		5
A ₁₀	1.0	2×10^{-4}		1.0			20		5
P ₁	1.0	1×10^{-3}	\mathcal{C}	0.5	1	2	50	6	2×50^d
P ₂	1.0	5×10^{-3}	\mathcal{A}	0.5	1	2	20	7	2×50^d
H ₁	1.0	5×10^{-2}	\mathcal{C}	0.5	1	2	100	6	100
H ₂	1.0	5×10^{-2}	\mathcal{A}	0.5	1	2	100	8	100

^a The blocked alanine monopeptide (A₁–A₁₀), unblocked polyalanine decapeptide (P₁ and P₂), and artificial hexopyranose (H₁ and H₂) systems, as well as the corresponding relevant conformational subspaces (two, nine, and seven dimensions, respectively) are described in sections 3.1–3.3. The different parameters are defined in sections 2.4–2.6. Unless noted otherwise, the distribution-alteration function was set to one (eq 22). The indicated parameters are the basis force constant k_{LE} , the EDS smoothing parameter *s*, the conformational region \mathcal{R} involved in the force-constant-reduction procedure (\mathcal{C} : union of all central grid points of the conformational spheres, \mathcal{A} : union of all grid points of all of the conformational lines and spheres), the force-constant reduction factor f_{LE} , the local visiting cutoff γ_{LE} , the global visiting cutoff n_{LE} , the duration of the LE-build-up phase t_{LE} , the value of the force-reduction counter at the end of the build-up phase $I_{\mathcal{R}}^{\max}$, and the duration of the US sampling phase t_{US} . The characteristics of the associated conformational objects (spheres or/and lines) are listed in Table 1. ^b Plain LEUS simulation from ref 37. ^c Using the distribution-alteration function of eq 23. ^d Two sampling trajectories of 50 ns each were concatenated.

dihedral angles corresponding to the atom sequences C_{ac}–N–C_α–C_{CO} and N–C_α–C_{CO}–N_{am}, respectively. Ten different forms of biasing potential were considered, labeled A₁–A₁₀, the corresponding parameters being reported in Tables 1 and 2.

3.2. Polyalanine Decapeptide. A solvated (unblocked) polyalanine decapeptide was used as an application of the B&S-LEUS scheme to the determination of the relative free energies of different types of helices (π , α , and β_{10}), based on a single simulation. This oligopeptide consists of a sequence of 10 alanine residues with free termini (unprotonated amine and protonated carboxylic acid) and was simulated in the presence of $N_{\text{H}_2\text{O}} = 3300$ water molecules (see section 3.4 for the simulation details). The nine-dimensional reduced conformational subspace was defined by the variables $Q_n = \sigma^{-1}\chi_n$ ($n = 1, \dots, 9$), where χ_n is the sum of the two dihedral angles ϕ_{n+1} and ψ_n encompassing the successive peptide bonds¹¹³ (see Figure 1 in ref 37). Note that these angles do not encompass information on ϕ_1 and ψ_{10} .

The three helical states were defined by common values $\chi_n = \chi_{\text{ref}}$ for all n (Figure 2). The values $\sigma\chi_{\text{ref}}$ were set to $\sigma\chi_{\pi} = -130^\circ$ (refs 114–117 suggest $\sigma\chi_{\pi} = -117^\circ, -119^\circ, -127^\circ$, and -131° , respectively), $\sigma\chi_{\alpha} = -105^\circ$ (refs 118, 118+119 and 118+120+121 suggest $\sigma\chi_{\alpha} = -103^\circ, -105^\circ$, and -111° , respectively), and $\sigma\chi_{\beta_{10}} = -75^\circ$ (refs 118+121, 118+119 and 118 suggest $-75^\circ, -78^\circ$, and -89° , respectively). In principle, the above reference values are not sufficient to define ideal regular helices, which are only associated with corresponding χ_n values in the upper right quadrant of the map in Figure 2 (shown in bold). However, the imposition of an additional restraint on the difference between the dihedral angles ϕ_{n+1} and ψ_n (which does not enter into the definition of the reduced conformational subspace) turned out to be unnecessary. In all simulations, these differences never left the range -90° to $+90^\circ$ (for any residue).

Two different forms of biasing potential were considered, labeled P₁ and P₂, the corresponding B&S-LEUS parameters being reported in Tables 1 and 2. The biasing potential P₁ is defined by three spheres centered at the ideal π , α , and β_{10} conformations, respectively, and connected by a unique (thin) line. The biasing potential P₂ is defined by a unique (thick) line passing through (and extending slightly beyond) the three ideal conformations.

3.3. The “Mother” of All D-Hexopyranoses. A solvated artificial D-hexopyranose was used as another application of the B&S-LEUS scheme to the determination of the relative free energies of the 32 D-hexopyranose stereoisomers (All, Alt, Glc, Man, Gul, Ido, Gal, or Tal; Figure 3), anomers (α or β), and chair conformers (⁴C₁ or ¹C₄), based on a single simulation. This artificial compound, termed here the “mother” of all D-hexopyranoses, consists of a D-hexopyranose where the harmonic improper-dihedral potentials normally controlling the stereochemistry of the hydroxyl groups at carbon atoms C₁, C₂, C₃, and C₄ have been changed to a potential form allowing for the interconversion between the two stereoisomers with a finite energy barrier. As a result, a MD simulation of this artificial compound would in principle provide a good reference ensemble for the perturbative evaluation of the relative free energies of the 32 isomers, by reweighting to the corresponding 16 physical ensembles, considering the two chair forms separately. In practice, however, the epimerization, anomerization, and chair-inversion processes are slow and are enhanced here using the B&S-LEUS approach. The artificial D-hexopyranose was simulated in the presence of $N_{\text{H}_2\text{O}} = 1200$ water molecules (see section 3.4 for simulation details). The seven-dimensional reduced mixed (conformational and alchemical) subspace was defined by the variables $Q_n = \sigma^{-1}\xi_n$ ($n = 1, \dots, 4$), where ξ_n is the improper-dihedral angle defining the stereochemistry at carbon atom C_n (C₁–O₅–C₂–O₁, C₂–C₁–O₂–C₃, C₃–C₂–C₄–O₃, and C₄–C₅–C₃–O₄), along with $Q_n = \sigma^{-1}\alpha_{n-3}$ ($n = 4, \dots, 7$), where α_1 , α_2 , and α_3 are the

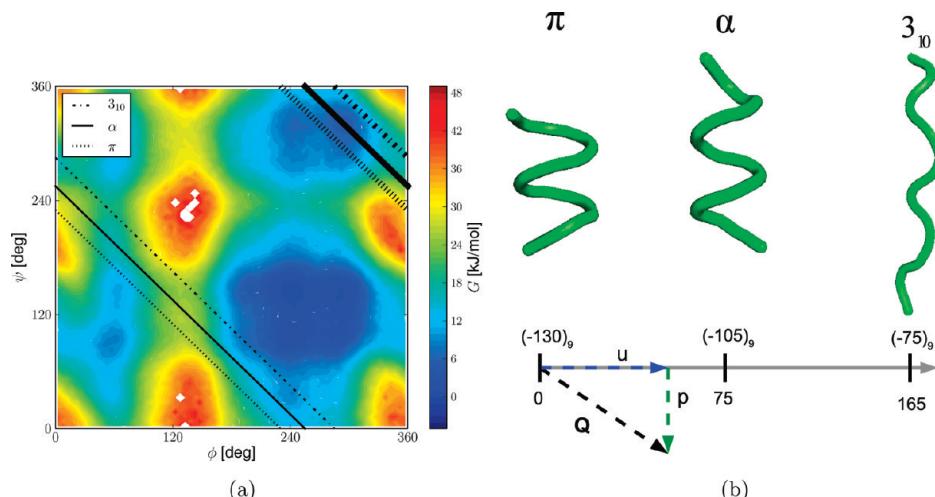


Figure 2. Definition of the reduced conformational subspace and relevant conformational states for the polyalanine decapeptide test system (section 3.2). (a) Definition of ideal conformations for the π , α , and 3_{10} helices, illustrated on a two-dimensional single-linkage basis considering the free-energy surface of a blocked alanine monopeptide ($\sigma Q_n = \chi_n = \phi_{n+1} + \psi_n$ for $n = 1, \dots, 9$; the line segments defining appropriate values for regular helices, upper right quadrant, are shown in bold). Note that the map is drawn considering a $[0^\circ, 360^\circ]$ dihedral-angle range rather than the more usual $[-180^\circ, 180^\circ]$ range. (b) Representative structures (backbone trace) for the ideal conformations of the three types of helices, and illustration of the longitudinal (u) and transverse (p) distances corresponding to a point Q in the nine-dimensional conformational subspace relative to line \mathcal{L}_1 of biasing potential P_1 (Table 1). The representative points for the ideal helical conformations are $(-130)_9$, $(-105)_9$, and $(-75)_9$ for the π , α , and 3_{10} helices, respectively, where $(Q)_9$ represents a nine-dimensional vector with identical components Q . The corresponding longitudinal distances along line \mathcal{L}_1 are 0, 75 ($(9 \cdot 25^2)^{1/2}$), and 165 ($(9 \cdot 55^2)^{1/2}$), respectively.

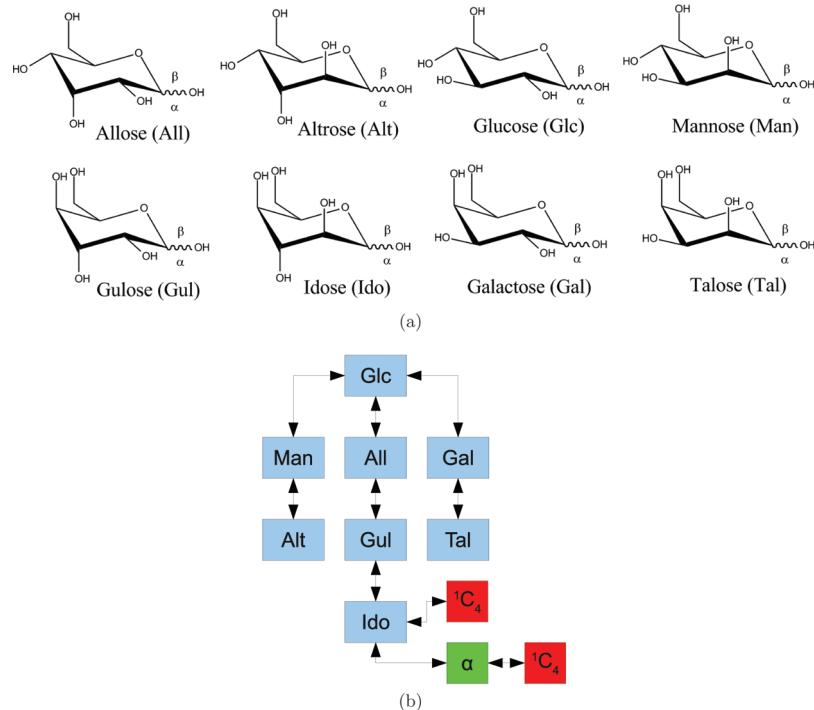


Figure 3. Structures of the hexopyranoses of the D-series and illustration of the maximum-spanning tree employed for the artificial hexopyranose system in protocol H₁ (section 3.3). (a) Structures and naming of the 8 D-hexopyranoses, represented in a 4C_1 chair conformation. (b) Illustration of the maximum-spanning tree employed in protocol H₁. In this tree, the 4C_1 - β -hexopyranoses are connected by 7 lines (drawn between the blue boxes). Each 4C_1 - β -hexopyranose is further connected to its α -anomer leading to 8 additional lines (schematized by the connection to the green box), and each of the 4C_1 - α - and 4C_1 - β -hexapyranose conformers is connected to the corresponding 1C_4 conformer leading to 16 additional lines (schematized by the connection to the red boxes).

three out-of-plane dihedral angles used to define the ring conformation according to Pickett and Strauss¹²² (values of the improper-dihedral angles $C_4-O_5-C_2-C_1$, $O_5-C_2-C_4-C_3$, and $C_2-C_4-O_5-C_5$, respectively, decreased by 180°).

The 32 states corresponding to the different stereoisomers, anomers, and chair conformers were defined by all possible combinations of the ideal values $\sigma Q_n = \pm 35^\circ$ ($n = 1, \dots, 7$) with the additional constraint $\sigma Q_5 = \sigma Q_6 = \sigma Q_7$ (ideal chair

Table 3. Representation of the 32 D-Hexopyranose Stereoisomers (Figure 3), Anomers (α or β), and Chair Conformers (4C_1 or 1C_4) in Terms of Integer Codes i and Corresponding Bit Strings b_n^a

i	b_n	isomer	i	b_n	isomer
0	00000	${}^4C_1\text{-}\alpha\text{-idose}$	16	10000	${}^4C_1\text{-}\beta\text{-idose}$
1	00001	${}^1C_4\text{-}\alpha\text{-idose}$	17	10001	${}^1C_4\text{-}\beta\text{-idose}$
2	00010	${}^4C_1\text{-}\alpha\text{-altrose}$	18	10010	${}^4C_1\text{-}\beta\text{-altrose}$
3	00011	${}^1C_4\text{-}\alpha\text{-altrose}$	19	10011	${}^1C_4\text{-}\beta\text{-altrose}$
4	00100	${}^4C_1\text{-}\alpha\text{-talose}$	20	10100	${}^4C_1\text{-}\beta\text{-talose}$
5	00101	${}^1C_4\text{-}\alpha\text{-talose}$	21	10101	${}^1C_4\text{-}\beta\text{-talose}$
6	00110	${}^4C_1\text{-}\alpha\text{-mannose}$	22	10110	${}^4C_1\text{-}\beta\text{-mannose}$
7	00111	${}^1C_4\text{-}\alpha\text{-mannose}$	23	10111	${}^1C_4\text{-}\beta\text{-mannose}$
8	01000	${}^4C_1\text{-}\alpha\text{-gulose}$	24	11000	${}^4C_1\text{-}\beta\text{-gulose}$
9	01001	${}^1C_4\text{-}\alpha\text{-gulose}$	25	11001	${}^1C_4\text{-}\beta\text{-gulose}$
10	01010	${}^4C_1\text{-}\alpha\text{-allose}$	26	11010	${}^4C_1\text{-}\beta\text{-allose}$
11	01011	${}^1C_4\text{-}\alpha\text{-allose}$	27	11011	${}^1C_4\text{-}\beta\text{-allose}$
12	01100	${}^4C_1\text{-}\alpha\text{-galactose}$	28	11100	${}^4C_1\text{-}\beta\text{-galactose}$
13	01101	${}^1C_4\text{-}\alpha\text{-galactose}$	29	11101	${}^1C_4\text{-}\beta\text{-galactose}$
14	01110	${}^4C_1\text{-}\alpha\text{-glucose}$	30	11110	${}^4C_1\text{-}\beta\text{-glucose}$
15	01111	${}^1C_4\text{-}\alpha\text{-glucose}$	31	11111	${}^1C_4\text{-}\beta\text{-glucose}$

^a The coding principle is described in section 3.3.

conformations; 4C_1 , -35° ; 1C_4 , $+35^\circ$). For simplicity, these conformations can be encoded into an integer index i ($i = 1, \dots, 32$) defined by a string of five bits b_n ($n = 1, \dots, 5$, ordered from the highest-weight to the lowest-weight bit), so that $Q_n = 35(2b_n - 1)$ for $n = 1, \dots, 4$ and $Q_5 = Q_6 = Q_7 = 35(2b_5 - 1)$. Thus, for instance, the ideal conformation $i = 17$ (bit string 10001) corresponds to $\xi_1 = +35^\circ$, $\xi_2 = \xi_3 = \xi_4 = -35^\circ$, and $\alpha_1 = \alpha_2 = \alpha_3 = +35^\circ$, that is, to ${}^1C_4\text{-}\beta\text{-Ido}$. For the ease of reference, the 32 correspondences between integer indices i and hexopyranose isomers are provided in Table 3.

Two different forms of biasing potential were considered, labeled H₁ and H₂, the corresponding parameters being reported in Tables 1 and 2. Both biasing potentials rely on 32 conformational spheres centered at the corresponding ideal values. In protocol H₁, the 32 spheres are connected by a maximum-spanning tree of 31 lines, illustrated in Figure 3b. Note that this tree is only one possible choice among many others. In protocol H₂ the 32 spheres are connected by a redundant tree of 80 lines motivated by the use of the Manhattan (or Taxicab) metric,¹²³ that is, each state is connected with the five states differing from it by the value of a single bit in terms of binary representation.

3.4. Simulation Details. All MD simulations were carried out using a modified version of the GROMOS05 program¹²⁴ together with the 53A6 force field³⁰ (peptides) and the 56A_{CARBO} force field⁹⁷ (hexopyranose). The systems considered consisted of 1 solute molecule and N_{H_2O} simple point charges¹²⁵ (SPC) water molecules, where N_{H_2O} was 1300, 3300, and 1200 for the blocked alanine monopeptide, polyalanine decapeptide, and artificial hexopyranose, respectively.

Newton's equations of motion were integrated using the leapfrog algorithm^{126,127} with a 2 fs time step. The SHAKE procedure¹²⁸ was applied to constrain all bond lengths as well as the full rigidity of the solvent molecules with a relative geometric tolerance of 10^{-4} . The simulations were performed under periodic boundary conditions based on a cubic computational box and in the isothermal–isobaric (NPT) ensemble at 298.15 K and 1 bar. The temperature

was maintained by weak coupling of the solute and solvent degrees of freedom (jointly) to a heat bath¹²⁹ with a relaxation time of 0.1 ps. The pressure was maintained by weak coupling of the atomic coordinates and box dimensions to a pressure bath¹²⁹ (isotropic coordinate scaling, group-based virial) with a relaxation time of 0.5 ps and an isothermal compressibility of 0.4575×10^{-3} (kJ mol⁻¹ nm⁻³)⁻¹ as appropriate for water.⁸⁸ The center of mass motion was removed every time step. Nonbonded interactions were handled using a twin-range cutoff scheme,^{2,130} with short- and long-range cutoff distances of 0.8 and 1.4 nm, respectively, and update frequency of 5 time steps for the short-range pairlist and intermediate-range interactions. The mean effect of the omitted electrostatic interactions beyond the long-range cutoff distance was approximately reintroduced using a reaction-field correction,¹³¹ based on a relative dielectric permittivity of 61 as appropriate for the SPC water model.⁴⁰ Configurations (along with U_{bias} in the sampling phase) were written to file every 0.02 ps (peptide) or 0.2 ps (hexopyranose) for subsequent analysis. The systems were preequilibrated using 500 ps MD simulation.

As described in section 3.3, the simulation of the “mother” of all D-hexopyranoses required a modification of the functional form for the terms controlling the stereochemistry at carbon atoms C₁, C₂, C₃, and C₄. In the GROMOS force field,^{88,89,124} the chirality of a center is determined by an improper-dihedral angle energy term of the form

$$V_{\xi}(\xi) = \frac{1}{2}k_{\xi}(\xi - \xi_0)^2 \quad (28)$$

where two opposite values of ξ_0 characterize the two stereoisomers. In the simulations of the “mother” of all D-hexopyranoses, this interaction term was modified on the basis of the EDS principle⁷⁸ to

$$V'_{\xi}(\xi) = -\frac{1}{\beta s'} \ln \left[\exp[-\beta s' \frac{1}{2}k_{\xi}(\xi - |\xi_0|)^2] + \exp[-\beta s' \frac{1}{2}k_{\xi}(\xi + |\xi_0|)^2] \right] \quad (29)$$

where s' is a barrier-smoothing parameter. These choices, along with the setting $\beta s' = 0.1$, led to smooth anomeration and epimerization transitions, with barriers on the order of 50–100 kJ mol⁻¹. The thermodynamic quantities appropriate for the physical ensemble considering one specific stereoisomer are then calculated on the basis of the configurations generated during the sampling phase of a B&S-LEUS simulation, via a reweighting procedure that now depends on the sum of the biasing potential U_{bias} and of the difference $V'_{\xi} - V_{\xi}$ for this specific stereoisomer.^{15,97} Note that this specific application of the EDS principle is functionally entirely distinct from its use in the unification of the single-object biasing potentials described in section 2.4.

4. Results

4.1. Blocked Alanine Monopeptide. The results of the 10 simulations (A₁–A₁₀; Tables 1 and 2) involving a (blocked) alanine monopeptide (section 3.1) are displayed in Figure 4 in the form of Ramachandran free-energy maps.

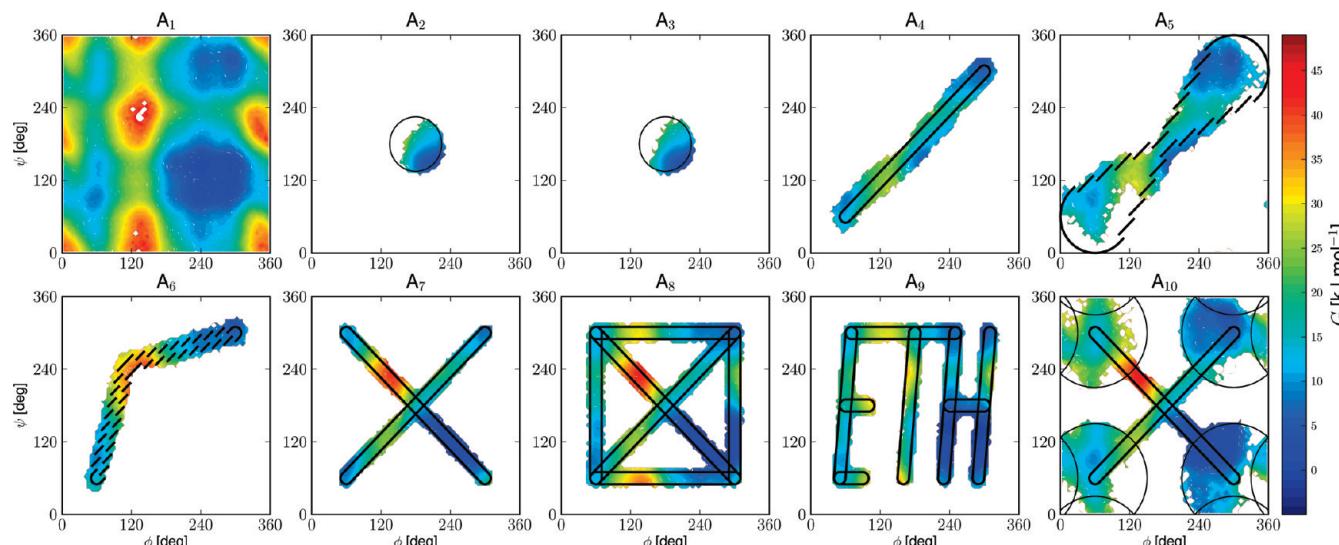


Figure 4. Free-energy maps obtained from the different B&S-LEUS simulations (along with one plain LEUS simulation) of a blocked alanine monopeptide in water (section 3.1). The relevant conformational subspace is defined by $\sigma(Q_1, Q_2) = (\phi, \psi)$, and the maps represent the calculated free energy. The characteristics of the conformational objects (spheres or/and lines) defining the B&S-LEUS biasing potentials and the parameters of the associated protocol are provided in Tables 1 and 2, respectively, for the different simulations (A_1 – A_{10}). The maps are calculated by reweighting of the biased probability distribution using a grid spacing of 5° , and anchored by fitting of the covered area onto the free-energy map for simulation A_1 , itself anchored to zero at the location of its global minimum. Note that the maps are drawn considering $[0^\circ, 360^\circ]$ dihedral-angle ranges rather than the more usual $[-180^\circ, 180^\circ]$ ranges.

Simulations A_2 – A_{10} are B&S-LEUS simulations, while simulation A_1 is a reference plain LEUS simulation from ref 37 showing the entire free-energy map.

The results illustrate, in the simple context of a reduced conformational subspace of low dimensionality, the versatility of the B&S-LEUS scheme in the design of different patterns of active subspaces (here, active surfaces) in terms of spheres (here, disks) and lines. They also underline that the B&S-LEUS sampling is nearly homogeneous radially but not necessarily directionally within the spheres (e.g., A_2 , A_3 , and A_{10}), as well as longitudinally but not necessarily transversally within the lines (e.g., A_5). Note that the latter inhomogeneities will be accentuated when considering problems of higher dimensionalities. The most remarkable feature in Figure 4 is that all maps are closely resemblant to each other within the active subspace. The achievement of the B&S-LEUS procedure is to permit a flexible (problem-adapted) definition of this subspace, guarantee an appropriate sampling within it, and exclude its exterior from the sampling.

Simulation A_{10} can be considered as a two-dimensional prototype of what the B&S-LEUS approach is meant to achieve in higher-dimensionality problems, namely the connection of a number of conformational states (here, four) by means of lines (here, two) ensuring a sufficient number of transitions between these states. Further details concerning this simulation are shown in Figure 5.

The number of visits N_v to different conformational points (squares of edge 5°) observed during the sampling phase of this simulation is displayed in Figure 5a (on a logarithmic scale). Projections along the line directions or along the sphere radii (data not shown) evidence, as expected, essentially homogeneous distributions of N_v , although the two-

dimensional distribution itself shows a significant extent of inhomogeneity within the active subspace, especially inside the spheres. The biasing potential $\mathcal{U}_{\text{bias}}$ obtained at the end of the LE build-up phase (and used during the sampling phase) is shown in Figure 5b. As expected, due to the action of the restraining potentials, any sampling outside the active subspace is essentially precluded. Within the active subspace, the memory-based component compensates for the physical bias in the free-energy surface (i.e., it is higher in regions where the free energy is lower). Although one might consider using the negative of $\mathcal{U}_{\text{bias}}$ as an approximate estimate for the free energy G (eq 25), the biased map $G + \mathcal{U}_{\text{bias}}$ (Figure 5d) shows that the corresponding error can be very large (up to about ± 10 kJ mol $^{-1}$ for specific conformational points).

Simulations A_5 and A_6 illustrate two variants of lines, the line with longitudinally dependent width and the displaced line (section 2.3). Simulations A_2 and A_3 illustrate two variants of biasing potentials within a sphere, either radially homogeneous or with a radial distribution proportional to the Jacobian factor (here, to the distance to the sphere center) using a nonunit distribution-alteration function (eq 23 instead of eq 22). Further details concerning the latter two simulations are shown in Figure 6.

The numbers of visits N_v per conformational point (squares of edge 5°) for the two simulations are shown in Figure 6a and b (on a logarithmic scale), and the corresponding radial projections are shown in Figure 6c and d. As expected, the nonunit distribution-alteration function biases the sampling of the sphere toward its periphery and changes the homogeneous radial distribution to an approximately linear one. However, in view of the low dimensionality and long

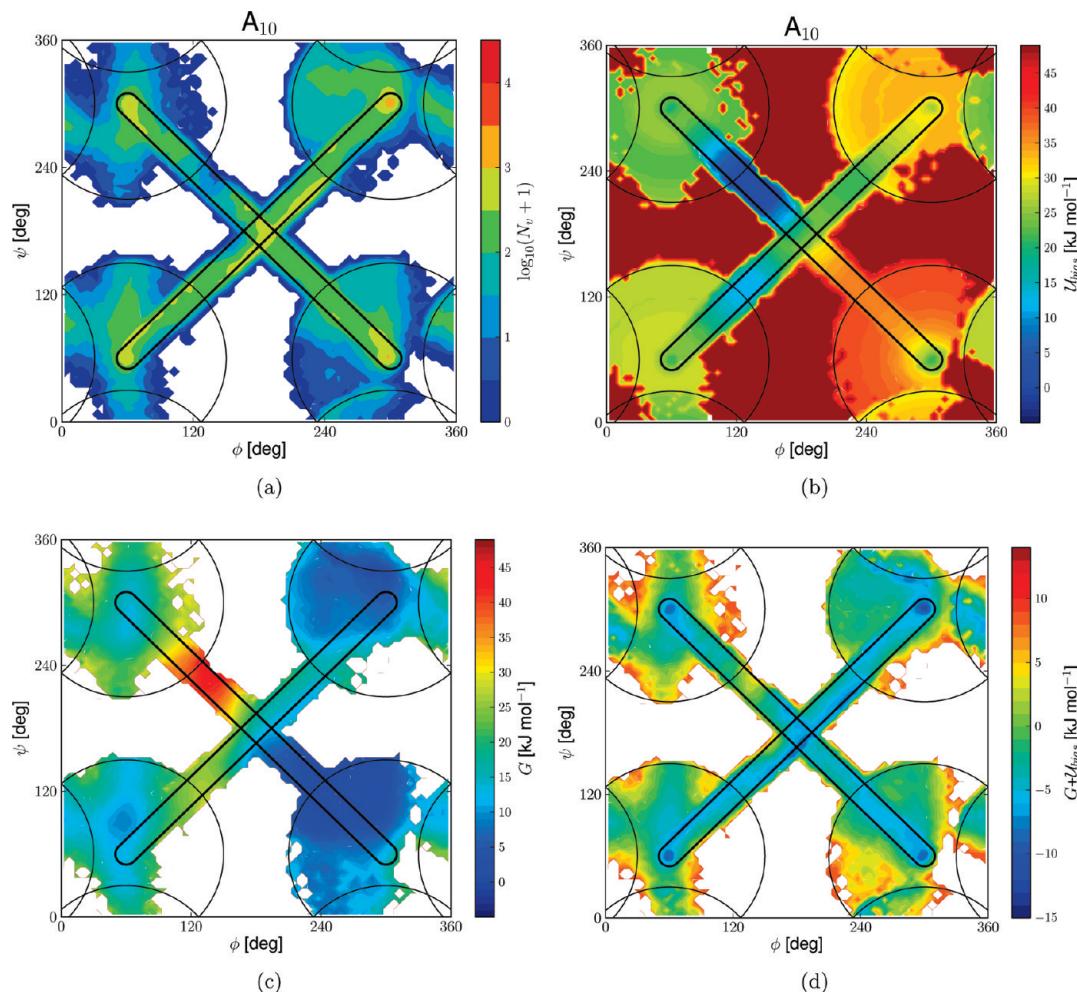


Figure 5. Illustrative details of the B&S-LEUS simulation of a blocked alanine monopeptide in water using the biasing potential A_{10} , see legend of Figure 4. (a) Number of visits N_v (displayed as $\log_{10}(N_v + 1)$) to a given conformational point during the sampling phase. (b) Optimized biasing potential U_{bias} (all points that were not sampled, or where $U_{\text{bias}} \geq 50 \text{ kJ mol}^{-1}$, are shown in red). (c) Calculated free-energy map G . Biased free-energy map $G + U_{\text{bias}}$. The maps and numbers of visits are calculated using a grid spacing of 5° . The map of panel (c) is identical to that displayed in Figure 4. The sum of N_v for panel (a) is equal to 250 000 (5 ns sampling phase, coordinates written every 0.02 ps).

sampling time, this change does not significantly alter the final free-energy map (Figure 4).

4.2. Polyalanine Decapeptide. The results of the two B&S-LEUS simulations (P_1 and P_2 ; Tables 1 and 2) involving an (unblocked) polyalanine decapeptide (section 3.2) are displayed in Figures 7 and 8. The biasing potential P_1 is defined by three spheres centered at the ideal π , α , and 3_{10} conformations, respectively, and connected by a unique (thin) line. The biasing potential P_2 is defined by a unique (thick) line passing through (and extending slightly beyond) the three ideal conformations.

The time evolutions and distributions of the longitudinal distance u and transverse distance p (with reference to line \mathcal{L}_1 of protocol P_1) are displayed in Figure 7a and b for simulations P_1 and P_2 , respectively.

The longitudinal distance u accounts for one degree of freedom of the reduced conformational subspace. In both simulations, it varies over a range (-26 to 262 in P_1 , -122 to 226 in P_2) appropriate to cover the three types of helices (π , 0; α , 75; 3_{10} , 165). However, the number of “sweeps” across this interval remains limited, and the corresponding (biased) distributions $P(u)$ are not homogeneous, suggesting

that a sampling phase longer than 2×50 ns (and possibly also a longer build-up phase) might be desirable. Note that a homogeneous distribution is expected in simulation P_2 upon full convergence, but not necessarily in simulation P_1 due to the presence of the three spheres overlapping with the line. In both simulations, the regions corresponding to the π and α helices are found to be significantly more sampled than that corresponding to the 3_{10} helix (the more limited sampling of the 3_{10} helix region is most pronounced in simulation P_2).

The transverse distance p accounts for the remaining eight degrees of freedom of the reduced conformational subspace. This distance displays much lower fluctuations along the trajectories, with relatively narrow (biased) distributions $P(p)$ centered at about 30 and 100 for simulations P_1 and P_2 , respectively. A lower value for P_1 is expected due to the inclusion of the spheres, which enforce a radial biasing relative to three reference points characterized by $p = 0$. The average values of 30 and 100 correspond to root-mean-square deviations of about 10° and 35° in terms of the individual dihedral-angle differences χ_n .

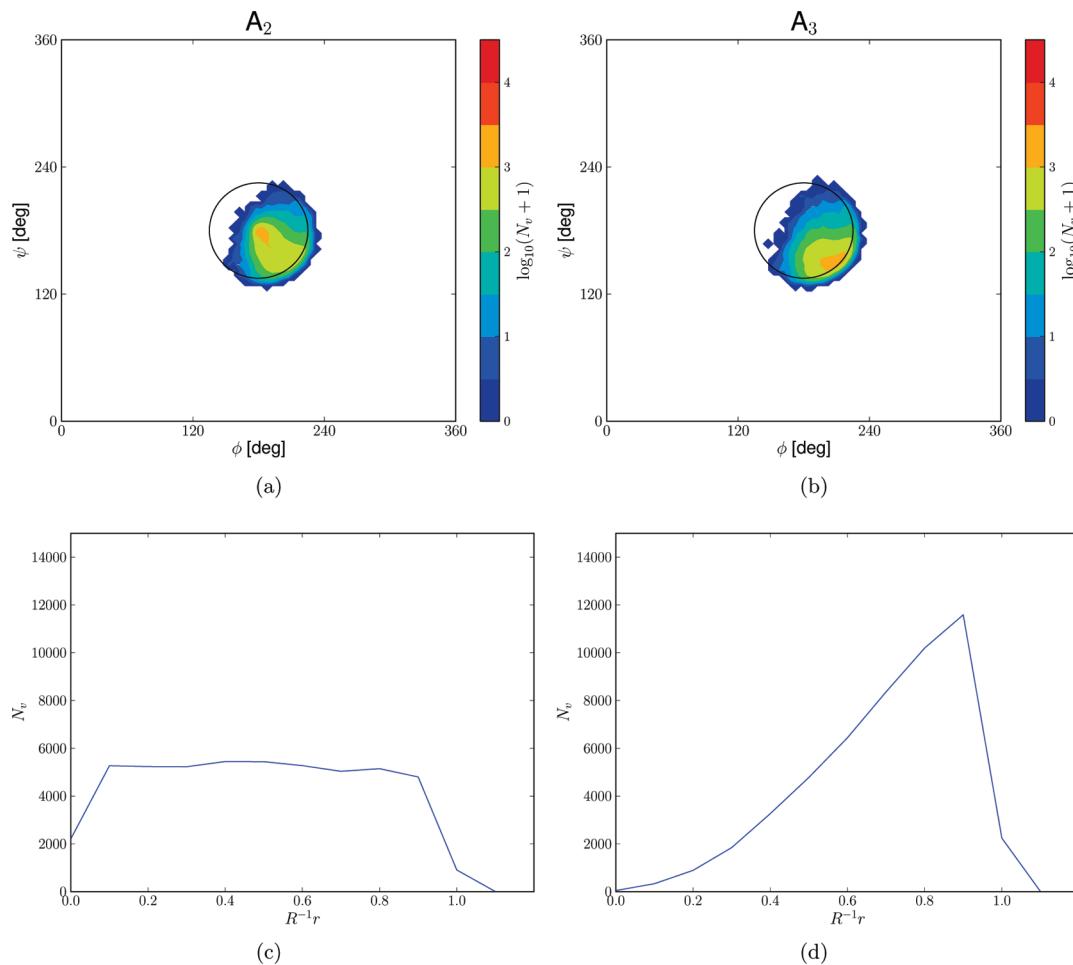


Figure 6. Illustrative details of the B&S-LEUS simulations of a blocked alanine monopeptide in water using the biasing potentials A_2 and A_3 , see legend of Figure 4. (a,b) Number of visits N_v (displayed as $\log_{10}(N_v + 1)$) to a given conformational point during the sampling phases of simulations A_2 (a) and A_3 (b). (c,d) Number of visits N_v as a function of the ratio of the distance from the sphere center to the sphere radius during the sampling phases of simulations A_2 (c) and A_3 (d). The numbers of visits are calculated using a grid spacing of 5° . The sum of N_v for panels (a) and (b) is equal to 250 000 (5 ns sampling phase, coordinates written every 0.02 ps).

The time evolutions of the nine dihedral-angle differences χ_n defining the reduced conformational subspace ($\sigma^{-1}Q_n = \chi_n$) and of the local DSSP secondary-structure assignment¹³² at the eight central residues m ($m = 2, \dots, 9$, where the residue m corresponds to ϕ_m and ψ_m and is thus comprised between χ_{m-1} and χ_m) are displayed in Figure 7c and d for simulations P_1 and P_2 , respectively.

The individual dihedral-angle differences χ_n evidence large and highly correlated fluctuations. This correlation is expected, considering that the biasing potential has been designed to drive the sampling along a “diagonal” line in the nine-dimensional reduced conformational subspace. The magnitude of the fluctuations is higher and the extent of correlation lower for the terminal χ_1 and χ_9 values. The probable reasons for this effect are: (i) a less pronounced dependence of the free energy on the orientation of the (more flexible) terminal residues; and (ii) a correlation of these dihedral-angle differences with the dihedral angles ϕ_1 and ψ_{10} that are not included in the biasing scheme.

The local DSSP secondary-structure assignment reveals that in a significant fraction of the configurations where the longitudinal distance u as well as the individual χ_n values are in the ideal π or α helical regions, a π or α helix is

actually formed, despite the nonzero transverse distance p . In contrast, only a very limited fraction of the configurations where u and the χ_n values are in the ideal 3_{10} region are actually recognized as 3_{10} helices, due to the transverse distance p . In other words, the nonzero transverse distance p predominantly accounts for a conformational variability of the simulated configurations in the neighborhood of ideal π or α structures, while it predominantly represents a conformational discrepancy of the simulated configurations with respect to an ideal 3_{10} helix structure.

The number of visits N_v to a given conformational point (rectangles of edges $\Delta u = 7$ and $\Delta p = 2.4$), as well as the free-energy maps in the subspace of the longitudinal and transverse (or longitudinal only) distances u and p , are displayed in Figure 8a,c,e,g and b,d,h,f for simulations P_1 and P_2 , respectively. These two-dimensional (or one-dimensional) maps provide simplified variants (projections) of the corresponding nine-dimensional maps in the full reduced conformational subspace, which obviously cannot be easily visualized.

For simulation P_1 , the number of visits (Figure 8a) reveals an extensive coverage of the two-dimensional subspace (within the bounds -50 to $+250$ for u and 0 to $+125$ for p).

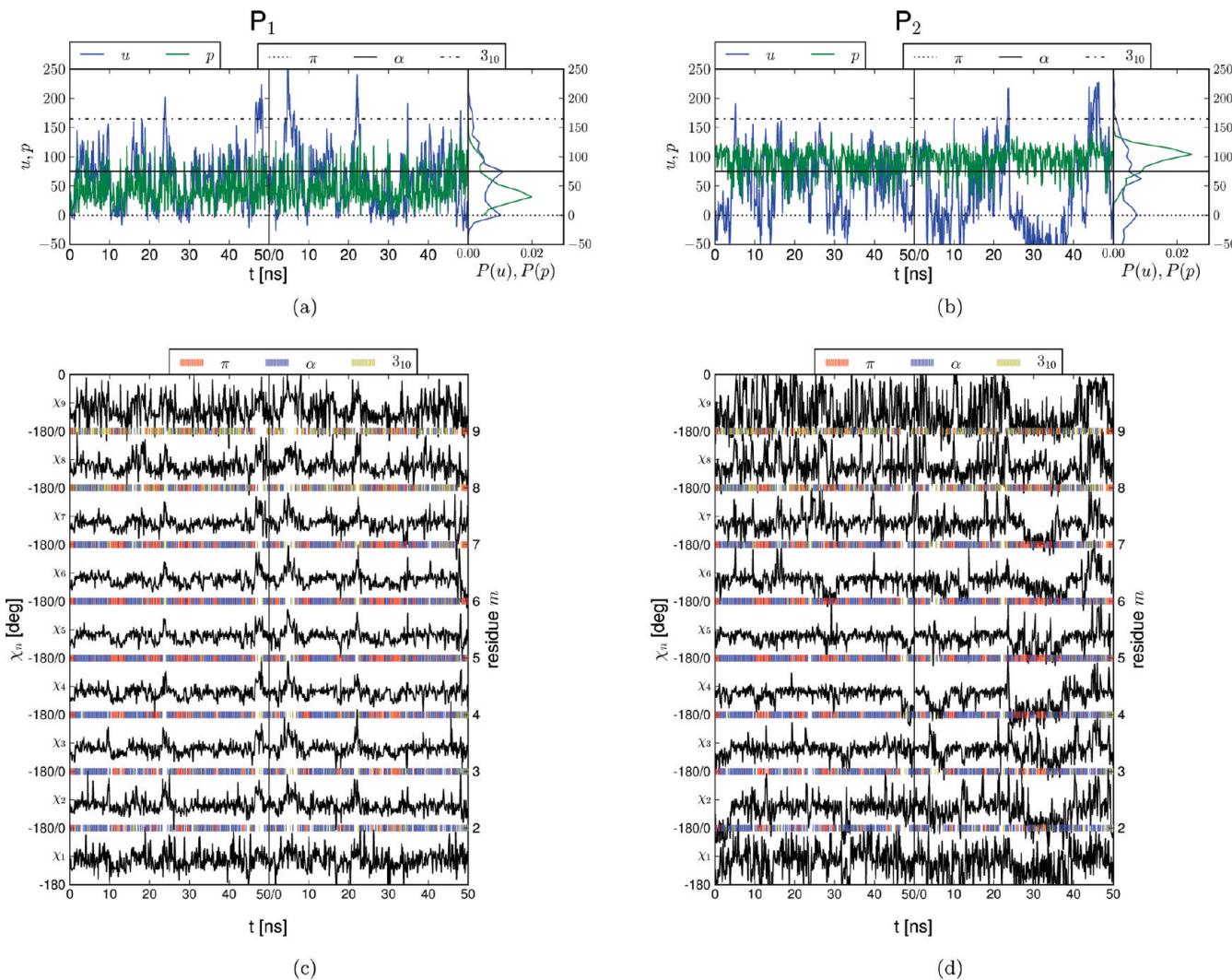


Figure 7. Time series of the longitudinal and transverse distances along a line connecting the relevant helical states, as well as of the conformational coordinates and local helical patterns, for the two B&S-LEUS simulations of a polyalanine decapeptide in water (section 3.2). The relevant conformational subspace is defined by $\sigma Q_n = \chi_n = \phi_{n+1} + \psi_n$ ($n = 1, \dots, 9$), see Figure 2. (a,b) Time series (and corresponding normalized distributions in the biased ensemble) of the longitudinal (u) and transverse (p) distances, referring to the line \mathcal{L}_1 of protocol P_1 (Table 1 and Figure 2), for the sampling phases of simulations P_1 (a) and P_2 (b). (c,d) Time series of the conformational coordinates χ_n , and of the local helical pattern (assigned according to DSSP¹³²), for the sampling phases of simulations P_1 (c) and P_2 (d). Note that the results correspond to the concatenated sampling phases of two independent 50 ns simulations.

As discussed above (Figure 7a), this coverage is not entirely homogeneous along u and shows a predominance of visited conformations around $p = 30$. For simulation P_2 (Figure 8b), however, a significant portion of the subspace in the neighborhood of the ideal 3_{10} helix conformation ($u = 165$, $p = 0$) is not visited.

The reason for this difference becomes obvious when comparing the corresponding free-energy maps (Figure 8c and d). As is clearly visible from the results of simulation P_1 , the close neighborhood of the 3_{10} helix corresponds to a region of very high free energy. In simulation P_1 , the sampling of this neighborhood is enforced by the inclusion of a corresponding conformational sphere. In simulation P_2 , the sampling of this region is not enforced, because the longitudinal sampling homogeneity within the line is compatible with high p values when u is close to 165. Besides the above difference in the extent of coverage, the maps issued from the two independent simulations are remarkably

similar inside the regions they both cover. This suggests a sufficient level of convergence in the two simulations, despite the residual inhomogeneity in $P(u)$ (see above).

The configurations recognized by the DSSP algorithm as π , α , and 3_{10} helices (for 7 out of 8 residues at least) are shown superimposed on the maps in Figure 8e and f for simulations P_1 and P_2 , respectively. A higher number of configurations are recognized as regular helices for simulation P_1 as compared to simulation P_2 , due to the lower average value of p (Figure 7a vs b), a consequence of the use of the three spheres in P_1 . Note again that configurations recognized as π or α helix can still present relatively high p values (variability relative to corresponding ideal conformations), while very few configurations are recognized as 3_{10} helix (almost none for simulation P_2).

Finally, the free energy and average value \bar{p} of p are shown as a function of u in Figure 8g and h for simulations P_1 and P_2 , respectively. Note that \bar{p} is not the straight average of p

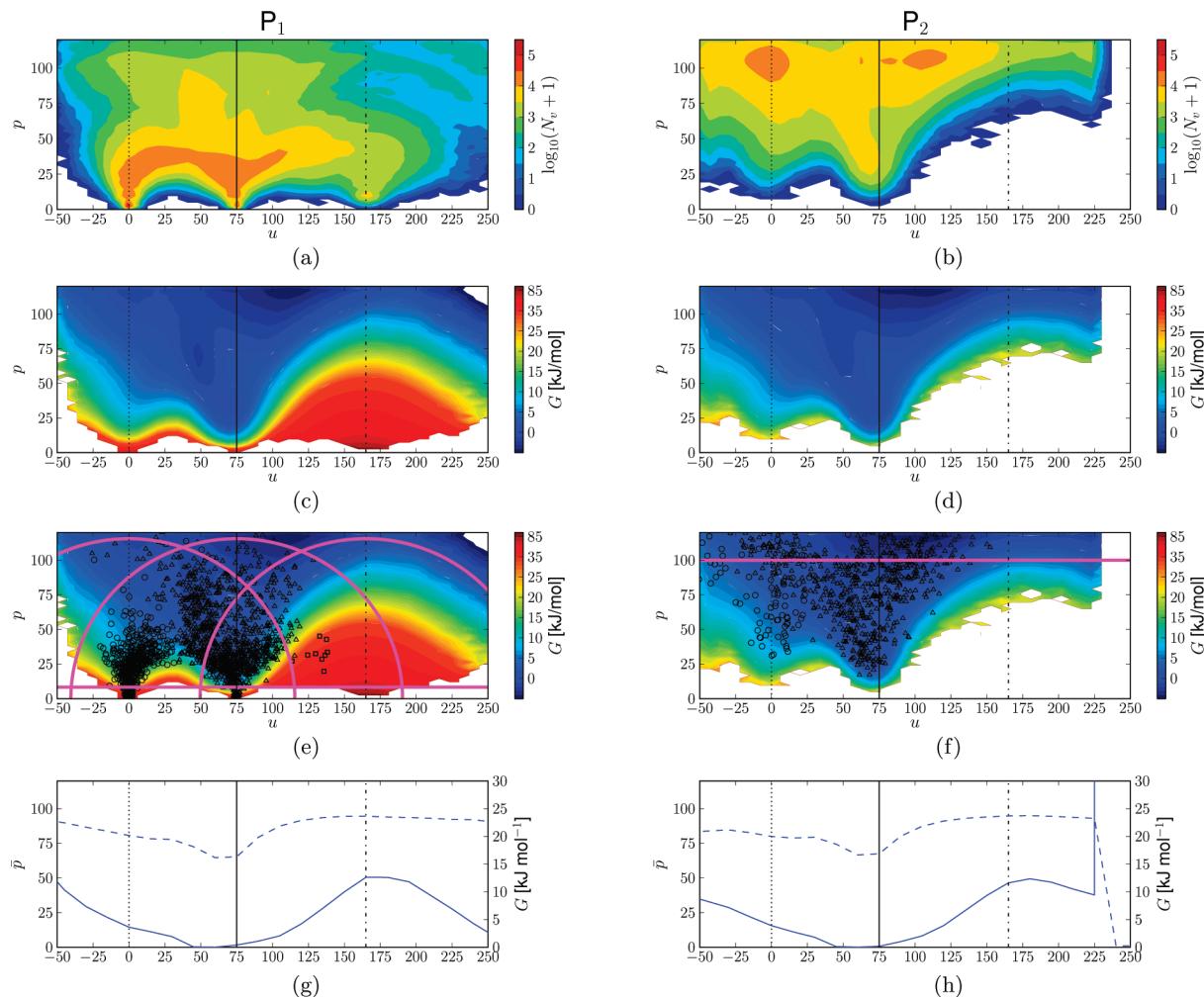


Figure 8. Number of visits N_v and free-energy maps (longitudinal-transverse and purely longitudinal projections) corresponding to a line connecting the relevant helical states for the two B&S-LEUS simulations of a polyalanine decapeptide in water (section 3.2). The relevant conformational subspace is defined by $\sigma Q_n = \chi_n = \phi_{n+1} + \psi_n$ ($n = 1, \dots, 9$), see Figure 2. The longitudinal (u) and transverse (p) components refer to the line \mathcal{L}_1 of protocol P₁ (Table 1 and Figure 2). (a,b) Number of visits N_v during the sampling phase of the simulations P₁ (a) or P₂ (b), displayed on a logarithmic scale. (c,d,e,f) Free energy $G(u, p)$ in a longitudinal-transverse projection for the simulations P₁ (c,e) or P₂ (d,f). (g,h) Free energy $G(u)$ (solid line) and average transverse distance $\bar{p}(u)$ (dashed line; reweighted to the physical ensemble) in a purely longitudinal projection for the simulations P₁ (g) or P₂ (h). The maps and number of visits are calculated using a grid spacing $\Delta u = 7$ and $\Delta p = 2.4$. The sum of N_v for panels (a) and (b) is equal to 5×10^6 (100 ns sampling phase, coordinates written every 0.02 ps). Panels (c) and (d) are identical to panels (e) and (f), respectively. However, in the latter, the spheres and lines involved in the B&S-LEUS potentials (lines representing their peripheries, based on the corresponding radii and widths) are shown as solid lines, as well as representative points of conformations assigned by DSSP¹³² as π (circles), α (triangles), and 3_{10} (squares) helices (at least seven linkages assigned to the corresponding helical conformation out of eight; Figure 7). Note that the color coding of the free energy is not linear for high free energies.

along the biased simulations (Figure 7a and b), due to reweighting to the physical ensemble. The two curves are closely similar in the two simulations (over the range they jointly sample). They indicate relative free energies of the order of 4, 0, and 12 kJ mol⁻¹ for the π , α , and 3_{10} helical forms of a polyalanine decapeptide in water based on the GROMOS 53A6 force field.³⁰ They also suggest the possibility of a smooth (barrier-free) interconversion pathway along the cooperative coordinate u . Of course, this implies neither that this path is the lowest free-energy path, nor that most interconversion trajectories follow this path. Finally, the average transverse distance \bar{p} is nearly constant along u (about 85, involving root-mean-square single-residue angular deviations of the order of 30°), except close to the α -helix (corresponding values of about 70 and 25°, respectively).

As a final note, the noticeable (but still limited) discrepancies between dihedral-angle and DSSP descriptions for helical structures suggest that it might be interesting to repeat the present simulations based on a conformational subspace defined by hydrogen-bonding rather than backbone dihedral-angle coordinates.

4.3. The “Mother” of All d-Hexopyranoses. The results of the two B&S-LEUS simulations (H₁ and H₂; Tables 1 and 2) involving the artificial d-hexopyranose (section 3.3) are displayed in Figures 9–11. Both biasing potentials rely on 32 conformational spheres centered at the corresponding ideal values. In protocol H₁, the 32 spheres are connected by a maximum-spanning tree of 31 lines, illustrated in Figure 3b. In protocol H₂, the 32 spheres are connected by a redundant tree of 80 lines motivated by the use of the

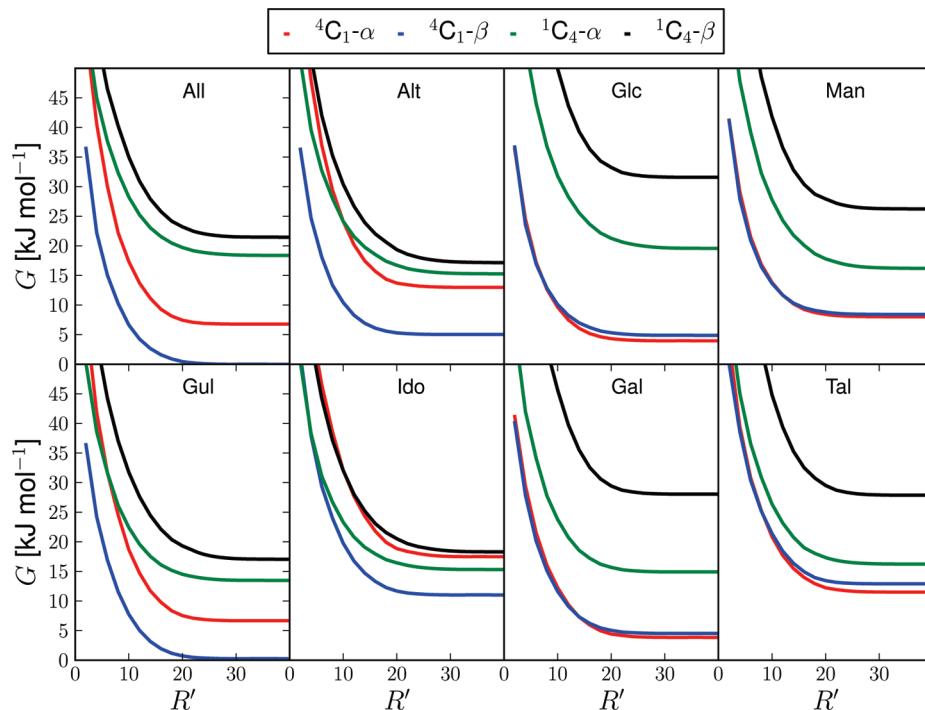


Figure 9. Relative free energies of the 32 different isomers as a function of the sphere radius used for the assignment of the conformations to states in the B&S-LEUS simulations of the “mother” of all D-hexopyranoses in water (section 3.3) using protocol H₂ (Tables 1 and 2). The assignment to states is based on seven-dimensional spheres of radius R' centered at the 32 ideal D-hexopyranose stereoisomers (Figure 3), anomers, and chair conformers. The free energies of the different states are calculated on the basis of the (reweighted) population of a state relative to the total (reweighted) population of the entire reduced conformational space, and offset so as to bring the free energy of ${}^4\text{C}_1\text{-}\beta\text{-All}$ to zero for large R' .

Manhattan (or Taxicab) metric,¹²³ that is, each state is connected with the five states differing from it by the value of a single bit in terms of binary representation.

The relative free energies calculated for the 32 isomers based on simulation H₂ are shown in Figure 9 as a function of the radius R' used to assign sampled conformations to states. The corresponding curves for simulation H₁ are qualitatively similar (data not shown). As discussed in section 2.7, the volumes used for the state assignment need not be identical to those used in the definition of the biasing potential. In the present case, both types of volumes are spheres centered at the 32 ideal conformations, with radii $R = 50$ for the former ones (Table 1) and R' for the latter ones. Note that the free-energy values in this figure are normalized relative to the total (reweighted) population of the reduced conformational subspace (and offset so as to bring the lowest value of ${}^4\text{C}_1\text{-}\beta\text{-All}$ to zero at large R').

For small R' , the free energies of the states diverge to $+\infty$ as the fractional populations they encompass become infinitesimally small. Note that in this limit, the corresponding differences remain in principle finite, but become increasingly inaccurate (and ultimately diverge) as a consequence of finite sampling. For sufficiently large R' , the free energies converge to well-defined values, indicating that a further extension of the sphere radii only adds high free-energy regions, which no longer contribute significantly to the free energy of the states, that is, the corresponding local free-energy basins have been fully encompassed. Note that the ideal conformations may not exactly correspond to the free-energy minima of these local basins. However, this is expected to have a minor

influence on the results as long as R' is chosen sufficiently large (and given that the basins corresponding to different states remain well separated). Based on the results of Figure 9, spheres of radii $R' = 30$ were chosen for all subsequent state assignments.

The results of the simulations of the artificial hexopyranose based on protocols H₁ and H₂ are shown in Figure 10a and b, respectively. Both simulations have visited the entire set of 32 states during their sampling phases. The fraction of intermediate configurations that could not be assigned to any state is about 30%. This number is actually remarkably low, that is, only about one-third of the computational effort has been invested in ensuring a sufficient number of transitions between the states via statistically irrelevant regions. Note that this fraction is not higher for simulation H₂ as compared to simulation H₁, although the latter involves more lines (80 vs 31). On the other hand, simulation H₂ achieved a much higher extent of homogeneity in terms of state populations within the biased ensemble and in terms of the numbers of transitions between these states. As a result, the relative free energies evaluated for the different states based on protocol H₂ are expected to be more precise (as supported by the lower error bars) and more accurate (within the employed force field and simulation methodology).

The convergence and accuracy properties of the two protocols are compared in Figure 11, in terms of the chair-inversion free-energy change $\Delta G_{\text{C}_1 \rightarrow \text{C}_4}^4$. Figure 11a shows the root-mean-square deviation (rmsd) between estimates based on the sampling phase up to time t and corresponding (well-converged) estimates from the 16 independent LEUS

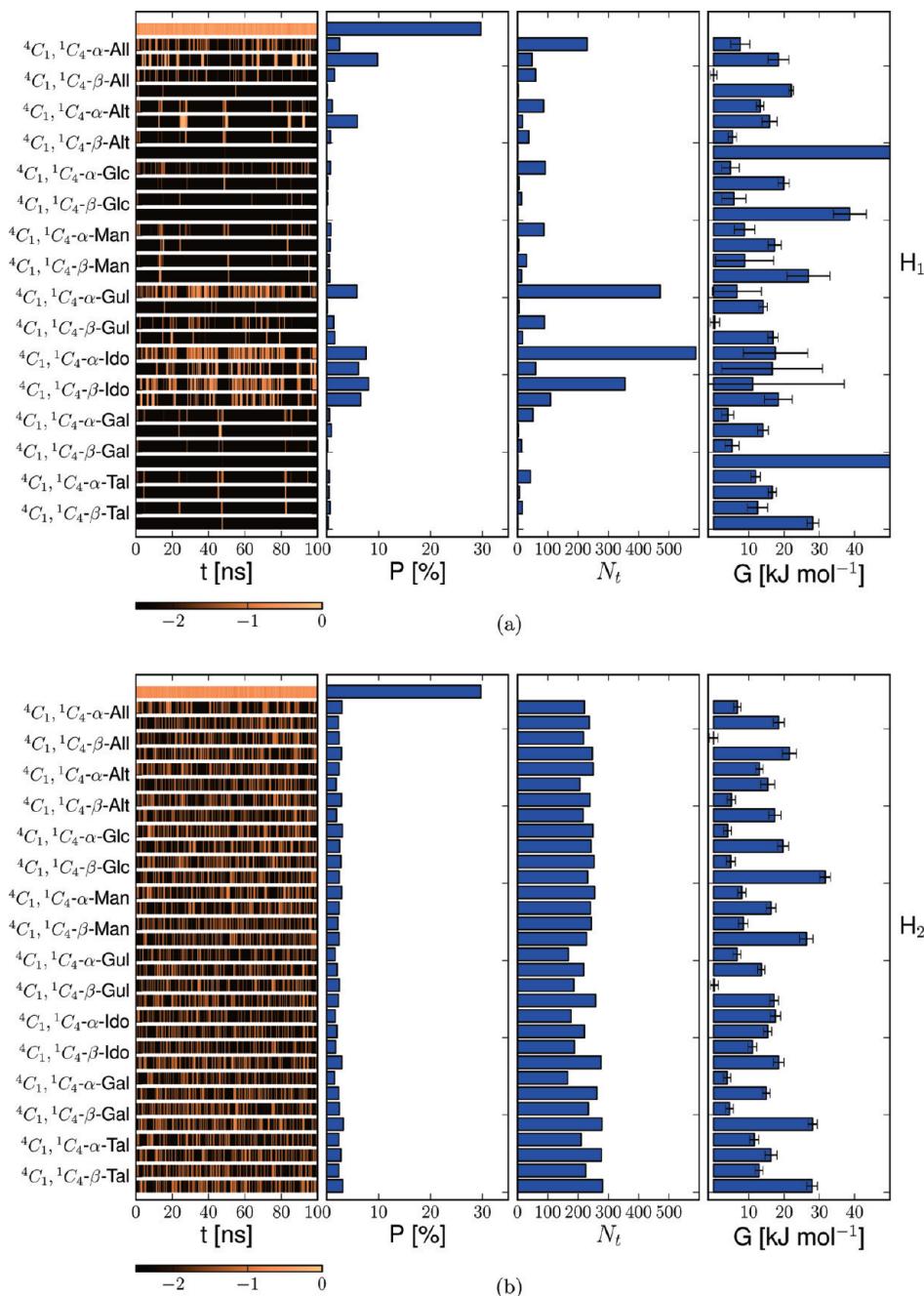


Figure 10. Results of the B&S-LEUS simulations of the “mother” of all D-hexopyranoses in water (section 3.3). (a) Results of the simulations H₁. (b) Results of the simulations H₂. The following quantities are displayed (from left to right), with reference to the 32 ideal D-hexopyranose stereoisomers (Figure 3), anomers, and chair conformers: the fraction (on a decimal logarithmic scale) of successive 40 ps intervals along the sampling phase spent visiting a given state, the relative probability P of a given state in the biased ensemble during the sampling phase, the number of transitions N_t observed to a given state from any other state during the sampling phase, and the relative free energies G of the states after reweighting. The visit fractions and free energies rely on a state assignment using seven-dimensional spheres of radius $R' = 30$ centered at the corresponding ideal conformations.

simulations reported in ref 97. Because the free energy of states that have not yet (or only poorly) been sampled at time t is formally infinite, the rmsd calculation only includes the N_p changes for which the number of visits to both conformers up to time t is at least 100. The detailed comparison in terms of values corresponding to the entire sampling phase is shown in Figure 11b.

Protocol H₁ shows a somewhat erratic convergence to a final rmsd value of 1.0 kJ mol⁻¹, with three identified

outliers (β -Alt, β -Glc, and β -Gal; absolute deviations of 43.7, 6.4, and 30.9 kJ mol⁻¹, respectively; not included in the rmsd), corresponding to the states that have been poorly visited during the biased simulation (Figure 10a). In contrast, protocol H₂ shows a smooth convergence to a final rmsd value of 0.7 kJ mol⁻¹ (all states included). Here, the deviation is homogeneously spread across the 16 free-energy changes, and within the statistical error bars in nearly all cases.

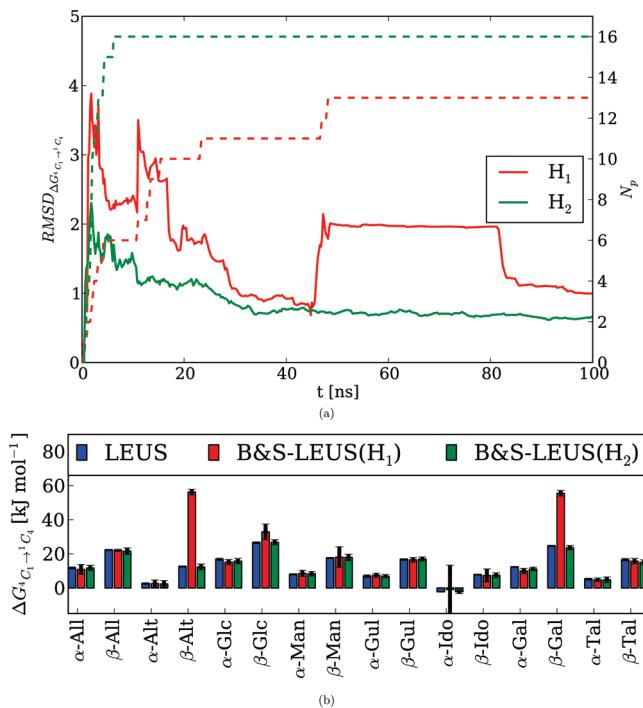


Figure 11. Comparison of the relative free energies of D-hexopyranose chair conformers in water calculated using a single B&S-LEUS simulations of the “mother” of all D-hexopyranoses (section 3.3) or using 16 plain LEUS simulations. The B&S-LEUS results correspond to simulations H₁ and H₂ (Tables 1 and 2). The plain LEUS results are from ref 97. (a) Root-mean-square deviation (rmsd; solid lines) between the free-energy difference $\Delta G^{4C_1 \rightarrow ^1C_4}$ as predicted by the B&S-LEUS simulation (protocols H₁ or H₂) considering a sampling time t and the corresponding values obtained using the 16 plain LEUS simulations. Only pairs of states that have been both visited more than 100 times are included in the rmsd calculation. The corresponding number of pairs N_p is also shown (dashed lines). (b) Free-energy difference $\Delta G^{4C_1 \rightarrow ^1C_4}$ between the ⁴C₁ and ¹C₄ ring conformers evaluated using the 16 plain LEUS simulations or the two B&S LEUS simulations.

The above results suggest that the use of a redundant tree (H₂) rather than a maximum-spanning tree (H₁) leads to improved convergence in the present case. Note, however, that an extension of the simulation length (sampling and, possibly, build-up) beyond 100 ns could almost certainly lead to converged results also using protocol H₁. Conversely, accurate results could be obtained using protocol H₂ with (sampling and, possibly, build-up) times shorter than 100 ns. Finally, it is worth emphasizing that the B&S-LEUS scheme has led to accurate estimates for the 16 relevant free-energy differences in a single simulation involving two phases of 100 ns each. In contrast, the plain LEUS estimates⁹⁷ required 16 independent simulations, each involving two phases of 3 and 40 ns each (total 688 ns) for a comparable precision. For comparison, corresponding estimates using plain US might well require 160 simulations (assuming 10 windows per simulation). In addition, the B&S-LEUS simulations also provide the relative free energies between anomers and epimers, the calculation of which would require 15 additional LEUS or sets of US simulations.

5. Conclusion

The present work is concerned with the development of a new method, ball-and-stick local elevation umbrella sampling (B&S-LEUS), to enhance the sampling in computer simulations of (bio)molecular systems. This approach enables in particular the calculation of conformational free-energy differences between states even in situations where the definition of these states relies on a conformational subspace involving more than a few degrees of freedom. The basic principle is to associate spheres (“balls”) to all relevant states and define a set of lines (“sticks”) that connect them, the union of all of these objects defining an active conformational subspace. A biasing potential involving confinement restraints (to restrict the sampling within the active subspace) and a memory-based term (to enforce a nearly homogeneous sampling of the active subspace, radially within spheres and longitudinally within lines) is then constructed using the local elevation (LE) procedure and applied in a subsequent umbrella sampling (US) simulation. The performance of the B&S-LEUS approach was tested here in the context of three illustrative examples: (i) the restriction of the sampling to arbitrary areas of the Ramachandran map of a (blocked) monopeptide in water; (ii) the evaluation of the relative free energies of three helical forms of a solvated polyalanine decapeptide; and (iii) the calculation of the relative free energies of the 32 isomers, anomers, and chair conformers of aqueous D-hexopyranoses. This new approach is appealing for the following three main reasons.

First, the B&S-LEUS method is generally applicable to virtually any type of reduced subspace definition and state properties to be evaluated. The degrees of freedom involved in the reduced subspace definition can be of conformational (internal coordinate of the physical system), alchemical (extended-state λ variable of λ -dynamics^{91–96}), or even thermodynamic (e.g., temperature or pressure) nature, or a any mixture of the three. The corresponding state properties can be the free energy (present work), but also any other thermodynamic (e.g., enthalpy, entropy, heat capacity, or volume) or structural (ensemble average of a given instantaneous observable) quantity.

Second, the B&S-LEUS method is a problem-oriented (engineering) scheme, which makes it suitable for the direct translation of precise scientific questions. Typical questions such as “what are the relative thermodynamic properties of a set of distinct conformational states of a given macromolecule”, “what are the relative solvation properties of a set of molecules in a given solvent”, “what are the relative binding properties of a set of molecules to a given receptor”, or “how are the simulated properties of a system altered upon changing a given set of force-field parameters” can in principle directly be translated into the language of “balls” and “sticks”, resulting in a question-specific scheme (i.e., a scheme requiring the minimal possible computational effort to answer only this specific question). Note that in the conformational context, the translation of what is meant experimentally by the word “state” (e.g., ensemble of conformations characterized by specific spectroscopic or functional properties) into a choice of reduced coordinates and conformational volumes may still hide a significant extent of complexity and ambiguity (see below).

Third, the B&S-LEUS method is in large parts “automated”. The human component resides in the choice of the reduced space, in the definition of the various “balls” and “sticks”, in the specification of a number of protocol parameters (prominently, the build-up basis force constant and force-reduction factor, the build-up and sampling durations, and, possibly, a state-assignment scheme), and in the convergence assessment and data analysis. As a simple example, the present simulations of the “mother” of all D-hexopyranoses returned 31 (converged) free-energy differences in one single simulation (2×100 ns). A corresponding evaluation of the 31 stepwise changes via thermodynamic integration or umbrella sampling, assuming 20 λ points or windows per change, would have required about 600 simulations (probably for about the same total sampling duration or more), and a considerable amount of human effort (adjustment of the staging, equilibration and sampling protocols, baby-sitting of the computer jobs, combination of the data from the individual simulations, and consideration of case-to-case problems).

Fourth, the B&S-LEUS method addresses simultaneously the searching efficiency problem (i.e., how to avoid the continuous revisiting of previously discovered configurations), the statistical efficiency problem (i.e., how to avoid compromising the optimal Boltzmann-weighted sampling of plain MD as much as possible), the transition problem (i.e., how to ensure a high number of transitions between the relevant states), and the computational efficiency problem (i.e., how to avoid an excessive computational overhead and memory requirement). As a result, for a given design of the problem geometry as defined by the various “balls” and “sticks”, this scheme is likely to provide a close-to-optimal compromise in terms of sampling efficiency, that is, a close-to maximal achievable accuracy for a given user-specified problem and a given amount of computer time.

The four above considerations suggest that the B&S-LEUS approach has a very high potential for practical applications in the area of molecular simulation. However, although it efficiently addresses the computational part of the problem, both the setup (definition of the reduced conformational subspace and active region thereof) and the post-processing (state assignment) may involve intricate issues related to the very definition of the concept of state. For example, depending on the adopted perspective, a state might be defined either as a single free-energy basin (thermodynamic definition), as a collection of configurations with low interconversion barriers (kinetic definition), as a collection of configurations with low mutual coordinate deviations (structural definition), or as a conformational region associated with given spectroscopic or functional properties (experimental definition). The choice of one type of definition or another is ultimately in the hands of the chemist, but it should be realized that different definitions may lead to very different interpretations of the simulation results.

Acknowledgment. We would like to thank Zhixiong Lin for his help during the setup of the polyalanine decapeptide simulations. Financial support from the Swiss National Science Foundation (Grant NF200021-121895) is also gratefully acknowledged.

References

- (1) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
- (2) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992–1023.
- (3) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064–4092.
- (4) Berendsen, H. J. C. *Simulating the Physical World*; Cambridge University Press: Cambridge, U.K., 2007.
- (5) Rick, S. W.; Stuart, S. J. *Rev. Comput. Chem.* **2002**, *18*, 89–146.
- (6) Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005**, *172*, 69–85.
- (7) Stern, H. A.; Berne, B. J. *J. Chem. Phys.* **2001**, *115*, 7622–7628.
- (8) Geerke, D. P.; Luber, S.; Marti, K. H.; van Gunsteren, W. F. *J. Comput. Chem.* **2008**, *30*, 514–523.
- (9) Hünenberger, P. H.; van Gunsteren, W. F. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer/Escom Science Publishers: Dordrecht, The Netherlands, 1997; pp 3–82.
- (10) Kastenholz, M.; Hünenberger, P. H. *J. Phys. Chem. B* **2004**, *108*, 774–788.
- (11) Reif, M. M.; Kräutler, V.; Kastenholz, M. A.; Daura, X.; Hnenberger, P. H. *J. Phys. Chem. B* **2009**, *113*, 3112–3128.
- (12) van Gunsteren, W. F.; Huber, T.; Torda, A. E. *AIP Conf. Proc.* **1995**, *330*, 253–268.
- (13) Berne, B. J.; Straub, J. E. *Curr. Opin. Struct. Biol.* **1997**, *7*, 181–189.
- (14) Christen, M.; van Gunsteren, W. F. *J. Comput. Chem.* **2008**, *29*, 157–166.
- (15) Hansen, H. S.; Hünenberger, P. H. *J. Comput. Chem.* **2010**, *31*, 1–23.
- (16) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
- (17) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
- (18) King, P. M. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; ESCOM Science Publishers: B.V., Leiden, The Netherlands, 1993; Vol. 26, pp 7–314.
- (19) van Gunsteren, W. F.; Beutler, T. C.; Fraternali, F.; King, P. M.; Mark, A. E.; Smith, P. E. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; ESCOM Science Publishers: B.V., Leiden, The Netherlands, 1993; Vol. 31, pp 5–367.
- (20) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (21) Straatsma, T. P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers Inc.: New York, 1996; Vol. 8, pp 1–127.

- (22) Chipot, C.; Pearlman, D. A. *Mol. Simul.* **2002**, *28*, 1–12.
- (23) Rodinger, T.; Pomès, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 164–170.
- (24) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
- (25) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798–2814.
- (26) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57–83.
- (27) van Gunsteren, W. F.; Gattin, Z. In *Foldamers: Structure, Properties and Applications*; Hecht, S., Huc, I., Eds.; Wiley: Weinheim, Germany, 2007.
- (28) Kastenholz, M. A.; Schwartz, T. U.; Hünenberger, P. H. *Biophys. J.* **2006**, *91*, 2976–2990.
- (29) Villa, A.; Mark, A. E. *J. Comput. Chem.* **2002**, *23*, 548–553.
- (30) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (31) Geerke, D. P.; van Gunsteren, W. F. *ChemPhysChem* **2006**, *7*, 671–678.
- (32) Oostenbrink, C.; van Gunsteren, W. F. *Proteins: Struct., Funct., Genet.* **2004**, *54*, 234–246.
- (33) Christ, C. D.; van Gunsteren, W. F. *J. Comput. Chem.* **2009**, *30*, 1664–1679.
- (34) Hünenberger, P. H.; Granwehr, J. K.; Aebsicher, J.-N.; Ghoneim, N.; Haselbach, E.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 7533–7544.
- (35) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (36) Valleau, J. P.; Torrie, G. M. In *Modern Theoretical Chemistry*; Berne, B. J., Ed.; Plenum Press: New York, 1977; Vol. 16, pp 9–194.
- (37) Hansen, H. S.; Daura, X.; Hünenberger, P. H. *J. Chem. Theory Comput.*, companion manuscript; DOI: 10.1021/ct1003059.
- (38) Beutler, T. C.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 1492–1497.
- (39) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–1350.
- (40) Heinz, T. N.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Chem. Phys.* **2001**, *115*, 1125–1136.
- (41) Piccinini, E.; Ceccarelli, M.; Affinito, F.; Brunetti, R.; Jacoboni, C. *J. Chem. Theory Comput.* **2008**, *4*, 173–183.
- (42) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1391–1436.
- (43) Mezei, M. *J. Comput. Phys.* **1987**, *68*, 237–248.
- (44) Hooft, R. W. W.; van Eijck, B. P.; Kroon, J. *J. Chem. Phys.* **1992**, *97*, 6690–6694.
- (45) Friedman, R. A.; Mezei, M. *J. Chem. Phys.* **1995**, *102*, 419–426.
- (46) Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450–1462.
- (47) Wang, J.; Gu, Y.; Liu, H. *J. Chem. Phys.* **2006**, *125*, 094907/1–094907/9.
- (48) Babin, V.; Roland, C.; Darden, T. A.; Sagui, C. *J. Chem. Phys.* **2006**, *125*, 204909/1–204909/9.
- (49) Marsili, S.; Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2006**, *110*, 14011–14013.
- (50) Lelièvre, T.; Rousset, M.; Stoltz, J. *Chem. Phys.* **2007**, *126*, 134111/1–134111/8.
- (51) van der Vaart, A.; Karplus, M. *J. Chem. Phys.* **2007**, *126*, 164106/1–164106/17.
- (52) Babin, V.; Roland, C.; Sagui, C. *J. Chem. Phys.* **2008**, *128*, 134101/1–134101/7.
- (53) Barnett, C. B.; Naidoo, K. *J. Mol. Phys.* **2009**, *107*, 1243–1250.
- (54) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (55) Laio, A.; Rodriguez-Forte, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.
- (56) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120/1–144120/13.
- (57) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (58) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175–1188.
- (59) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (60) Tobias, D. J.; Brooks, C. L., III. *Chem. Phys. Lett.* **1987**, *142*, 472–476.
- (61) Mitchell, M. J.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 271–275.
- (62) Simonson, T. *Mol. Phys.* **1993**, *80*, 441–447.
- (63) Beutler, T. C.; Mark, A. E.; van Schaik, R.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (64) Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3050–3053.
- (65) Helfand, E. *J. Chem. Phys.* **1979**, *71*, 5000–5007.
- (66) Boresch, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 5145–5154.
- (67) Straatsma, T. P.; Zacharias, M.; McCammon, J. A. *Chem. Phys. Lett.* **1992**, *196*, 297–302.
- (68) den Otter, W. K.; Briels, W. J. *J. Chem. Phys.* **1998**, *109*, 4139–4146.
- (69) Hermans, J.; Shankar, S. *Isr. J. Chem.* **1986**, *27*, 225–227.
- (70) Oostenbrink, C.; van Lipzig, M. M. H.; van Gunsteren, W. F. In *Comprehensive Medicinal Chemistry II Computer-Assisted Drug Design*; Taylor, J. B., Triggle, D. J., Eds.; Elsevier: Amsterdam, The Netherlands, 2007; Vol. 65, pp 1–668.
- (71) Pitera, J. W. *Curr. Opin. Drug Discovery Dev.* **2009**, *12*, 388–396.
- (72) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 577–585.
- (73) Schäfer, H.; van Gunsteren, W. F.; Mark, A. E. *J. Comput. Chem.* **1999**, *20*, 1604–1617.
- (74) Pitera, J. W.; van Gunsteren, W. F. *J. Phys. Chem.* **2001**, *105*, 11264–11274.
- (75) Oostenbrink, C.; van Gunsteren, W. F. *J. Comput. Chem.* **2003**, *24*, 1730–1739.
- (76) Oostenbrink, C.; van Gunsteren, W. F. *Chem.-Eur. J.* **2005**, *11*, 4340–4348.

- (77) Oostenbrink, C.; van Gunsteren, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6750–6754.
- (78) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Phys.* **2007**, *126*, 184110/1–184110/10.
- (79) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Phys.* **2008**, *128*, 174112/1–174112/12.
- (80) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2009**, *5*, 276–286.
- (81) Crippen, G. M.; Scheraga, H. A. *Chemistry* **1969**, *64*, 42–49.
- (82) Levy, A. V.; Montalvo, A. *SIAM J. Sci. Stat. Comput.* **1985**, *6*, 15–29.
- (83) Glover, F. *ORSA J. Comput.* **1989**, *1*, 190–206.
- (84) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695–708.
- (85) Grubmüller, H. *Phys. Rev. E* **1995**, *52*, 2893–2906.
- (86) Engkvist, O.; Karlström, G. *Chem. Phys.* **1996**, *213*, 63–76.
- (87) Fukunishi, Y.; Mikami, Y.; Nakamura, H. *J. Phys. Chem. B* **2003**, *107*, 13201–13210.
- (88) van Gunsteren, W. F. Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P. Mark, A. E. Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 manual and user guide*; Verlag der Fachvereine: Zürich, Switzerland, 1996.
- (89) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- (90) Perić-Hassler, L.; Hansen, H. S.; Baron, R.; Hünenberger, P. H. *Carbohydr. Res.* **2010**, *345*, 1781–1801.
- (91) Tidor, B. *J. Phys. Chem.* **1993**, *97*, 1069–1073.
- (92) Kong, X.; Brooks, C. L., III. *J. Chem. Phys.* **1996**, *105*, 2414–2423.
- (93) Guo, Z.; Brooks, C. L., III; Kong, X. *J. Phys. Chem. B* **1998**, *102*, 2032–2036.
- (94) Guo, Z.; Brooks, C. L., III. *J. Am. Chem. Soc.* **1998**, *120*, 1920–1921.
- (95) Leitgeb, M.; Schröder, C.; Boresch, S. *J. Chem. Phys.* **2005**, *122*, 084109/1–084109/15.
- (96) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*, 074115/1–074115/12.
- (97) Hansen, H. S.; Hünenberger, P. H. *J. Comput. Chem.*, in press.
- (98) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
- (99) Depaepe, J.-M.; Ryckaert, J.-P.; Paci, E.; Ciccotti, G. *Mol. Phys.* **1993**, *79*, 515–522.
- (100) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–308.
- (101) Gö, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535–542.
- (102) Gottlieb, M.; Bird, R. B. *J. Chem. Phys.* **1976**, *65*, 2467–2468.
- (103) Kannan, S.; Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 697–706.
- (104) Xu, C.; Wang, J.; Liu, H. *J. Chem. Theory Comput.* **2008**, *4*, 1348–1359.
- (105) Liu, Z.; Berne, B. J. *J. Chem. Phys.* **1993**, *99*, 6071–6077.
- (106) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603/1–020603/4.
- (107) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978–9985.
- (108) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 024106/1–024106/15.
- (109) Sheppard, D.; Terrel, R.; Henkelman, G. *J. Chem. Phys.* **2008**, *128*, 134106/1–134106/10.
- (110) Vanden-Eijnden, E.; Venturoli, M. *J. Chem. Phys.* **2009**, *130*, 194103/1–194103/17.
- (111) Schlitter, J.; Swegat, W.; Mülders, T. *J. Mol. Model.* **2001**, *7*, 171–177.
- (112) Schlitter, J.; Engels, M.; Krüger, P. *J. Mol. Graphics* **1994**, *12*, 84–89.
- (113) Makino, Y.; Itoh, N. *BMC Struct. Biol.* **2008**, *8*, 46/1–46/15.
- (114) Fodje, M. N.; Al-Karadaghi, S. *Protein Eng.* **2002**, *15*, 353–358.
- (115) Al-Karadaghi, S.; Cedergren-Zeppezauer, E. S.; Hovmöller, S.; Petratos, K.; Terry, H.; Wilson, K. S. *Acta Crystallogr., Sect. D* **1994**, *50*, 793–807.
- (116) Low, B. W.; Grenville-Wells, H. *J. Chemistry* **1953** 9785801.
- (117) Lee, K.-H.; Benson, D. R.; Kuczera, K. *Biochemistry* **2000**, *39*, 13737–13747.
- (118) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1988** 20/601619.
- (119) Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1951** 37235240.
- (120) Arnott, S.; Wonacott, A. J. *J. Am. Chem. Soc.* **1966**, *88*, 2598–2599.
- (121) Perutz, M. F. *Nature* **1951**, *167*, 1053–1054.
- (122) Pickett, H. M.; Strauss, H. L. *J. Am. Chem. Soc.* **1970**, *92*, 7281–7290.
- (123) Krause, E. F. *Taxicab Geometry*; Dover Publications: New York, 1986.
- (124) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719–1751.
- (125) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; Vol. 33, pp 1–342.
- (126) Feynman, R. P.; Leighton, R. B.; Sands, M. *The Feynman Lectures on Physics*; Addison-Wesley: Boston, MA, 1963.
- (127) Hockney, R. W. *Methods Comput. Phys.* **1970**, *9*, 136–211.
- (128) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (129) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (130) Berendsen, H. J. C.; van Gunsteren, W. F.; Zwinderman, H. R. J.; Geurtsen, R. G. *Ann. N. Y. Acad. Sci.* **1986** 482269285.
- (131) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.
- (132) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.