JCTC Journal of Chemical Theory and Computation

Enhanced Conformational Sampling in Molecular Dynamics Simulations of Solvated Peptides: Fragment-Based Local Elevation Umbrella Sampling

Halvor S. Hansen,[†] Xavier Daura,[‡,§] and Philippe H. Hünenberger*[,†]

Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland, Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain, and Catalan Institution for Research and Advanced Studies (ICREA), E-08010 Barcelona, Spain

Received June 5, 2010

Abstract: A new method, fragment-based local elevation umbrella sampling (FB-LEUS), is proposed to enhance the conformational sampling in explicit-solvent molecular dynamics (MD) simulations of solvated polymers. The method is derived from the local elevation umbrella sampling (LEUS) method [Hansen and Hünenberger, *J. Comput. Chem.* **2010**, *31*, 1−23], which combines the local elevation (LE) conformational searching and the umbrella sampling (US) conformational sampling approaches into a single scheme. In LEUS, an initial (relatively short) LE build-up (searching) phase is used to construct an optimized (grid-based) biasing potential within a subspace of conformationally relevant degrees of freedom, which is then frozen and used in a (comparatively longer) US sampling phase. This combination dramatically enhances the sampling power of MD simulations but, due to computational and memory costs, is only applicable to relevant subspaces of low dimensionalities. As an attempt to expand the scope of the LEUS approach to solvated polymers with more than a few relevant degrees of freedom, the FB-LEUS scheme involves an US sampling phase that relies on a superposition of low-dimensionality biasing potentials optimized using LEUS at the fragment level. The feasibility of this approach is tested using polyalanine (poly-Ala) and polyvaline (poly-Val) oligopeptides. Two-dimensional biasing potentials are preoptimized at the monopeptide level, and subsequently applied to all dihedral-angle pairs within oligopeptides of 4, 6, 8, or 10 residues. Two types of fragment-based biasing potentials are distinguished: (i) the basin-filling (BF) potentials act so as to "fill" free-energy basins up to a prescribed free-energy level above the global minimum; (ii) the valley-digging (VD) potentials act so as to "dig" valleys between the (four) free-energy minima of the two-dimensional maps, preserving barriers (relative to linearly interpolated free-energy changes) of a prescribed magnitude. The application of these biasing potentials may lead to an impressive enhancement of the searching power (volume of conformational space visited in a given amount of simulation time). However, this increase 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). As a result, it appears difficult to engineer FB-LEUS schemes representing a significant improvement over plain MD, at least for the systems considered here.

1. Introduction

Classical atomistic simulations, in particular molecular dynamics (MD), represent nowadays a powerful tool complementary to experiment for investigating the properties of atomic and molecular systems relevant in physics, chemistry, and biology.[1−4]

The success of these methods in the context of condensed-phase systems results in particular from a favorable trade-off between model resolution and computational cost. On the one hand, classical atomistic models, although they represent an approximation to quantum mechanics, can still provide a realistic description of many molecular systems

* Corresponding author. Phone: +41 44 632 5503. Fax: +41 44 632 1039. E-mail:phil@igc.phys.chem.ethz.ch.

[†] ETH Zürich.

[‡] Universitat Autònoma de Barcelona.

[§] Catalan Institution for Research and Advanced Studies (ICREA).

10.1021/ct1003059
Published on Web 08/18/2010

at spatial and temporal resolutions on the order of 0.1 nm and 1 fs, respectively. On the other hand, their computational cost remains tractable at present for system sizes and time scales on the order of 10 nm and 100 ns, respectively. These scales are sufficient to enable in many cases (i) an appropriate description of bulk-like solvation (discrete solvent molecules, sufficiently large solvation range), (ii) a reliable calculation of thermodynamic properties via statistical mechanics (converged ensemble averages and free energies), and (iii) a direct comparison with experimental data (structural, thermodynamic, transport, and dynamic observables measured on similar spatial and temporal scales).

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] (neglect or mean-field treatment of electronic and quantum effects), (ii) the approximate force-field representation of interatomic interactions[2−4,9] (potential energy function with simplified functional form and empirical parameters, various parameter transferability and combination assumptions), (iii) the presence of finite-size and surface effects[10,11] (related to the still microscopic size of the simulated systems), and (iv) the insufficient conformational sampling[9,12−15] (related to the still very short time scale of the simulations). The reduction of the last type of errors can be viewed as a first-priority target in the improvement of simulation methodologies, because insufficient sampling results in errors that are predominantly nonsystematic, while the three other types of errors are systematic.

Among the applications of MD simulation, the calculation of the relative free energies associated with different conformational states of a (bio)molecular system (and, possibly, of corresponding free-energy profiles or maps) is typically very sensitive to sampling errors. For example, the direct evaluation of the relative free energies corresponding to different structural motives of a solvated oligopeptide (e.g., various helix or sheet motives, folded and unfolded states) using plain MD simulations is only possible nowadays for short peptides,[16−18] requires a sizable amount of computing time, and leads to results still affected by large uncertainties. These difficulties result directly from the long time scale associated with the corresponding interconversion processes, for example, for α-peptides in water based on refs 17, 19, and 20, α-helix formation[21] (∼200 ns), loop closing[19,22] (∼0.05−1 μs), β-hairpin folding[23] (∼1−10 μs), and mini-protein folding[24] (∼1−10 μs).

In the numerous cases where a direct-counting approach[16,25,26] based on plain MD simulations is not applicable, one may resort to umbrella sampling[27,28] (US). The US approach relies on the use of a time-independent biasing potential in a relevant conformational subspace (US subspace) that forces the sampling of the conformational states under consideration with a sufficient number of interconversion transitions. In this case, the free-energy differences can be calculated from the ratio of the reweighted numbers of conformations assigned to the different states, the reweighting acting as a correction for the effect of the biasing. This approach is easily extended to the evaluation of free-energy profiles or maps. In practice, however, the direct design of a biasing potential fulfilling the above conditions for significantly differing conformational states is difficult. There exist three basic approaches to overcome this problem.

In multiple-windows schemes,[29−32] a set of MD simulations is performed using different local biasing potentials (typically harmonic) that restrict the sampling to specific regions of the US subspace. The data from the different (overlapping) biased ensembles is then assembled, either manually or automatically (e.g., using the weighted-histogram[33−36] or umbrella-integration[37,38] methods) to construct the complete free-energy profile or map. In adaptive (iterative) schemes,[35,39−48] successive equilibrium MD simulations are used to progressively optimize a nonlocal biasing potential (typically nonharmonic) with the goal of achieving a nearly complete sampling of the US subspace. This can be done by examining the conformational probability at each step and adjusting the biasing potential for the next step so as to remove undersampled regions. The data from the different steps is then combined to construct the final free-energy profile or map. Finally, in memory-based schemes,[15,49−52] the strategy is similar to that of the adaptive approach, but a single (relatively short) nonequilibrium (memory-building) MD simulation is used as a preoptimization tool for the nonlocal biasing potential, followed by a single (comparatively longer) equilibrium US simulation for production.

The last approach relies on methods previously 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, without any requirement on the probability distribution of these configurations. Memory-based methods are probably the most efficient types of MD-based searching methods available nowadays. They rely on the progressive build-up of a time-dependent penalty potential, which prevents the continuous revisiting of previously discovered configurations. Many closely related variants of this approach can be found in the literature, including (chronologically) the deflation,[53] tunneling,[54] tabu search,[55] local elevation,[56] conformational flooding,[57] Engkvist−Karlström,[58] adaptive reaction coordinate force,[49] adaptive biasing force,[52] metadynamics,[50,51] and filling potential[59] methods. The first practically useful implementation of a memory-based searching scheme in the context of (bio)molecular systems with explicit solvation is probably the local elevation (LE) method of Huber, Torda, and van Gunsteren,[56] as implemented in the GROMOS96 program.[60,61] 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 weighted sum of local (grid-based, short-ranged) repulsive Gaussian functions, the corresponding weights being made proportional to the number of previous visits to the specific conformation (grid cell). Because memory-based searching methods (such as the LE method) have a time-dependent Hamitonian, 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 after reweighting. This can be done by observing that at the end of a memory-based search, the penalty potential has essentially "flattened" the free-energy surface in the considered subspace up to a certain threshold value above the lowest minimum discovered.[58] 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[15,62] (LEUS) approach. This scheme consist of two steps: (i) a LE build-up (searching) phase, that is used to construct an optimized memory-based biasing potential within a LE subspace of N_{LE} conformationally relevant degrees of freedom; (ii) an US sampling phase, where the (frozen) memory-based potential is used to generate a biased ensemble with extensive coverage of the US subspace defined by the same $N_{US} = N_{LE}$ degrees of freedom. A successful build-up phase will produce a biasing potential resulting in a nearly homogeneous coverage of the relevant subspace (up to a given free-energy level) during the subsequent sampling phase. Thermodynamic information appropriate for the physical (unbiased) ensemble can then be recovered from the simulated data by means of a simple reweighting procedure.[15,27,28] Note that the LEUS approach bears some analogies with other memory-based sampling approaches[14,15] such as the adaptive umbrella sampling,[40,41,63] adaptive biasing force,[52,64,68] adaptive reaction coordinate force,[49,66−68] and metadynamics[50,51] methods, but within a two-step implementation where the US sampling phase is used to correct for the inaccuracy of the (nonequilibrium) LE build-up phase, leading to a number of advantages in terms of efficiency, accuracy, and robustness.[15] A similar two-step approach can also be found in related schemes developed by Babin et al.,[44,48] Ensing et al.,[69] and Li et al.[70]

The LEUS approach was previously applied to the calculation of the relative free energies of β-D-glucopyranose ring conformers in water,[15] to the calculation of Ramachandran free-energy maps for the 11 glucose-based disaccharides in water,[62] and, more recently, to the parametrization of a new GROMOS carbohydrate force field.[71] This scheme was found to dramatically enhance the sampling power of MD simulations and to permit the calculation of accurate free-energy differences between relevant conformational states (as well as free-energy profiles or maps) for LEUS subspaces of low dimensionalities ($N_{LE} = N_{US} = 1-4$). This approach is efficient, nearly all the computational effort being invested in the actual sampling phase rather than in searching and equilibration, and robust, the method being only weakly sensitive to the details of the build-up protocol.[15,62]

The LEUS approach represents a powerful sampling-enhancement tool in cases where the relevant conformational subspace is of low dimensionality. However, it becomes inapplicable as such for high-dimensional problems. Consider, for example, a decapeptide with 20 relevant degrees of freedom (successive ϕ and ψ backbone dihedral angles). If each degree of freedom is discretized into 32 bins (as in refs 15 and 62) and if the biasing potential is expected to map out about 50% of the relevant conformational subspace, the number of local functions required is approximately 10^{30}, which is cleary intractable in terms of both memory and build-up duration requirements.

A tentative solution to this problem relies on the optimization of fragment-based biasing potentials of low dimensionalities (e.g., in the peptide case, two-dimensional potentials optimized for the ϕ and ψ dihedral angles corresponding to a specific type of residue pair, with a build-up involving the corresponding dipeptide, that is, $N_{LE} = 2$) and their simultaneous application to each of the corresponding fragments in a molecule (e.g., 10 successive ϕ and ψ pairs along the peptide backbone, that is, $N_{US} = 20$), following a similar principle as that developed in refs 72 and 73. The resulting combined biasing potential would eliminate the influence of the local conformational preferences of the successive fragments, without affecting nonlocal (longer-ranged) influences.

Such an extension of the LEUS approach to solvated polymers with more than a few relevant degrees of freedom is the goal of the present work. The resulting method will be termed fragment-based LEUS (FB-LEUS). More specifically, the FB-LEUS scheme involves an US sampling phase that relies on a superposition of low-dimensionality biasing potentials optimized using LEUS at the fragment level, that is, a situation with $N_{US} = N_F N_{LE}$, where N_F is the number of fragments in the molecule considered. This new combination appears to be very versatile, because optimized biasing potentials can in principle be precalculated for fragments (e.g., in the peptide case, ϕ and ψ backbone dihedral angles for all possible dipeptides), stored in a database, and later applied to larger molecules (e.g., on the corresponding dipeptide units of an oligopeptide with specified sequence), so as to enhance the sampling with a limited additional computational and memory cost.

The feasibility of the FB-LEUS approach is investigated here in the context of polyalanine (poly-Ala) and polyvaline (poly-Val) oligopeptides in water. Two-dimensional biasing potentials are preoptimized at the monopeptide level, distinguishing between alanine and valine, as well as between N-blocked, C-blocked, and N&C-blocked monopeptides, assumed representative for the C-terminal, N-terminal, and intermediate residues of the oligopeptides, respectively. These potentials are then applied to all dihedral-angle pairs within (unblocked) oligopeptides of 4, 6, 8, or 10 residues. Two types of fragment-based biasing potentials are distinguished: (i) the basin-filling (BF) potentials act so as to "fill" free-energy basins up to a prescribed free-energy level above the global minimum; (ii) the valley-digging (VD) potentials act so as to "dig" valleys between the (four) relevant free-energy minima of the two-dimensional maps, preserving barriers (relative to linearly interpolated free-energy changes) of a prescribed magnitude. Two important differences between the BF and VD biasing potentials at the fragment level are that (i) the VD potential opens up a much smaller volume of irrelevant conformational space (i.e., conformations associated with low Boltzmann weights in the physical ensemble) and (ii) the VD potential preserves the relative

free energies of the (four) minima, while the BF potential equalizes their values. However, both potentials will promote an increased number of tranistions between these minima. The results of the simulations involving different FB-LEUS biasing schemes are analyzed in terms of searching efficiency (volume of conformational space visited in a given amount of simulation time), statistical efficiency (representativeness of the biased ensemble in terms of the conformational distribution appropriate for the physical ensemble), and reliability of the predicted low free energy conformations and associated free energies (convergence with time, number of interconversion transitions, and mutual overlap across schemes).

2. Computational Details

2.1. Simulation Procedure. All MD simulations were carried out using a modified version of the GROMOS05 program,[74] together with the GROMOS 53A6 force field[75] and the simple-point charges (SPC) water model.[76] They were performed under periodic boundary conditions based on cubic computational boxes and in the isothermal−isobaric (*NPT*) ensemble at 300 K and 1 bar. Newton's equations of motion were integrated using the leapfrog algorithm[77,78] with a 2 fs time step. The SHAKE procedure[79] 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 temperature was maintained by weakly coupling the solute and solvent degrees of freedom (jointly) to a temperature bath[80] at 300 K with a relaxation time of 0.1 ps. The pressure was maintained by weakly coupling the atomic coordinates and box dimensions (isotropic coordinate scaling, group-based virial) to a pressure bath[80] at 1 bar 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.[60] The center of mass motion was removed every time step. The nonbonded interactions were handled using a twin-range cutoff scheme,[2,81] based on short- and long-range cutoff distances of 0.8 and 1.4 nm, respectively, and an 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,[82] with a relative dielectric permittivity of 61 as appropriate for the SPC water model.[31] Values of the successive backbone ϕ and ψ dihedral angles, as well as of the biasing potential, were written to file every time step for subsequent analysis.

The simulated systems (computational box) consisted of one solute molecule and N_{H_2O} water molecules. The solute molecules considered (Figure 1) were (partially) blocked alanine or valine monopeptides (Ala_1^N, Ala_1^C, Ala_1^{NC}, Val_1^N, Val_1^C or Val_1^{NC}, along with $N_{H_2O} = 1300$) and unblocked alanine or valine oligopeptides (Ala_n or Val_n with $n = 4,\ 6,\ 8$, or 10, along with $N_{H_2O} = 2100,\ 3500,\ 5500$, and 8000, respectively). Note that the term "dipeptide" is sometimes used rather than "monopeptide" for the compounds Ala_1 and Val_1, the latter term appearing more consistent to the authors. Unblocked termini were modeled using parameters appropriate for the uncharged forms of the corresponding free groups (unprotonated amine and protonated carboxylic acid). Blocked termini were modeled using parameters appropriate for the acetylated (N-terminus) and methylamidated (C-terminus) forms of these groups. Although the termini of unblocked peptides are expected to be ionized in aqueous solution at neutral pH, the choice was made here to consider uncharged termini for the oligopeptides Ala_n and Val_n ($n = 4,\ 6,\ 8$, or 10), so as to avoid conformational ensembles dominated by (nonlocal) interactions between terminal charges.[11,83] In addition to providing a simpler context for the testing of a fragment-based sampling-enhancement approach, this situation is also more representative of an experimental setup involving a blocked peptide, an excess of counterions or a peptide segment within a protein. The simulations were initiated from fully extended peptide structures, subsequently relaxed by 500 ps MD equilibration.

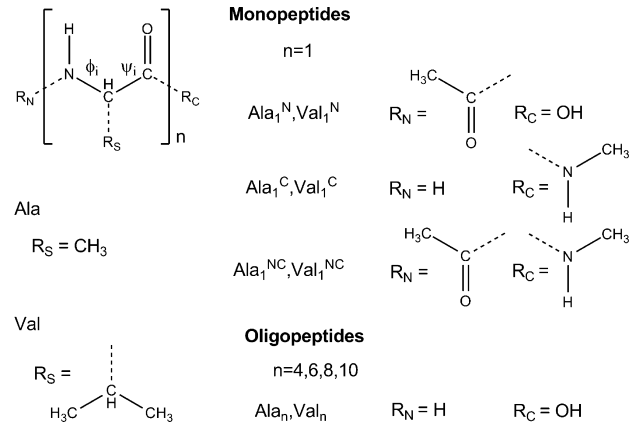

Figure 1. Solute molecules considered in the present study. These include (partially) blocked alanine or valine monopeptides (Ala_1^N, Ala_1^C, Ala_1^{NC}, Val_1^N, Val_1^C or Val_1^{NC}) and unblocked alanine or valine oligopeptides (Ala_n or Val_n with $n = 4,\ 6,\ 8$, or 10). Unblocked termini correspond to unprotonated amine (N-terminus) and protonated carboxylic acid (C-terminus) groups. Blocked termini correspond to acetylated (N-terminus) and methylamidated (C-terminus) groups.

The oligopeptide simulations (Ala_n or Val_n with $n = 4,\ 6,\ 8$, or 10) were carried out using the proposed FB-LEUS approach. Corresponding unbiased simulations were also undertaken for comparison. The fragment-based biasing potentials were constructed at the monopeptide level (Ala_1^N, Ala_1^C, Ala_1^{NC}, Val_1^N, Val_1^C, and Val_1^{NC}) following different approaches (section 2.2). The design of these potentials relied on Ramachandran free-energy maps for the monopeptides, evaluated using the standard LEUS method.[15] Note that truncated-polynomial local functions (of widths equal to the grid spacing) were used in the present work rather than the previously employed Gaussian functions,[15,56,62] as described in Appendix A.

2.2. FB-LEUS Schemes. If the essentially stiff peptide bonds are omitted, the backbone conformation of a mono- or oligopeptide, Ala_n or Val_n ($n = 1,\ 4,\ 6,\ 8$, or 10, irrespective of its terminal blocking), is defined by n pairs of dihedral angles ϕ_i and ψ_i with $i = 1, ..., n$, numbered starting from the N-terminus and simply noted ϕ and ψ for $n = 1$ (Figure 1).

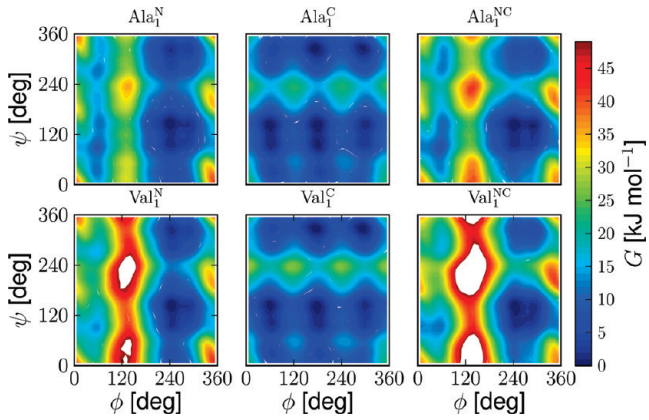

Figure 2. Ramachandran free-energy maps, $G(\phi, \psi)$, for the monopeptides (Ala_1^N, Ala_1^C, Ala_1^{NC}, Val_1^N, Val_1^C, or Val_1^{NC}, Figure 1), evaluated using LEUS simulations ($t_{LE} = 15$ ns for alanine or 20 ns for valine, $t_{US} = 50$ ns). All maps are anchored by the condition $G(\phi, \psi) = 0$ at the global minimum. Regions displayed in white are associated with very high relative free energies (>50 kJ mol^{-1}) due to steric clashes. Note that the maps are drawn considering [0°, 360°] dihedral-angle ranges rather than the more usual [−180°, 180°] ranges.

Simulations of the six (partially) blocked alanine or valine monopeptides were used to preoptimize a library of fragment-based two-dimensional biasing potentials. To this purpose, Ramachandran free-energy maps, $G(\phi, \psi)$, were first evaluated for the six compounds using the LEUS method,[15] with a dimensionality $N_{LE} = N_{US} = 2$ (ϕ and ψ), a number of grid points per dimension $N_G = 32$ (angular spacing 11.25°), a force-constant increment per visit $k_{LE} = 2 \times 10^{-3}$ kJ mol^{-1}, a build-up duration $t_{LE} =$ 15 ns (alanine) or 20 ns (valine), and a sampling duration $t_{US} = 50$ ns (Appendix A). The resulting maps, obtained from the sampling phases of these simulations after application of the proper reweighting[15] are displayed in Figure 2. These maps were calculated with the same grid spacing as used in the LEUS scheme and anchored by the condition $G(\phi, \psi) = 0$ at the global minimum. The maps for N-blocked and N&C-blocked monopeptides present four free-energy basins centered at ϕ values of about 60° and 260°, along with ψ values of about 120° and 300°, the main (deepest and widest) basin corresponding to $(\phi, \psi) = (260°, 120°)$. For these monopeptides, the rotation around ψ is associated with low barrieres (similar for alanine and valine), while the rotation around ϕ is associated with a higher barrier at $\phi \approx 120°$ (significantly higher for valine compared with alanine). In contrast, the maps for C-blocked monopeptides present six free-energy basins centered at ϕ values of about 60°, 180°, and 300°, along with ψ values of about 120° and 300°, all of comparable depths and widths. For these monopeptides, the rotations around both ϕ and ψ are associated with low barriers (similar for alanine and valine). The fragment-based biasing potentials were derived from these maps according to two different procedures.

The basin-filling (BF) biasing potentials are defined as

$$\mathscr{U}_{bias}^{BF}(\phi, \varphi; h) = \begin{cases} h - G(\phi, \varphi) & \text{if } G(\phi, \varphi) \le h \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where h denotes a free-energy level (relative to the global minimum of the map) up to which the biased map $G + \mathscr{U}_{bias}^{BF}$ is "flattened". In practice, $\mathscr{U}_{bias}^{BF}$ is constructed from the calculated free-energy surface by using a grid-based version of eq 1, as detailed in Appendix B. The BF approach is illustrated schematically in Figure 3a. Three values of h were considered, $h = 10$, 20, or 30 kJ mol^{-1}, resulting in BF biasing potentials labeled F_{10}, F_{20}, and F_{30}, respectively. These three types of biasing potentials were evaluated separately for Ala_1^N, Ala_1^C, Ala_1^{NC}, Val_1^N, Val_1^C, and Val_1^{NC}. Those for Ala_1^{NC} and Val_1^{NC} are illustrated in Figure 3b in the form of biased maps $G + \mathscr{U}_{bias}^{BF}$.

The valley-digging (VD) biasing potentials are defined as

$$\mathscr{U}_{bias}^{VD}(\phi, \psi; h) = \begin{cases} h - \Delta G_k(\phi, \psi) & \text{if } (\phi, \psi) \rightarrow k \text{ and } \Delta G_k(\phi, \psi) \ge h \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Here, the biasing potential is defined in terms of eight line segments k, each extending along either ϕ or ψ, and connecting the points (ϕ, ψ) of the set {(61.875°, 118.125°), (275.625°, 118.125°),(61.875°, 298.125°),(275.625°, 298.125°)}, four of them directly and four of them across a period. These points belong to the LEUS grid, so that the connecting segments encompass a series of grid points. For a segment k extending from $\phi_k^{(b)}$ to $\phi_k^{(e)}$ along ϕ at ψ_k, the quantity $\Delta G_k(\phi, \psi)$ in eq 2 is defined as

$$\Delta G_k(\phi, \psi) = G(\phi, \psi) - \{[1 - \lambda_k(\phi)]G(\phi_k^{(b)}, \psi_k) + \lambda_k(\phi)G(\phi_k^{(e)}, \psi_k)\} \quad (3)$$

where

$$\lambda_k(\phi) = \frac{\phi - \phi_k^{(b)}}{\phi_k^{(e)} - \phi_k^{(b)}} \quad (4)$$

and the notation $(\phi, \psi) \rightarrow k$ denotes a point belonging to segment k, that is, satisfying

$$2|\psi - \psi_k| \le \sigma \quad \text{and} \quad \phi_k^{(b)} + |\psi - \psi_k| < \phi < \phi_k^{(e)} - |\psi - \psi_k| \quad (5)$$

Similar definitions apply to a segment k extending from $\psi_k^{(b)}$ to $\psi_k^{(e)}$ along ψ at ϕ_k. When a point belongs to segment k, λ_k measures its fractional longitudinal distance from the beginning point of the segment relative to the full segment length, and $\Delta G_k(\phi, \psi)$ is the difference between the free energy at this point and a free energy value linearly interpolated from those at the beginning and end points of the segment. Setting h to zero will lead to a biased map $G + \mathscr{U}_{bias}^{VD}$ presenting a connection of the four above points via eight line segments associated with linearly varying free energies (valleys). The actual value of h denotes a free-energy level (relative to this linear free-energy variation) down to which the valleys are "dug". Note that the second condition in eq 5 is formulated in such a way that a given (ϕ,ψ) point can belong to at most one segment, even close to the connecting points, that is, that the eight segments are nonoverlapping. In practice, $\mathscr{U}_{bias}^{VD}$ is constructed from the calculated free-energy surface by using grid-based versions of eqs 2−5, as detailed in Appendix B. In

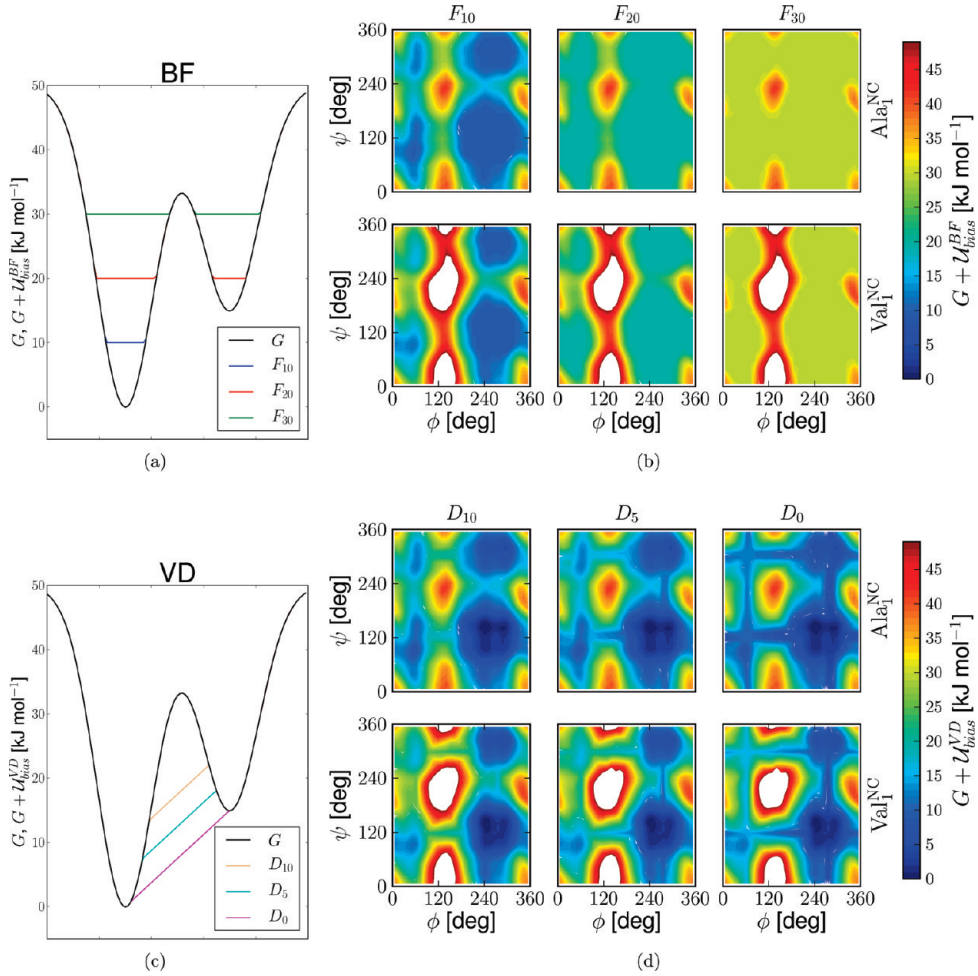

Figure 3. Schematic illustrations of the BF and VD procedures for generating fragment-based biasing potentials and corresponding biased maps for the monopeptides $\mathrm{Ala}_1^{\mathrm{NC}}$ and $\mathrm{Val}_1^{\mathrm{NC}}$ (Figure 1) associated with different parameters h of the BF and VD potentials. The biased maps are defined by $G + \mathcal{U}_{\mathrm{bias}}$, where $G(\phi, \psi)$ is the free energy of the physical system (Figure 2) and $\mathcal{U}_{\mathrm{bias}}(\phi, \psi; h)$ is the biasing potential (eqs 1 or 2). The biasing potentials F_{10}, F_{20}, and F_{30} correspond to the BF procedure with $h = 10$, 20, or 30 kJ mol^{-1}, respectively. The biasing potentials D_{10}, D_5, and D_0 correspond to the VD procedure with $h = 10$, 5, or 0 kJ mol^{-1}, respectively. Note that the maps are drawn considering [0°, 360°] dihedral-angle ranges rather than the more usual [−180°, 180°] ranges.

addition, to avoid spurious transverse oscillatory motions within the segments, the grid-based functions corresponding to all lines parallel to the eight above segments are adjusted so as to level off the component of the biasing force orthogonal to the lines at a maximal magnitude of 1 kJ mol^{-1} deg^{-1}. The VD approach is illustrated schematically in Figure 3c. Three values of h were considered, $h = 10$, 5, or 0 kJ mol^{-1}, resulting in VD biasing potentials labeled D_{10}, D_5, and D_0, respectively. These three types of biasing potentials were evaluated separately for $\mathrm{Ala}_1^{\mathrm{N}}$, $\mathrm{Ala}_1^{\mathrm{C}}$, $\mathrm{Ala}_1^{\mathrm{NC}}$, $\mathrm{Val}_1^{\mathrm{N}}$, $\mathrm{Val}_1^{\mathrm{C}}$, and $\mathrm{Val}_1^{\mathrm{NC}}$. Those for $\mathrm{Ala}_1^{\mathrm{NC}}$ and $\mathrm{Val}_1^{\mathrm{NC}}$ are illustrated in Figure 3d in the form of biased maps $G + \mathcal{U}_{\mathrm{bias}}^{\mathrm{VD}}$.

For comparison purposes, unbiased simulations were also undertaken. The corresponding "zero" biasing potential will be noted U (unbiased, an equivalent notation would be F_0). In this case, the biased maps analogous to those reported in Figure 3 for the BF and VD potentials evaluated for $\mathrm{Ala}_1^{\mathrm{NC}}$ and $\mathrm{Val}_1^{\mathrm{NC}}$ are simply the unbiased maps of Figure 2 (right panels).

After the definition of the different fragment-based biasing potentials, these were applied in simulations of the longer (unblocked) oligopeptides Ala_n and Val_n ($n = 4$, 6, 8, or 10) in the following way. The potentials derived for $\mathrm{Ala}_1^{\mathrm{C}}$ and $\mathrm{Val}_1^{\mathrm{C}}$ were applied to (ϕ_1, ψ_1) in Ala_n and Val_n, respectively, the potentials derived for $\mathrm{Ala}_1^{\mathrm{N}}$ and $\mathrm{Val}_1^{\mathrm{N}}$ to (ϕ_n, ψ_n) in Ala_n and Val_n, respectively, and the potentials derived for $\mathrm{Ala}_1^{\mathrm{NC}}$ and $\mathrm{Val}_1^{\mathrm{NC}}$ to all other dihedral angles (ϕ_i, ψ_i) with $i = 2, ..., n - 1$ in Ala_n and Val_n, respectively. Note that if, for simplicity, the same terminal blocking groups have been used here for Ala_1 and Val_1, the use of groups involving an isopropyl (rather than a methyl) termination would possibly be more appropriate for Val_1, being more representative for Val−Val interactions within a peptide. An alternative approach would involve unblocked Ala_2 and Val_2 fragments, with ψ_1 and ϕ_2 taken as representative for (ψ_i,ϕ_{i+1}) with $i = 1, ..., n - 1$ within Ala_n and Val_n, respectively (leaving ϕ_1 and ψ_n unbiased).

The different FB-LEUS simulations were carried out for a sampling duration $t_{\mathrm{US}} = 50$ ns (100 ns for the unbiased simulations). This resulted in a set of 56 simulations, depending on the type of oligopeptide considered (Ala or

Val), number of residues ($n = 4,\ 6,\ 8$, or 10), and type of biasing potential (F_{10}, F_{20}, F_{30}, D_{10}, D_5, D_0, or U).

2.3. Analysis Procedure. The overall efficiency achieved by a specific sampling-enhancement scheme is a combination of two factors: (i) the searching efficiency, that is, the ability of the scheme to search for low free energy regions across a wide extent of conformational space, thereby escaping local free-energy minima and overcoming free-energy barriers; (ii) the statistical efficiency, that is, the fraction of the sampled configurations actually relevant in terms of the conformational (Boltzmann) distribution characteristic of the physical system.

The searching efficiency was assessed for the Ala_n and Val_n oligopeptides by monitoring the cumulative number $N_n(t)$ of unique peptide backbone conformations discovered up to time t during the sampling phases of the simulations involving different biasing potentials. In the present study, a unique backbone conformation is defined by an integer corresponding to a string of $2n$ bits (arranged from the highest-weight to the lowest-weight bit). Given a numbering of the backbone dihedral angles from the N-terminus to the C-terminus of the peptide, bit $2i - 1$ ($i = 1, ..., n$) is set to zero if the backbone dihedral angle ϕ_i is in the range [130°; 360°] and to one otherwise, while bit $2i$ ($i = 1, ..., n$) is set to zero if the backbone dihedral angle ψ_i is in the range [0°; 230°] and to one otherwise. These intervals were selected by consideration of the free-energy maps for the monopeptides as approximately defining four distinct free-energy basins (Figure 2). At the monopeptide level ($n = 1$), the bit strings 00 and 01 correspond to the sheet and helical regions of the Ramachandran map, respectively, while 10 and 11 correspond to typically less populated regions (in proteins). Based on this assignment, there are in total $N_n^{\max} = 4^n$ unique backbone conformations for Ala_n or Val_n, defining the corresponding exhaustive-search upper bound for $N_n(t)$. The different $N_n(t)$ curves were fitted to stretched-exponential functions of the form

$$N_n(t) = N_n^{\max}\{1 - \exp[-(\tau_n^{-1}t)^{\alpha_n}]\} \tag{6}$$

The resulting parameters α_n and τ_n are further referred to as the stretching exponent and the characteristic searching time, respectively, of the simulation involving a specific biasing potential. The stretching exponent α_n is expected to be one for trajectories representing an entirely random configuration generation process. Negative deviations ($0 < \alpha_n < 1$) account for two effects: (i) the presence of time correlations in real trajectories, which represent a diffusion process in configuration space; (ii) the presence of a probability bias in real trajectories, which generate configurations according to a (possibly biased) Boltzmann distribution in configuration space. Although the two effects are not clearly separable in practice in terms of their influence on the searching rate, they both induce a tendency of the system to revisit previously discovered configurations, leading to a slower evolution of $N_n(t)$ toward $N_n^{\max}$. Irrespective of the value of α_n, the characteristic searching time τ_n represents the time required for the trajectory to cover a fraction $1 - e^{-1}$ (63%) of the entire configuration space accessible to the system. This value can be reexpressed in terms of an effective visiting time $\tilde{\tau}_n$, defined as

$$\tilde{\tau}_n = \frac{\tau_n}{(1 - e^{-1})N_n^{\max}} \tag{7}$$

Under the assumption of an entirely random configuration generation process (i.e., when $\alpha_n = 1$), $\tilde{\tau}_n$ can be interpreted as the average time separating the visit of two conformations (including newly discovered and revisited ones) or, equivalently, as the average time the trajectory spends in a given conformation before leaving it.

The statistical efficiency was assessed for the Ala_n and Val_n oligopeptides by calculating, for the simulations involving different biasing potentials, the quantity F_n defined as[62]

$$F_n = N_f^{-1} \exp[-\sum_k^{N_f} p_k \ln p_k] \tag{8}$$

where N_f is the number of considered trajectory frames (25×10^6 for the 50 ns sampling phases of the present simulations with configurations stored every 2 fs) and p_k is the statistical (unbiasing) weight associated with frame k, defined as

$$p_k = \left[\sum_{l=1}^{N_f} \exp[\beta \mathscr{U}_{\text{bias},l}]\right]^{-1} \exp[\beta \mathscr{U}_{\text{bias},k}] \tag{9}$$

where $\beta = (k_B T)^{-1}$, k_B being Boltzmann's constant and T the absolute temperature (300 K), and $\mathscr{U}_{\text{bias},k}$ is the value of the biasing potential associated with trajectory frame k. The limiting case of an unbiased simulation corresponds to $p_k = N_f^{-1}$ for all frames, leading to $F_n = 1$ (maximum statistical efficiency). The limiting case of a biased simulation where a single frame k entirely dominates the reweighted probability distribution corresponds to $p_l = 0$ for $l \neq k$ along with $p_k = 1$, leading to $F_n = N_f^{-1}$ (minimum statistical efficiency, very close to zero). The quantity $F_n N_f$ can thus be viewed as an effective number of frames of the biased trajectory contributing to the statistics in terms of the properties of the unbiased ensemble. Although one usually has $F_n \ll 1$ in a biased simulation, the sampling efficiency may still be greatly enhanced in practice when this effective number of frames spans a much wider (i.e., more representative) volume of the configuration space accessible to the unbiased system, that is, when the searching efficiency has been increased. The significance of the statistical efficiency factor F_n is discussed in more detail in Appendix C. Note that a number of other measures for the statistical efficiency have been proposed previously.[25,84−87]

Considering the above discussion, the opposing effects of an enhancement of the searching efficiency and a deterioration of the statistical efficiency upon applying a biasing potential can be characterized by means of a combined sampling efficiency parameter $C_n(t_0)$ associated to a given time scale t_0, defined as

$$C_n(t_0) = \tilde{\tau}_n^{-1} \frac{1 - \exp[-(\tau_n^{-1}t_0)^{\alpha_n}]}{1 - \exp[-(\tau_n^{-1}t_0)]} F_n \tag{10}$$

The first factor accounts for the rate at which successive conformations (including newly discovered and revisited ones) are produced (eq 7). The second factor accounts for the tendency to revisit known conformations on a time scale t_0, compared to a purely random configuration generation process (eq 6). The third factor accounts for the statistical efficiency, that is, the relevance of the generated configurations in terms of the physical ensemble (eq 8). In practice, because a sampling enhancement or deterioration is always defined by reference to plain MD simulation for a given system, the combined sampling efficiency parameter will be reported as $\tilde{C}_n(t_0)$, defined as the quotient of $C_n(t_0)$ for a specific biasing scheme to the corresponding value for the unbiased simulation.

For practical applications, a more directly relevant point to be addressed concerns the reliability with which the low free energy conformations and associated free energies can be determined using a given sampling scheme. To this purpose, the lowest free energy conformations predicted by a given sampling scheme were also analyzed in terms of (i) convergence with time, (ii) number of interconversion transitions, and (iii) mutual overlap across schemes.

As a measure for the convergence of the calculated free energies over time, a convergence score $S_n(t)$ was defined as

$$S_n(t) = \left\{ \sum_{m=1}^{N_n^{\max}} [\exp[-\beta G_m(t)] - \exp[-\beta G_m(t_s)]]^2 \right\}^{1/2} \quad (11)$$

where t_s is a reference total simulation time (set to 50 ns for all simulations), and $G_m(t)$ is the free energy predicted for backbone conformation m based on the simulation time t, defined as

$$G_m(t) = -\beta^{-1} \ln\left\{ \left[\sum_{l=1}^{N_f'(t)} \exp[\beta \mathscr{U}_{\text{bias},l}] \right]^{-1} \sum_{k=1,k\in\mathscr{T}_m}^{N_f'(t)} \exp[\beta \mathscr{U}_{\text{bias},k}] \right\} \quad (12)$$

where $N_f'(t)$ denotes the number of trajectory frames generated up to time t and $\mathscr{T}_m$ the subset of these trajectory frames assigned to conformation m. Note that any conformation that has not been sampled at time t or at time t_s is characterized by an infinite free energy and leads to a corresponding exponential factor of zero in eq 11. The score $S_n(t)$ accounts for the root-sum-square error in the predicted conformational populations for a simulation of duration t compared with the full simulation of duration t_s. Because population differences will be largest for the conformations that are most populated at time t or at time t_s, this score gives more weight to unconverged free-energy estimates within the lowest free energy conformations. The convergence score will typically evolve from a value somewhat below one for $t \approx 0$ to exactly zero at $t = t_s$ and indicates how well the predicted lowest free energy conformations and associated free energies are converged at time t.

As a measure for the number of interconversion transitions between the lowest free energy conformations, an average number of (direct or indirect) transitions $T_n(G_c)$ was monitored, considering the $L_n(G_c)$ conformations within a free-energy cutoff G_c of the lowest free energy conformation discovered. More precisely, after evaluation of the relative free energies of the different conformations based on the entire simulation (after reweighting), the set of $L_n(G_c)$ lowest free energy conformations was identified. The number of transitions occurring along the trajectory between any two distinct conformations of this set (possibly via other conformations not included in the set) was then calculated and divided by $L_n(G_c)$, leading to the average number of transitions per conformation $T_n(G_c)$. This number provides an indication concerning the expected accuracy with which the relative free energies of the low free energy conformations are estimated, because the relative populations of two states are only likely to be representative of an equilibrium situation if interconversions are frequent on the simulation time scale. The calculation of $T_n(G_c)$ was performed separately for $G_c = 5$, 10, or 15 kJ mol^{-1} based on 50 ns simulations (first 50 ns for U).

Finally, as a measure for the (dis)agreement between the lowest free energy conformations generated using different schemes, overlap measures $O_n^{AB}(G_c)$ and $O_n^{BA}(G_c)$ were monitored for all pairs of simulations (A and B) involving different biasing potentials, considering the corresponding sets of $L_n^A(G_c)$ and $L_n^B(G_c)$ conformations within a free-energy cutoff G_c of the lowest free energy conformation discovered. The overlaps $O_n^{AB}(G_c)$ and $O_n^{BA}(G_c)$ are defined as the number of common configurations of the two sets, divided by $L_n^A(G_c)$ or $L_n^B(G_c)$, respectively. These overlap measures are comprised between 0 (no overlap) and 1 ($O_n^{AB}(G_c)$, set B encompasses all conformations of set A; $O_n^{BA}(G_c)$, set A encompasses all conformations of set B; both, full overlap). The calculation was performed separately for $G_c = 5$, 10, or 15 kJ mol^{-1} based on 50 ns simulations (first 50 ns for U).

Finally, to further illustrate the extent of overlap between the different simulations, a set of five consensus lowest free energy conformations was identified, considering simultaneously the seven 50 ns simulations (first 50 ns for U) performed for each of the eight systems (Ala$_n$ or Val$_n$ with $n = 4$, 6, 8, or 10). For a given system, each backbone conformation m was attributed a consensus rank R_m defined by

$$R_m = \sum_{i=1}^{7} (N_i - R_{m,i}) \exp[-\beta G_{m,i}] \quad (13)$$

where N_i is the number of conformations visited in simulation i, $R_{m,i}$ is the free-energy rank of conformation m in this simulation, and $G_{m,i}$ the corresponding free energy (the latter two measured relative to the minimum free energy conformation predicted by this simulation). The five configurations with the lowest consensus rank were then identified, and their ranks $R_{m,i}$ and free energies $G_{m,i}$ in the seven individual simulations are reported.

3. Results

The curves representing the cumulative number $N_n(t)$ of unique peptide backbone conformations discovered up to

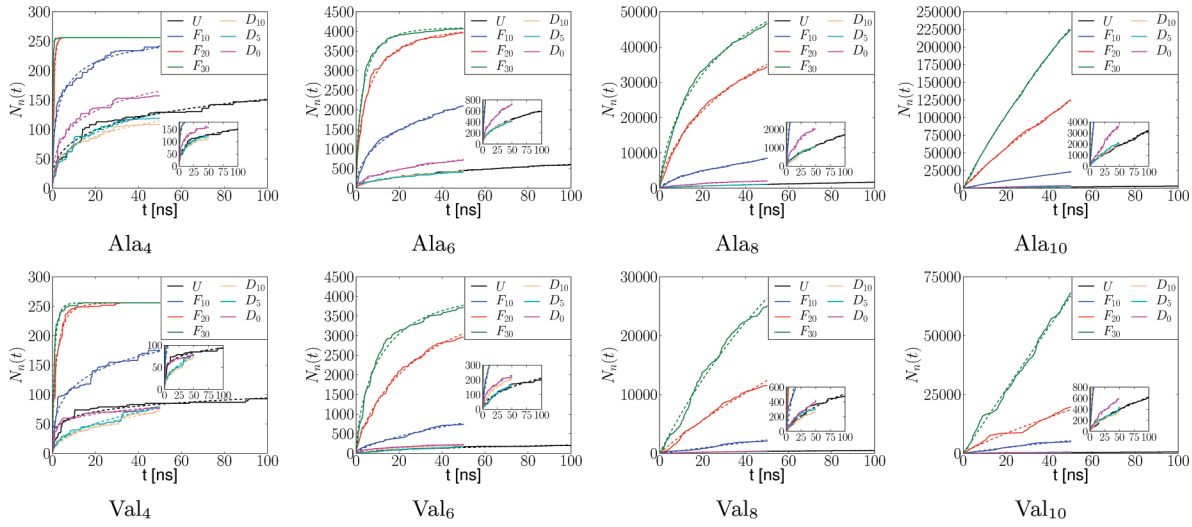

Figure 4. Number, $N_n(t)$, of unique peptide backbone conformations visited as a function of the sampling time t for Ala$_n$ and Val$_n$ oligopeptide simulations performed with different FB-LEUS biasing potentials. The biasing potentials considered are illustrated in Figure 3. The $N_n(t)$ curves are displayed using solid lines. Stretched-exponential fits (eqs 6 and 7) are displayed using dashed lines, the corresponding parameters α_n, τ_n and $\tilde{\tau}_n$ being reported in Table 1 and displayed graphically in Figure 5a,b (α_n,$\tilde{\tau}_n$).

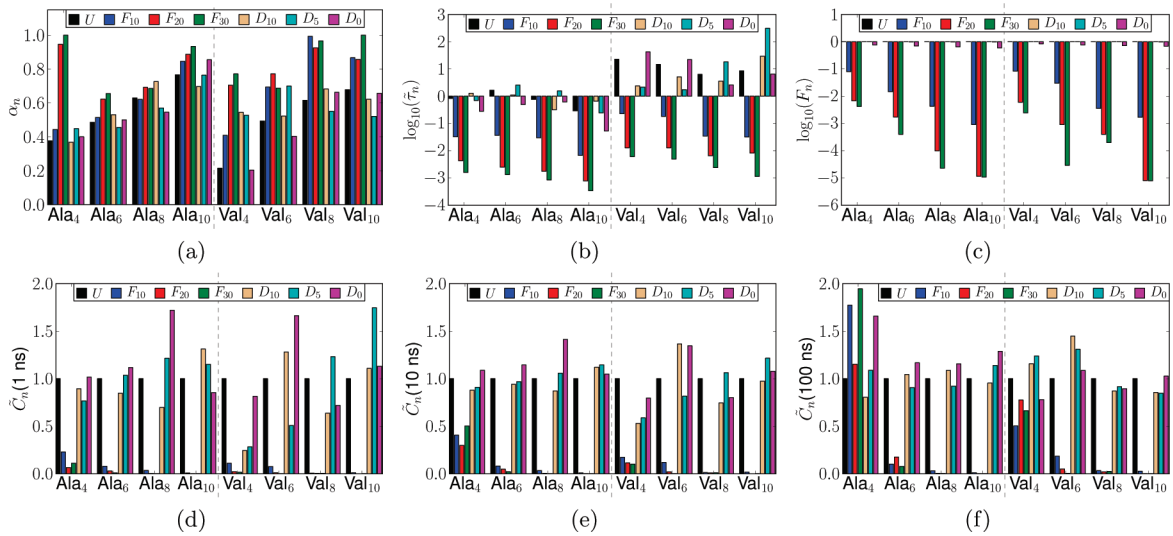

Figure 5. Stretching exponent α_n, effective visiting time $\tilde{\tau}_n$, statistical efficiency F_n, and combined sampling efficiency $\tilde{C}_n(t_0)$ for Ala$_n$ and Val$_n$ oligopeptide simulations performed with different FB-LEUS biasing potentials. The biasing potentials considered are illustrated in Figure 3. The parameters α_n, $\tilde{\tau}_n$, F_n, and $\tilde{C}_n(t_0)$ are defined by eqs 6−10, where $\tilde{C}_n(t_0)$ is the ratio of $C_n(t_0)$ for a given simulation to the corresponding value for the unbiased simulation U, and reference times t_0 = 1, 10, or 100 ns are considered. Note that the quantities $\tilde{\tau}_n$ (in units of ns) and F_n are displayed on a (decimal) logarithmic scale. Corresponding numerical values can be found in Table 1.

time t during the sampling phase of the different Ala$_n$ and Val$_n$ oligopeptide simulations are displayed in Figure 4. In terms of searching efficiency, the application of the FB-LEUS biasing potentials may lead to a spectacular enhancement compared with the unbiased MD simulation U.

For all oligopeptides, this enhancement is most pronounced when using the BF biasing potentials, systematically increasing along the series F_{10}, F_{20}, and F_{30} of increasingly "flattened" monopeptide free-energy surfaces (Figure 3a,b). For Ala$_4$ and Val$_4$, the exhaustive-search upper bound $N_4^{\max}$ = 256 for $N_4(t)$ is reached within 2.8 (Ala$_4$) or 13.8 (Val$_4$) ns when the biasing potential F_{30} is used and within 4.4 (Ala$_4$) or 31.4 (Val$_4$) ns when the biasing potential F_{20} is used, while plain MD only visits 150 (Ala$_4$) or 93 (Val$_4$) conformations within 100 ns. For Ala$_6$, the corresponding upper bound $N_6^{\max}$ = 4096 for $N_6(t)$ is also reached close to 50 ns when using the biasing potential F_{30}, while plain MD only visits 598 conformations within 100 ns. Considering all systems, the numbers of conformations that have been visited within 50 ns is increased by factors of 1.9−12.7 (F_{10}), 2.0−65.6 (F_{20}), and 2.0−172.9 (F_{30}) compared to the unbiased simulation U, these factors systematically increasing with the oligopeptide length n and being tendentially larger for Val$_n$ compared with Ala$_n$ at a given n.

The searching enhancement is much less pronounced when the VD biasing potentials are used, tendentially increasing along the series D_{10}, D_5, and D_0 of increasingly "dug" valleys on the monopeptide free-energy surfaces (Figure 3c,d).

Table 1. Stretching Exponent α_n, Characteristic Searching Time τ_n, Effective Visiting Time $\tilde{\tau}_n$, Statistical Efficiency F_n, and Combined Sampling Efficiency $\tilde{C}_n(t_0)$ for Ala_n and Val_n Oligopeptide Simulations Performed with Different FB-LEUS Biasing Potentials[a]

bias	α_n	τ_n [ns]	$\tilde{\tau}_n$ [ns]	F_n	$\tilde{C}_n$ (1)	$\tilde{C}_n$ (10)	$\tilde{C}_n$ (100)	bias	α_n	τ_n [ns]	$\tilde{\tau}_n$ [ns]	F_n	$\tilde{C}_n$ (1)	$\tilde{C}_n$ (10)	$\tilde{C}_n$ (100)
				Ala_4								Val_4			
U	0.38	1.3×10^2	8.3×10^{-1}	1.00×10^0	1.0000	1.0000	1.0000	U	0.22	3.7×10^3	2.3×10^1	1.00×10^0	1.0000	1.0000	1.0000
F_{10}	0.44	5.2×10^0	3.2×10^{-2}	7.99×10^{-2}	0.2288	0.4057	1.7728	F_{10}	0.41	3.7×10^1	2.3×10^{-1}	8.33×10^{-2}	0.1101	0.1726	0.5035
F_{20}	0.95	7.0×10^{-1}	4.3×10^{-3}	6.73×10^{-3}	0.0655	0.2976	1.1525	F_{20}	0.70	2.1×10^0	1.3×10^{-2}	6.01×10^{-3}	0.0218	0.1138	0.7763
F_{30}	1.00	2.6×10^{-1}	1.6×10^{-3}	4.23×10^{-3}	0.1116	0.5024	1.9454	F_{30}	0.77	9.9×10^{-1}	6.1×10^{-3}	2.44×10^{-3}	0.0158	0.1011	0.6634
D_{10}	0.37	2.1×10^2	1.3×10^0	1.00×10^0	0.8943	0.8796	0.8040	D_{10}	0.54	3.9×10^2	2.4×10^0	1.00×10^0	0.2444	0.5295	1.1568
D_5	0.45	1.1×10^2	7.0×10^{-1}	9.89×10^{-1}	0.7656	0.9092	1.0881	D_5	0.53	3.5×10^2	2.2×10^0	9.95×10^{-1}	0.2831	0.5878	1.2388
D_0	0.40	4.6×10^1	2.8×10^{-1}	7.55×10^{-1}	1.0175	1.0882	1.6594	D_0	0.20	6.8×10^3	4.2×10^1	8.38×10^{-1}	0.8149	0.7969	0.7791
				Ala_6								Val_6			
U	0.49	4.3×10^3	1.7×10^0	1.00×10^0	1.0000	1.0000	1.0000	U	0.49	3.8×10^4	1.5×10^1	1.00×10^0	1.0000	1.0000	1.0000
F_{10}	0.51	9.5×10^1	3.7×10^{-2}	1.47×10^{-2}	0.0792	0.0812	0.1014	F_{10}	0.69	4.7×10^2	1.8×10^{-1}	2.99×10^{-2}	0.0761	0.1191	0.1847
F_{20}	0.62	6.5×10^0	2.5×10^{-3}	1.72×10^{-3}	0.0292	0.0481	0.1757	F_{20}	0.77	3.3×10^1	1.3×10^{-2}	9.15×10^{-4}	0.0110	0.0205	0.0505
F_{30}	0.66	3.4×10^0	1.3×10^{-3}	3.98×10^{-4}	0.0097	0.0207	0.0769	F_{30}	0.69	1.3×10^1	4.9×10^{-3}	2.92×10^{-5}	0.0009	0.0014	0.0043
D_{10}	0.53	2.9×10^3	1.1×10^0	1.00×10^0	0.8470	0.9407	1.0443	D_{10}	0.52	1.3×10^4	5.1×10^0	1.00×10^0	1.2808	1.3647	1.4474
D_5	0.46	6.6×10^3	2.6×10^0	9.86×10^{-1}	1.0360	0.9678	0.9058	D_5	0.70	4.5×10^3	1.7×10^0	9.93×10^{-1}	0.5069	0.8157	1.3100
D_0	0.50	1.3×10^3	4.9×10^{-1}	6.96×10^{-1}	1.1170	1.1450	1.1678	D_0	0.40	5.7×10^4	2.2×10^1	7.59×10^{-1}	1.6628	1.3459	1.0870
				Ala_8								Val_8			
U	0.63	3.2×10^4	7.6×10^{-1}	1.00×10^0	1.0000	1.0000	1.0000	U	0.62	2.6×10^5	6.3×10^0	1.00×10^0	1.0000	1.0000	1.0000
F_{10}	0.62	1.2×10^3	3.0×10^{-2}	4.30×10^{-3}	0.0352	0.0340	0.0322	F_{10}	0.99	1.4×10^3	3.4×10^{-2}	3.59×10^{-3}	0.0057	0.0136	0.0326
F_{20}	0.69	7.3×10^1	1.8×10^{-3}	9.83×10^{-5}	0.0034	0.0038	0.0049	F_{20}	0.93	2.7×10^2	6.5×10^{-3}	3.98×10^{-4}	0.0048	0.0098	0.0200
F_{30}	0.69	3.5×10^1	8.5×10^{-4}	2.29×10^{-5}	0.0013	0.0015	0.0023	F_{30}	0.97	9.9×10^1	2.4×10^{-3}	1.99×10^{-4}	0.0051	0.0113	0.0256
D_{10}	0.73	1.3×10^4	3.2×10^{-1}	1.00×10^0	0.6976	0.8709	1.0872	D_{10}	0.68	1.5×10^5	3.6×10^0	1.00×10^0	0.6379	0.7448	0.8699
D_5	0.57	6.5×10^4	1.6×10^0	9.84×10^{-1}	1.2146	1.0578	0.9217	D_5	0.55	7.5×10^5	1.8×10^1	9.90×10^{-1}	1.2326	1.0629	0.9168
D_0	0.55	2.6×10^4	6.2×10^{-1}	6.40×10^{-1}	1.7197	1.4129	1.1560	D_0	0.66	1.1×10^5	2.6×10^0	7.20×10^{-1}	0.7190	0.8018	0.8937
				Ala_{10}								Val_{10}			
U	0.77	1.9×10^5	2.9×10^{-1}	1.00×10^0	1.0000	1.0000	1.0000	U	0.68	5.7×10^6	8.6×10^0	1.00×10^0	1.0000	1.0000	1.0000
F_{10}	0.85	4.5×10^3	6.7×10^{-3}	9.17×10^{-4}	0.0084	0.0101	0.0120	F_{10}	0.87	2.1×10^4	3.2×10^{-2}	1.69×10^{-3}	0.0115	0.0177	0.0273
F_{20}	0.89	5.2×10^2	7.8×10^{-4}	1.16×10^{-5}	0.0005	0.0007	0.0009	F_{20}	0.86	5.5×10^3	8.2×10^{-3}	8.02×10^{-6}	0.0002	0.0003	0.0004
F_{30}	0.93	2.3×10^2	3.4×10^{-4}	1.09×10^{-5}	0.0008	0.0011	0.0017	F_{30}	1.00	7.6×10^2	1.1×10^{-3}	7.92×10^{-6}	0.0004	0.0008	0.0018
D_{10}	0.70	4.4×10^5	6.6×10^{-1}	1.00×10^0	1.3129	1.1197	0.9550	D_{10}	0.62	1.9×10^7	2.9×10^1	1.00×10^0	1.1080	0.9739	0.8561
D_5	0.76	1.6×10^5	2.4×10^{-1}	9.79×10^{-1}	1.1511	1.1452	1.1391	D_5	0.52	2.1×10^8	3.1×10^2	9.86×10^{-1}	1.7464	1.2158	0.8465
D_0	0.86	3.5×10^4	5.3×10^{-2}	5.90×10^{-1}	0.8537	1.0485	1.2871	D_0	0.66	4.3×10^6	6.5×10^0	6.81×10^{-1}	1.1313	1.0779	1.0269

[a] The biasing potentials considered are illustrated in Figure 3. The parameters α_n, $\tilde{\tau}_n$, F_n, and $\tilde{C}_n(t_0)$ are defined by eqs 6−10, where $\tilde{C}_n(t_0)$ is the ratio of $C_n(t_0)$ for a given simulation to the corresponding value for the unbiased simulation U, and reference times $t_0 = 1$, 10, or 100 ns are considered. The results are also illustrated graphically in Figure 5.

However, in most simulations, the effects of the biasing potentials D_{10} and D_5 are comparable, and the corresponding enhancement is marginal (for Ala_4 and Val_4, the searching is even slightly slower than that of plain MD). Only the biasing potential D_0 enhances the searching nearly systematically (with the possible exception of Val_4). Considering all systems, the numbers of conformations that have been visited within 50 ns changes by factors of 0.7−1.2 (D_{10}), 0.8−1.1 (D_5), and 0.9−1.9 (D_0) compared to the unbiased simulation U. For the biasing potential D_0, this factor tendentially increases with the oligopeptide length n and is typically larger for Ala_n compared with Val_n at a given n.

The different $N_n(t)$ curves of Figure 4 could be adequately fitted to stretched-exponential functions (eqs 6 and 7). The values of the resulting parameters α_n (stretching exponent), τ_n (characteristic searching time), and $\tilde{\tau}_n$ (effective visiting time) for the different Ala_n and Val_n oligopeptide simulations are reported in Table 1 and displayed graphically in Figure 5a,b. The α_n values range from 0.22 (high revisiting tendency) to 1.00 (random configuration generation process), the τ_n values span 9 orders of magnitude from 0.26 ns to 0.21 s, and the $\tilde{\tau}_n$ values span 6 orders of magnitude from 0.34 ps to 0.31 μs for the different systems and biasing potentials considered.

The α_n values associated with the unbiased MD simulations U range from 0.22 to 0.77. They systematically increase with n for a given oligopeptide type and are systematically higher for Ala_n compared with Val_n at a given n. This indicates a high revisiting tendency, expectedly more pronounced for systems of lower dimensionality (low n, implying a higher probability to diffuse back into previously visited conformations) and systems involving stronger nonlocal interactions (more important hydrophobic side chain−side chain interactions in Val_n compared with Ala_n, inducing a more significant conformational probability bias). The τ_n values for the unbiased simulations increase roughly exponentially with the oligopeptide length n, approximately by one order of magnitude upon increasing n by two for both Ala_n and Val_n. The values are also systematically higher by about one order of magnitude for Val_n compared with Ala_n. Thus, for example, a coverage of a fraction $1 - e^{-1}$ (63%) of the entire conformational space accessible to the Ala_{10} and Val_{10} oligopeptides can be estimated to require plain MD simulations on the 0.2 and 6 ms time scales, respectively. These estimates should probably be regarded as lower bounds, because they rely on α_n and τ_n values characteristic of the (low free energy) regions visited on the 50 ns time scale. In reality, the conformational probability bias is likely to increase when the sampling is extended to other (higher free energy) regions, probably resulting in a decrease of the α_n value and an increase of the τ_n value (or even a breakdown of the stretched-exponential fitting). Note also that these time scales are far above the suggested time scales for secondary-structure formation in polypeptides (section 1), indicating that these processes are by no means random search processes and require the dynamical sampling of a much smaller volume fraction of conformational space (e.g., compared with $1 - e^{-1}$) for their occurrence (Levinthal's paradox[88]). Finally, the $\tilde{\tau}_n$ values for the unbiased simulations evidence much smaller variations (compared to τ_n) across the different systems considered, ranging from 0.29 to 1.65 ns for Ala_n and from 6.32 to 22.8 ns for Val_n. These times nearly systematically decrease with n for a given oligopeptide type and are about one order of magnitude higher for Val_n compared with Ala_n at a given n. This suggests effective times associated with backbone torsional angle transitions (between wells as defined in section 2.3) on the order of 1 and 10 ns for the two types of peptides. The effective times $\tilde{\tau}_n$ are expectedly lower for systems of higher dimensionality (high n, because a conformational transition results from a transition in any of the n linkages) and higher for systems involving more important transition barriers (more bulky side chains in Val_n compared with Ala_n, inducing higher interconversion barriers; Figure 2).

The α_n values for the simulations with the BF biasing potentials range from 0.41 to 1.00. For a given oligopeptide, they nearly systematically increase along the series F_{10}, F_{20}, and F_{30} and are nearly systematically higher than the corresponding value for the unbiased simulation U. Here also, the values tendentially increase with n for a given oligopeptide type (the comparison of the Ala_n to the corresponding Val_n oligopeptide does not reveal systematic trends). For some systems (Ala_4, Ala_{10}, Val_8, and Val_{10}), α_n may become very close to one, suggesting that the FB-LEUS scheme leads in this case to a trajectory that is very similar to a random configuration generation process. Simultaneously, the τ_n and $\tilde{\tau}_n$ values are dramatically decreased compared to the plain MD simulation, by factors of about 25−250 (F_{10}), 200−1700 (F_{20}), or 500−7500 (F_{30}) for the different systems. For example, the coverage of a fraction $1 - e^{-1}$ (63%) of the entire conformational space accessible to the Ala_{10} or Val_{10} oligopeptides using the F_{30} biasing potential can be estimated to require simulations on the 200 and 800 ns time scales, respectively. These estimates are probably more realistic than the corresponding estimates (0.2 and 6 ms, respectively) for the unbiased simulation U (see above), considering that the F_{30} biasing potential largely reduces the conformational probability bias in the simulated ensemble (at least the local single-linkage component of this bias). The $\tilde{\tau}_n$ values are in the ranges 6.7−32 ps (F_{10}), 0.78−4.3 ps (F_{20}), and 0.34−1.6 ps (F_{30}) for Ala_n, and 32−230 ps (F_{10}), 6.5−13 ps (F_{20}), and 1.1−6.1 ps (F_{30}) for Val_n. Here also, for each of the three biasing potentials, these times decrease with n for a given oligopeptide type and are shorter for Ala_n compared with Val_n at a given n. The effective time associated with backbone torsional angle transitions has thus been brought from the 1−10 ns range for the unbiased simulations to the 1−100 ps range for the biased simulations.

The α_n values for the simulations with the VD biasing potentials range from 0.20 to 0.86. For a given oligopeptide, the values for the series D_{10}, D_5, and D_0 are generally similar, and comparable to the corresponding value for the unbiased simulation U. Here also, the values tendentially increase with n for a given oligopeptide type (the comparison of the Ala_n to the corresponding Val_n oligopeptide does not reveal systematic trends). Similarly, the τ_n and $\tilde{\tau}_n$ values are only moderately altered and in a nonsystematic way compared to the plain MD simulation. Note, however, that the values for

the biasing potential D_0 are nearly systematically lower than those for the unbiased simulation (except for Val_4 and Val_6), in agreement with the searching enhancement generally observed for this scheme (Figure 4).

The results of the simulations in terms of the statistical efficiency parameter F_n (eq 8) for the different Ala_n and Val_n oligopeptide simulations are reported in Table 1 and displayed graphically in Figure 5c. In terms of statistical efficiency, the application of the FB-LEUS biasing potentials may lead to a dramatic deterioration compared to the unbiased MD simulation U.

Because plain (thermostatted) MD samples in the canonical ensemble, the corresponding value of F_n is one for all n (maximal statistical efficiency). The value corresponding to a simulation where a single configuration entirely dominates the reweighted probability distribution is $F_n = N_f^{-1} = 4 \times 10^{-8} \approx 0$ (minimal statistical efficiency). For the different oligopeptides considered, the values of F_n for the biased simulations range from 8.0×10^{-6} to 1.0. Values lower than one indicate a tendency to generate irrelevant configurations in terms of the conformational distribution characteristic of the physical system, that is, high-energy configurations with low Boltzmann weights in the physical ensemble.

The deterioration of the statistical efficiency is most pronounced when using the BF biasing potentials, the efficiency systematically decreasing along the series F_{10}, F_{20}, and F_{30} of increasingly "flattened" monopeptide free-energy surfaces (Figure 3a,b). For example, considering Ala_4 with the biasing potential F_{30}, only about 0.4% of the trajectory configurations actually contribute to the statistics relevant for the physical ensemble, the corresponding fraction becoming as low as 0.001% for Ala_{10}. The deterioration is much less important when using the VD biasing potentials, the efficiency systematically decreasing along the series D_{10}, D_5, and D_0 of increasingly "dug" valleys on the monopeptide free-energy surfaces (Figure 3c,d). For example, considering all systems, at least 97% of the trajectory configurations contribute to the relevant statistics for the biasing potentials D_{10} and D_5, while this fraction is still at least 59% for the biasing potential D_0. This large qualitative difference between the BF and VD biasing potentials is easily understood on the basis of the very different extents of irrelevant conformational regions they open up to sampling (large basins vs narrow valleys; Figure 3). The trends observed within the two series of potentials are easily rationalized using similar arguments.

The value of F_n systematically decreases with n for a given oligopeptide and biasing potential type (the comparison of the Ala_n to the corresponding Val_n oligopeptide reveals comparable values, but no further systematic trends). This decrease results from two main effects. First, distinguishing between relevant and irrelevant conformational regions at the monopeptide level (in terms of the distribution appropriate for the physical system), the probability that all peptide bonds are simultaneously found in relevant regions at the oligopeptide level decreases exponentially with n (combinatorial effect). Second, the extension of the peptide length increases the likelihood that the location of the free-energy minima at the oligopeptide level differ from the corresponding locations at the monopeptide level, as a consequense nonlocal interactions (i.e., beyond the local conformational preferences of the individual linkages), resulting in a tendency of the FB-LEUS biasing potential to drive the system toward conformational regions that do not correspond to free-energy basins at the oligopeptide level.

As illustrated in Figure 6, the combinatorial effect is dominant in the context of the BF biasing potentials. In this case, F_n decreases as a single exponential function of n (no prefactor), that is, as $F_n \approx 10^{-an}$, where the value of a is nearly identical for Ala_n and Val_n, namely, about -0.29, -0.49, and -0.54 for F_{10}, F_{20}, and F_{30}, respectively. The existence of such a relationship suggests in particular that a good estimate for F_n at the oligopeptide level can already be formulated from the knowledge of the statistical efficiency F_1 at the monopeptide level. Note also that a does not decrease linearly with the free-energy cutoff h used in the definition of the BF potential (section 2.2 and Figure 3a). The dependence of a on h is expected to level off at a value h_{max} equal to the minimum-to-maximum difference in the fragment free-energy map. For any h above h_{max}, the most statistically inefficient BF potential has been reached (entirely homogeneous sampling of the relevant conformational subspace at the fragment level). Considering Figure 3b, this limit is essentially reached for the biasing potential F_{30} (at least for poly-Ala) and one has thus $F_n \approx 10^{-0.54n}$ for any $h \geq 30$ kJ mol^{-1}.

The opposing effects of the enhancement in the searching efficiency and of the deterioration in the statistical efficiency can be characterized by means of the combined sampling efficiency parameter $C_n(t_0)$ associated with a given time scale t_0 (eq 10). The results of the different Ala_n and Val_n oligopeptide simulations in terms of the corresponding relative values $\tilde{C}_n(t_0)$, with unbiased simulations U taken as reference, are reported in Table 1 and displayed graphically in Figure 5d,e,f for $t_0 = 1$, 10, or 100 ns. The values of $\tilde{C}_n(t_0)$ for the unbiased simulations are all equal to one by definition. Most of the biased simulations present values lower than one, indicating a decrease of the combined sampling efficiency compared with plain MD. Exceptions involve the BF biasing potentials for Ala_4 and $t_0 = 100$ ns, as well as the VD biasing potentials in about half of the considered systems and t_0 combinations. However, even in these cases, the sampling enhancement never exceeds a factor two.

For the simulations relying on BF biasing potentials, $\tilde{C}_n(t_0)$ tendentially decreases upon comparing Val_n to Ala_n, along the series F_{10}, F_{20}, and F_{30}, upon increasing n, or upon decreasing t_0. The results suggest that the BF scheme may become competitive with plain MD for the systems considered when the simulation time scale t_0 is of the same order as the effective searching time τ_n of the unbiased simulation (e.g., $t_0 = 100$ ns compared with $\tau_n = 134$ ns for Ala_4, where the BF schemes are more efficient than plain MD). This is of course a disappointing conclusion since this time scale is precisely the one beyond which a sampling enhancement is in principle no longer needed.

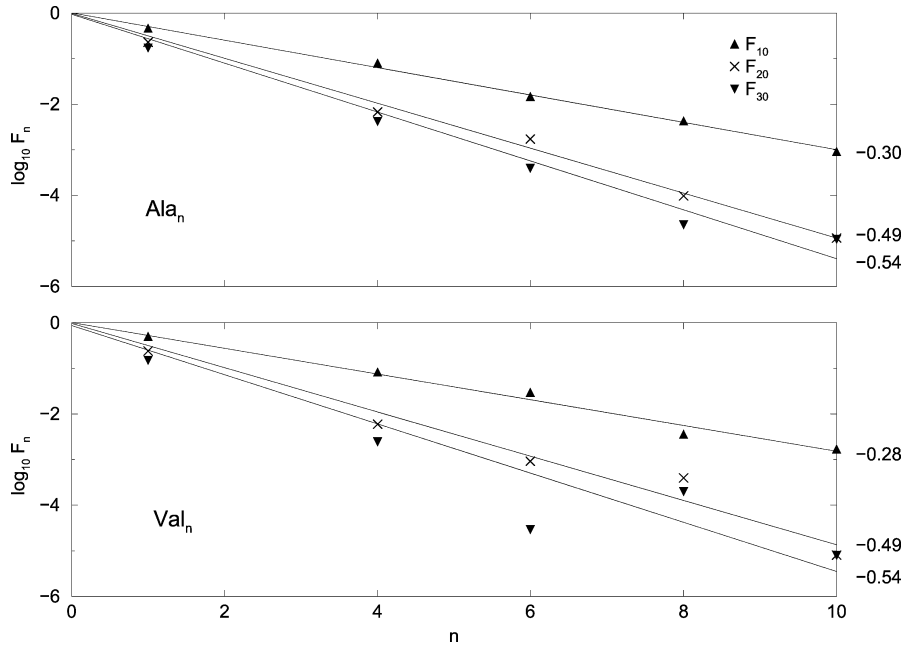

Figure 6. Statistical efficiency F_n for Ala_n and Val_n oligopeptide simulations performed with different FB-LEUS biasing potentials of the BF type. The BF biasing potentials considered are illustrated in Figure 3a,b. The parameter F_n is defined by eq 8. It is displayed on a (decimal) logarithmic scale, along with least-squares-fit regression lines anchored at the origin (single exponential fit, no prefactor). The corresponding slopes are also indicated. Corresponding numerical values can be found in Table 1 except for $n = 1$ (**$Ala_1 = 0.47$, 0.23, 0.17, and $Val_1 = 0.51$, 0.24, 0.15, for biasing potentials F_{10}, F_{20} and, F_{30}, based on the N&C-blocked monopeptides).

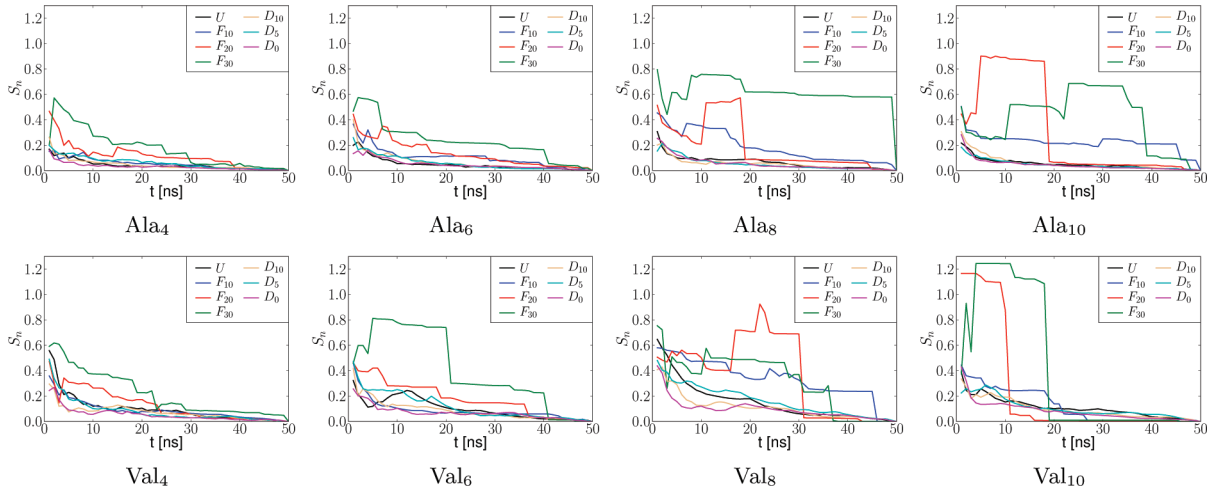

Figure 7. Free-energy convergence score $S_n(t)$ for Ala_n and Val_n oligopeptide simulations performed with different FB-LEUS biasing potentials. The biasing potentials considered are illustrated in Figure 3. The score $S_n(t)$ is defined by eq 11.

For the simulations relying on the VD biasing potentials, $\tilde{C}_n(t_0)$ is typically close to one, the dependence on the type of peptide and choice of biasing potential, as well as on n and t_0, being rather nonsystematic. Note, however, that $\tilde{C}_n(t_0)$ is nearly always higher than one when the biasing potential D_0 is used (except for Val_4 and Val_8, as well as Ala_{10} with $t_0 = 1$ ns), with values ranging from 0.72 to 1.72. In other words, this biasing potential generally leads to a modest sampling enhancement at all time scales considered.

The above results clearly illustrate the problem of the trade-off between searching and statistical efficiencies as determinants of the overall sampling efficiency of a scheme. For practical applications, a more directly relevant point to be addressed concerns the reliability with which the low free energy conformations and associated free energies can be determined using a given sampling scheme (convergence with time, number of interconversion transitions, and mutual overlap across schemes). These properties are investigated in turn below.

The results of the simulations in terms of the score $S_n(t)$ measuring the free-energy convergence (eq 11) for the different Ala_n and Val_n oligopeptide simulations are displayed in Figure 7. The score $S_n(t)$ accounts for the root-sum-square error in the predicted conformational populations for a simulation of duration t compared with the full simulation of duration t_s (50 ns). All the $S_n(t)$ curves start somewhat below one for $t \approx 0$ and converge to exactly zero at $t = t_s$. However, the rapidity as well as

the degree of monotonicity and smoothness of this evolution provides information concerning the convergence of the predicted lowest free energy conformations and of the associated free energies. For example, a rapid variation at a given time t suggests that the sudden visit to a new low free energy conformation has radically altered the population distribution within the most populated states of the (reweighted) ensemble, that is, that the distribution just before time t was unconverged. The curves corresponding to the unbiased simulations are rapidly converging as well as essentially smooth and monotonic. The same applies to the curves corresponding to the simulations relying on the VD biasing potentials. A convergence improvement is visible in some systems (e.g., Val_4 and Val_8) for some of these potentials. In contrast, the curves corresponding to the simulations relying on the BF biasing potentials are erratic. This behavior becomes increasingly pronounced upon increasing the oligopeptide length n. In other words, the predicted most stable conformations and associated free energies are steadily reshuffled by the sporadic encounter of new low free energy conformations and certainly not converged after 50 ns of simulation (with the possible exception of Ala_4 and Val_4).

The results of the simulations in terms of the average number of (direct or indirect) interconversion transitions $T_n(G_c)$ between the lowest free energy conformations (section 2.3) for the different Ala_n and Val_n oligopeptide simulations are displayed in Figure 8, considering three free-energy cutoffs $G_c = 5$, 10, or 15 kJ mol^{-1}. These numbers are correlated with the accuracy with which the relative free energies of the lowest free energy conformations will be predicted by a given scheme, because a sufficient number of interconversion transitions is a prerequisite for the evaluation of an accurate free-energy difference. For the cutoff values considered, the average number of transitions to each low free energy conformation is on the order of 10^2-10^4 for the unbiased simulations and the simulations relying on the VD biasing potentials. The $T_n(G_c)$ values tendentially decrease upon increasing the oligopeptide length n and upon increasing G_c, and are systematically slightly higher for the simulations employing the biasing potential D_0 compared to the corresponding unbiased simulations (increase by factors 1.1−5.9 for the different systems and G_c values considered). In contrast, for the simulations relying on the BF biasing potentials, the $T_n(G_c)$ values decrease much more abruptly upon increasing the peptide length, as well as along the series F_{10}, F_{20}, and F_{30}. As a result, if the numbers of transitions are comparable to those of the unbiased and VD simulations for Ala_4 and Val_4, they may decrease to very few (or even a single one) for some of the Ala_{10} and Val_{10} simulations. For these simulations, the ranking of the low free energy conformations and the corresponding relative free-energy estimates are likely to be incorrect. A second observation is that, in contrast to the unbiased and VD simulations, the $T_n(G_c)$ values for the BF simulations sometimes increase upon increasing G_c. For the unbiased and VD simulations, a systematic decrease of $T_n(G_c)$ upon increasing G_c is expected, because it extends the averaging to less populated conformations involved in fewer transitions. For the BF simulations, nonsystematic changes result from the fact that, especially for high n and low G_c, the number of states $L_n(G_c)$ below G_c is very low (poor statistics).

The results of the simulations in terms of the overlaps $O_n^{AB}(G_c)$ and $O_n^{BA}(G_c)$, between the lowest free energy conformations generated by all pairs (A and B) of schemes (section 2.3) for the different Ala_n and Val_n oligopeptide simulations are displayed in Figure 9, considering three free-energy cutoffs $G_c = 5$, 10, or 15 kJ mol^{-1}. Expectedly, the extent of overlap between the low free energy conformations predicted by the different schemes decreases upon increasing n. However, this decrease is much more pronounced for the schemes relying on the BF biasing potentials and, among these, in the sequence F_{10}, F_{20}, and F_{30}. For Ala_4 and Val_4, all schemes present a very high extent of mutual overlap. However, for Ala_{10} and Val_{10}, only simulations U, D_{10}, D_5, and D_0, and to a lesser extent F_{10}, present a reasonable extent of mutual overlap. In contrast, simulations F_{20} and F_{30} have produced two different sets of low free energy conformations, presenting essentially no overlap with the latter sets and with each other.

Similar observations can be made based on the ranking $R_{m,i}$ and free energies $G_{m,i}$ attributed by a given scheme i to the five conformations m with the lowest consensus rank R_m (section 2.3). The values of $R_{m,i}$ and $G_{m,i}$ for the different schemes are reported in Table 2. The conformations m associated with the lowest R_m are systematically found to be those with $m = 0$, corresponding (via their binary codes) to the n successive dihedral-angle pairs (ϕ_i, ψ_i) within the lowest free energy well at the monopeptide level (Figure 2). Bundles of 20 illustrative trajectory configurations assigned to these conformations are represented in Figure 10. The four other conformations with lowest consensus ranks differ from these ones by one or two bits only (with reference to the corresponding binary codes), commonly the first or last ones or both, indicating a different free-energy well at the monopeptide level for dihedral-angle pairs (ϕ_1, ψ_1) or (ϕ_n, ψ_n). This observation suggests that nonlocal interactions (beyond single-linkage conformational preferences) are weak in the oligopeptides considered (and given their relatively short lengths). This is confirmed by a DSSP analysis[89] (data not shown) of the corresponding configurations, which suggests a quasi-total absence of secondary structure (beyond turns and bends, that is, no helices or sheets). For Ala_4 and Val_4, nearly all schemes have predicted conformation 0 as the lowest free energy one (except simulation F_{30} for Val_4, for which this conformation has rank 2). Most simulations also encompass a significant fraction of the four other conformations among their lowest free energy conformations. For Ala_n and Val_n with $n = 6$, 8, or 10, the plain MD simulations and the simulations relying on the VD biasing potentials have all predicted conformation 0 as the lowest free energy one. However, the agreement concerning the ranking and relative free energies of the four other conformations progressively worsens upon increasing n. In contrast,

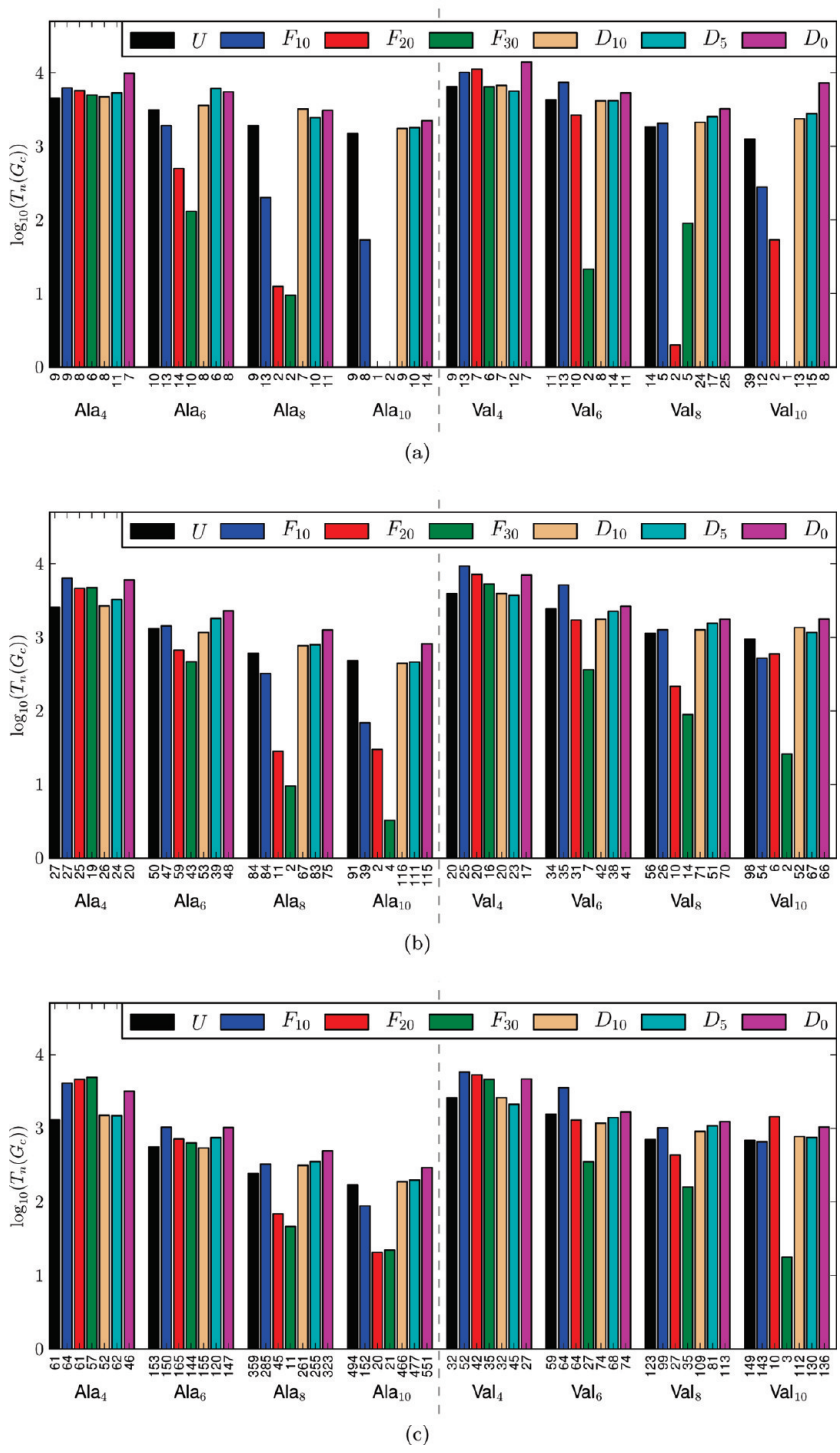

Figure 8. Average number of (direct or indirect) transitions $T_n(G_c)$ between the $L_n(G_c)$ lowest free energy conformations for Ala_n and Val_n simulations performed with different FB-LEUS biasing potentials. The biasing potentials considered are illustrated in Figure 3. The quantities $L_n(G_c)$, reported below the individual bars, and $T_n(G_c)$ are defined in section 2.3. The values are calculated using free-energy cutoffs $G_c = 5$, 10, or 15 kJ mol^{-1}. Note the use of a (decimal) logarithmic scale.

except for Ala_6 with biasing potentials F_{10} and F_{20}, the simulations relying on the BF biasing potentials fail to produce the same lowest free energy conformation and to agree with each other concerning this conformation beyond $n = 4$. In other words, the biased ensemble is in this case crowded with irrelevant configurations, leading to very high uncertainties in the calculated relative free energies for the few relevant (low free energy) conformations and, for longer peptides, to the complete omission of most of these conformations.

4. Conclusion

The goal of the present study was to expand the scope of the LEUS method to solvated polymers with more than a few relevant degrees of freedom, by means of a fragment-

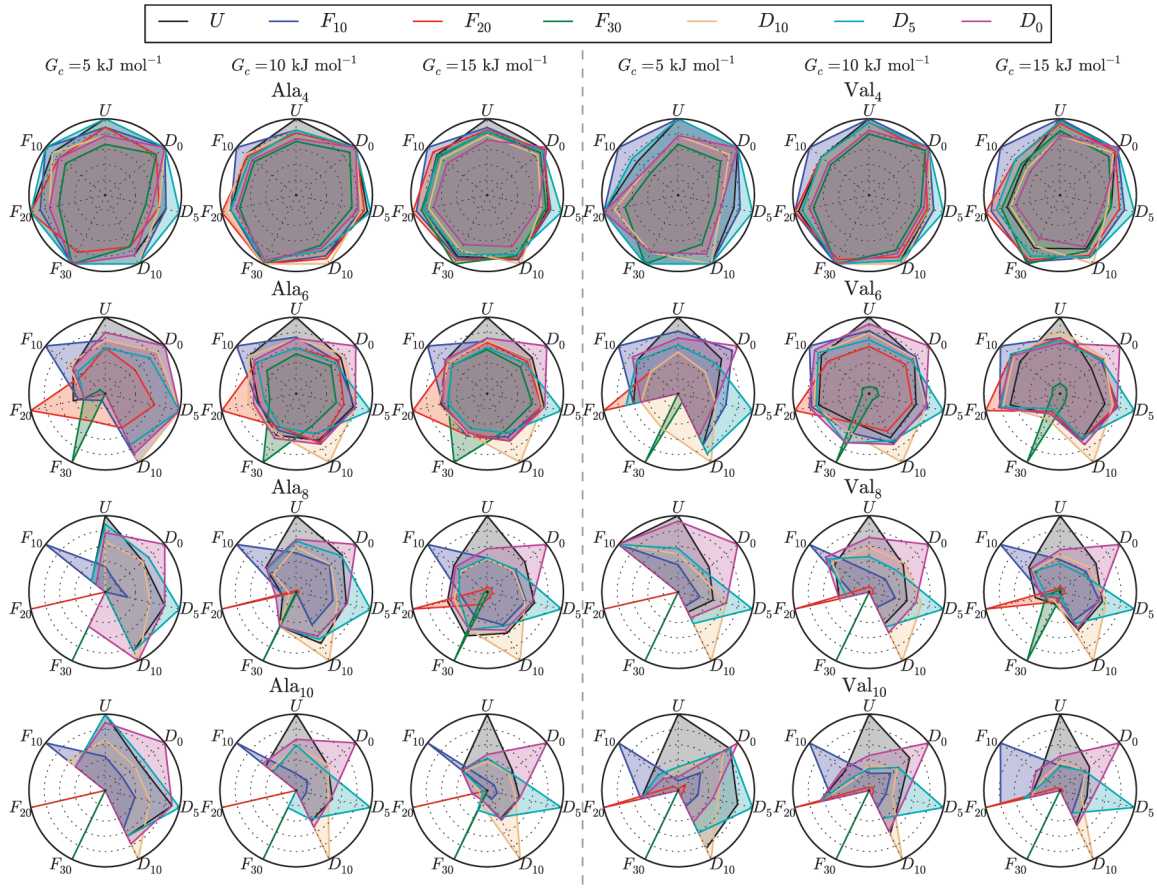

Figure 9. Overlaps $O_n^{AB}(G_c)$ and $O_n^{BA}(G_c)$ between the lowest free energy conformations predicted by pairs (*A* and *B*) of simulations of the Ala_n and Val_n oligopeptides performed using different FB-LEUS biasing potentials. The biasing potentials considered are illustrated in Figure 3. The quantities $O_n^{AB}(G_c)$ and $O_n^{BA}(G_c)$ are defined in section 2.3. Radial labels indicate the scheme *A* and line colors the scheme *B*. The intercept between a given radial line (*A*) and a given colored line (*B*) represents the overlap, $O_n^{AB}(G_c)$. The overlap $O_n^{BA}(G_c)$ can be read from the inverted pair of radial and colored lines. The origin of the circle corresponds to an overlap of zero, while its perimeter corresponds to an overlap of one. The values are calculated using free-energy cutoffs $G_c = 5$, 10, or 15 kJ mol^{-1}.

based approach. The resulting scheme, FB-LEUS, can be applied along with basin-filling (BF) or valley-digging (VD) fragment-based biasing potentials. The feasibility of this scheme was tested in the context of (unblocked) polyalanine (poly-Ala) and polyvaline (poly-Val) oligopeptides, using fragment-based biasing potentials optimized at the (partially) blocked monopeptide level. The results show that the application of the FB-LEUS biasing potentials may lead to an impressive enhancement of the searching power (volume of conformational space visited in a given amount of simulation time). However, this enhancement 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). As a result, it appears difficult to engineer FB-LEUS schemes representing a significant improvement over plain MD, at least for the systems considered here.

This no-free-lunch conclusion might seem disappointing at first sight, and it is interesting to analyze in more detail the reasons for this failure. The problems encountered by any sampling-enhancement method relying on a memory-based biasing potential, such as the LEUS approach, in the context of a relevant conformational subspace of high dimensionality can be summarized as follows:

A. Internal Dimensionality Problem. The dimensionality of the subspace involved in the construction of the biasing potential, referred to here as its internal dimensionality (as opposed to the true dimensionality, that is, that of the relevant conformational subspace), cannot be too high due to memory costs and build-up duration requirements. This problem is addressed in the FB-LEUS scheme by using a fragment-based approach, where the internal dimensionality N_{LE} refers to the fragments and the true dimensionality $N_{US} = N_F N_{LE}$ to the real system encompassing N_F fragments.

B. Irrelevant Volume Problem. In order to enhance the rate of discovery of new relevant conformational states (i.e., the conformational-searching efficiency), as well as the statistics concerning the relative populations of these states (i.e., on their relative free energies), the biasing potential must facilitate interconversion transitions between states. This implies the creation of low free energy pathways connecting these states and, thus, the opening of an irrelevant volume of conformational space to sampling, irrelevant meaning that the accessed configurations are characterized by very low Boltzmann weights in the physical ensemble. The size of this irrelevant volume is the main determinant of the

Table 2. Free-Energy Ranking $R_{m,i}$ and Relative Free Energy $G_{m,i}$ (in parentheses, in kJ mol^{-1}) of the Five Consensus Lowest Free Energy Conformations m (Lowest Consensus Rank R_m) for Ala_n and Val_n Simulations Performed with the Seven Different FB-LEUS Biasing Potentials[a]

	m	Δ_m	U	F_{10}	F_{20}	F_{30}	D_{10}	D_5	D_0		m	Δ_m	U	F_{10}	F_{20}	F_{30}	D_{10}	D_5	D_0
					Ala_4										Val_4				
1	0		1(0.0)	1(0.0)	1(0.0)	1(0.0)	1(0.0)	1(0.0)	1(0.0)	1	0		1(0.0)	1(0.0)	1(0.0)	2(0.7)	1(0.0)	1(0.0)	2(0.3)
2	64	2	2(1.2)	2(1.6)	2(0.7)	6(3.7)	3(1.7)	2(0.8)	2(1.2)	2	64	2	2(1.3)	2(0.5)	2(0.8)	4(2.8)	3(1.5)	2(0.1)	1(0.0)
3	128	1	3(1.5)	3(1.9)	3(1.4)	2(2.4)	2(1.2)	3(1.6)	3(1.5)	3	128	1	3(1.4)	3(1.3)	3(1.5)	1(0.0)	2(1.3)	3(1.5)	3(1.5)
4	1	8	5(2.4)	4(2.6)	5(3.4)	3(2.6)	4(2.6)	4(2.2)	4(2.3)	4	1	8	4(2.0)	4(1.8)	4(2.8)	3(2.3)	4(2.3)	4(1.8)	5(2.8)
5	16	4	4(2.3)	5(3.7)	7(3.9)	11(6.1)	5(3.9)	5(3.1)	7(4.1)	5	65	2,8	6(3.6)	5(2.1)	6(3.2)	9(6.0)	5(3.9)	5(2.6)	4(2.8)
					Ala_6										Val_6				
1	0		1(0.0)	1(0.0)	1(0.0)	27(8.1)	1(0.0)	1(0.0)	1(0.0)	1	0		1(0.0)	2(0.0)	4(1.8)	6(9.7)	1(0.0)	1(0.0)	1(0.0)
2	1024	2	2(0.9)	2(1.5)	9(4.3)	24(7.8)	2(1.6)	2(1.5)	3(1.3)	2	1024	2	2(0.7)	1(0.0)	1(0.0)	3(7.6)	2(1.6)	2(0.7)	2(1.1)
3	2048	1	3(1.6)	3(1.7)	15(5.0)	16(6.1)	3(1.7)	3(1.5)	2(1.2)	3	2048	1	3(1.5)	3(1.4)	6(2.3)	110(22.5)	3(2.1)	3(1.0)	3(1.2)
4	1	12	4(2.4)	4(2.6)	2(1.6)	18(6.5)	4(2.4)	4(2.7)	4(2.7)	4	1	12	4(2.2)	4(1.9)	8(3.2)	83(20.9)	4(3.1)	6(2.6)	4(2.4)
5	1040	2,8	16(6.1)	10(4.5)	3(3.0)	3(0.3)	17(6.7)	15(7.4)	10(5.2)	5	64	6	144(29.6)	16(5.7)	2(0.6)	97(21.7)	6(4.2)	4(1.2)	13(5.5)
					Ala_8										Val_8				
1	0		1(0.0)	25(6.5)	1283(31.5)	2307(41.9)	1(0.0)	1(0.0)	1(0.0)	1	0		1(0.0)	2(2.3)	35(16.5)	101(22.2)	1(0.0)	1(0.0)	1(0.0)
2	16384	2	2(1.2)	71(9.3)	409(24.8)	5077(49.3)	2(1.3)	2(1.3)	3(2.0)	2	16384	2	2(1.5)	37(10.8)	13(11.1)	16(10.2)	3(0.7)	2(0.5)	2(0.6)
3	32768	1	3(1.6)	26(6.5)	511(26.1)	2832(43.7)	3(2.1)	3(1.4)	2(1.7)	3	32768	1	3(1.5)	5(4.5)	4(8.0)	766(38.2)	7(1.8)	3(1.2)	4(1.0)
4	1	16	4(2.9)	33(6.9)	1415(32.2)	561(31.6)	4(2.6)	4(2.5)	4(2.2)	4	1	16	4(2.1)	3(2.5)	138(24.1)	41(17.3)	6(1.5)	8(3.2)	7(2.2)
5	16	12	18(6.1)	321(15.6)	43(14.6)	2(0.4)	11(5.8)	13(5.1)	9(4.2)	5	32769	1,16	9(3.9)	1(0.0)	99(22.0)	396(32.2)	9(2.8)	16(4.8)	14(4.1)
					Ala_{10}										Val_{10}				
1	0		1(0.0)	61(11.2)	4422(43.9)	1111(38.0)	1(0.0)	1(0.0)	1(0.0)	1	0		1(0.0)	7(3.3)	3(6.4)	13920(94.3)	1(0.0)	1(0.0)	2(0.3)
2	524288	1	3(1.8)	14(6.4)	106(22.7)	9350(55.8)	2(1.5)	2(1.1)	2(1.1)	2	262144	2	4(0.2)	6(3.1)	4(6.6)	X	2(1.8)	2(0.1)	1(0.0)
3	262144	2	2(1.2)	8(5.0)	2020(38.1)	11245(57.8)	3(1.8)	3(1.7)	3(1.2)	3	524288	1	5(0.9)	3(0.5)	2(4.8)	X	3(1.8)	3(1.5)	3(1.7)
4	1	20	4(2.5)	3(2.5)	59129(75.9)	4032(47.9)	4(2.5)	4(3.0)	4(1.8)	4	1	20	14(2.2)	17(5.9)	22(19.0)	5027(78.5)	4(2.5)	4(2.5)	4(2.3)
5	262145	2,20	5(3.2)	6(4.0)	28318(63.2)	33176(71.6)	5(3.8)	8(4.5)	6(3.4)	5	4096	8	245(24.3)	1(0.0)	11113(88.0)	3514(74.0)	295(30.5)	13(4.8)	11(5.6)

[a] The biasing potentials considered are illustrated in Figure 3. The quantities m (integer corresponding to the binary code of a unique peptide backbone conformation), R_m, and $G_{m,i}$ are defined in section 2.3. The column Δ_m lists the bits of the binary code represented by m ($2n$ bits in total) that differ from zero. An "X" indicates that a conformation was never visited.

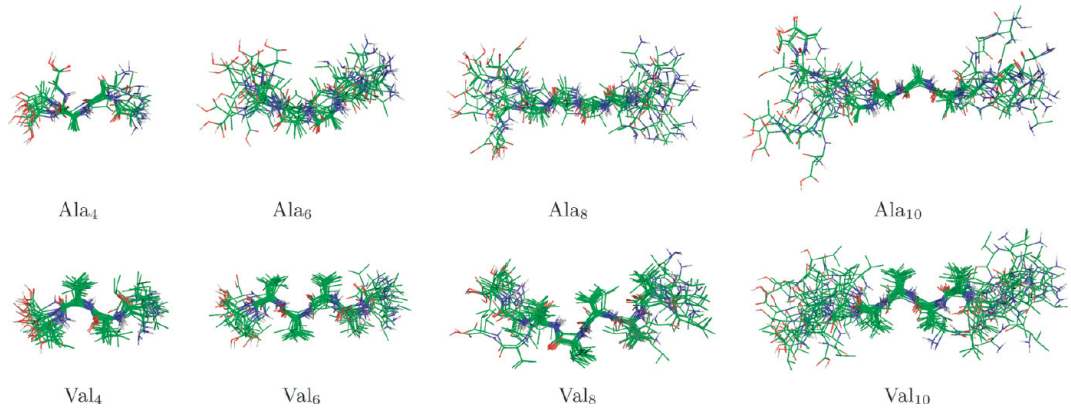

Figure 10. Illustrative configurations assigned to the backbone conformation $m = 0$ with the lowest consensus rank R_m for Ala_n and Val_n simulations performed with the seven different FB-LEUS biasing potentials. The biasing potentials considered are illustrated in Figure 3. The quantities m (integer corresponding to the binary code of a unique peptide backbone conformation) and R_m are defined in section 2.3. For each system, a bundle of 20 representative structures was extracted from successive 0.5 ns blocks along the first 10 ns of the simulation U. The structures were superimposed by rototranslational least-squares-fitting based on all backbone atoms of the four central residues (for Ala_4 the two central residues).

statistical efficiency of a biasing approach. In the context of the FB-LEUS scheme, the limitation of the irrelevant volume at the fragment level is particularly important because the corresponding volume for the real system increases exponentially with the number N_F of fragments in this system (combinatorial effect). In this respect, VD-type biasing potentials appear far superior to BF-type biasing potentials, because they promote transitions between states while opening a minimal amount of irrelevant volume to sampling. However, the oligopeptide test systems considered here do not present very high transition barriers (poly-Val slightly more than poly-Ala), so that transition rates are only moderately increased by the VD scheme.

The additional problems encountered specifically by the FB-LEUS approach can be summarized as follows:

C. Free-Energy Additivity Approximation. The LEUS scheme in terms of BF-type biasing potentials relies on the principle that a biasing potential that is approximately equal to the negative of the free-energy hypersurface in the relevant conformational subspace up to a given threshold above the global minimum, will lead to a nearly homogeneous coverage of this region in the biased simulation. However, a FB-LEUS biasing potential will only satisfy this condition if the free-energy function of the real system can be expressed as a sum of independent contributions from the N_F fragments. In reality, this sum can only account for local (single-fragment) contributions, thereby neglecting nonlocal (fragment-coupling) effects. Similar considerations apply to the VD-type biasing potentials in terms of the locations of the free-energy minima and magnitudes of transition barriers at the fragment and real-system levels. The presence of nonlocal effects has both negative and positive consequences for the FB-LEUS scheme. On the negative side, the FB-LEUS biasing potential is not strictly appropriate for the homogeneous (BF) or accelerated (VD) sampling of the real system, which may lead to a decrease in its searching efficiency. On the positive side, the presence of residual nonlocal driving forces may steer the biased system away from irrelevant conformational regions (e.g., by secondary structure formation or folding), thereby increasing its statistical efficiency. However, the oligopeptide test systems considered here do not present strong nonlocal interactions (poly-Val slightly more than poly-Ala) and appear to be significantly affected by neither of the two effects.

D. Combinatorial Problem. In cases where nonlocal effects are negligible, the application of the FB-LEUS scheme with a BF-type biasing potential will promote facilitated interconversions between a finite number M of equally deep free-energy basins at the fragment level (e.g., $M = 4$ for the present oligopeptide systems). Thus, it will induce the sampling of M^{N_F} equally populated states for a real system encompassing N_F fragments. Even if the searching rate is increased, visiting as many states, if tractable at all, requires a sizable amount of simulation time for all but the lowest N_F. Furthermore, nearly all these states must be visited before converged relative free energies can be obtained, because the biased sampling removes any discrimination between these states in terms of their relative free energies in the physical ensemble. In other words, the FB-LEUS scheme with BF-type biasing potentials (and assuming that nonlocal effects are negligible) is not very different from a systematic or random scanning approach of the relevant conformational subspace. For example, scheme F_{30} applied to Ala_{10} produces about 230 000 unique peptide backbone conformations within 50 ns of simulation. The evaluation of as many conformations by systematic or random scanning within the same simulation time would permit a sampling time of 0.2 ps per conformation, which is very close to the effective visiting time of 0.34 ps for this simulation. The situation is different for the VD-type biasing potentials, where the relative free energies of the M different states at the fragment level remain unaltered compared with the physical system, and thus generally different (only the transition barriers are changed). As a result the corresponding M^{N_F} states at the real-system level

also have different free energies, and the biased sampling retains a statistical efficiency comparable to that of plain MD (and much higher than in the BF case).

Considering points B and D above, as well as the results of the present study, it appears that the use of the FB-LEUS scheme along with BF-type biasing potentials is not a viable approach to address high-dimensionality problems. On the other hand, the FB-LEUS scheme relying on VD-type biasing potentials successfully addresses points A, B, and D, while point C seems to be of little relevance in the context of the oligopeptide test systems considered here. Nevertheless, the results of the present study suggest that biasing potentials of this type do not represent a significant improvement over plain MD for poly-Ala and poly-Val oligopeptides.

The reason for this failure is probably in large part related to the choice of these test systems. Because the peptide linkage is relatively flexible, there is only little gain in conformational searching efficiency upon "digging" valleys between the corresponding free-energy basins. This moderate gain is in turn partly offset by a minimal but unavoidable loss of statistical efficiency upon opening a small irrelevant volume to sampling. In other words, for these systems, it appears nearly impossible to significantly improve over plain MD in terms of combined sampling efficiency.

Work is currently in progress to investigate the performance of the FB-LEUS approach in the context of oligosaccharide test systems. In view of the much more "rigid" nature of the glycosidic linkage,[62] a significant sampling enhancement could be achieved in this case using VD-type biasing potentials. Finally, the development of an alternative approach, ball-and-stick LEUS (B&S-LEUS), that combines biasing potentials of low internal dimensionalities, minimal irrelevant volumes, and problem-adapted geometries (and is also generalizable to the calculation of "alchemical" free-energy differences) will be reported in a subsequent article.[90]

Acknowledgment. Financial support from the Swiss National Science Foundation (Grant NF200021-121895) is gratefully acknowledged. X.D. also acknowledges support from the Spanish MICINN/FEDER (Grant BIO2007-62954).

Appendix A. LEUS Method with Truncated Polynomials

The local elevation umbrella sampling[15,62] (LEUS) method consists of two steps: (i) a LE build-up (searching) phase that is used to construct an optimized memory-based biasing potential within a LE subspace of N_{LE} conformationally relevant degrees of freedom; (ii) an US sampling phase, where the (frozen) memory-based potential is used to generate a biased ensemble with extensive coverage of the US subspace defined by the same $N_{US} = N_{LE}$ degrees of freedom. During the LE build-up phase, the searching is carried out for a duration t_{LE} using (thermostatted and possibly barostatted) molecular dynamics (MD) based on the time-dependent potential energy function

$$\mathcal{U}_{LE}(\mathbf{r}, \mathcal{V}, t) = \mathcal{U}_{phys}(\mathbf{r}, \mathcal{V}) + \mathcal{U}_{bias}(\mathbf{Q};\mathbf{M}(t)) \quad (A.1)$$

During the subsequent US sampling phase, the sampling is carried out for a duration t_{US} using (thermostatted and possibly barostatted) MD based on the time-independent potential energy function

$$\mathcal{U}_{US}(\mathbf{r}, \mathcal{V}) = \mathcal{U}_{phys}(\mathbf{r}, \mathcal{V}) + \mathcal{U}_{bias}(\mathbf{Q};\mathbf{M}(t_{LE})) \quad (A.2)$$

In eqs A.1 and A.2, $\mathcal{U}_{phys}$ is the physical potential energy function of the system (force field), $\mathcal{U}_{bias}$ is the memory-based biasing potential, $\mathcal{V} = \mathcal{V}(t)$ is the system volume at time t, $\mathbf{r} = \mathbf{r}(t)$ is the vector containing the Cartesian coordinates of all atoms in the system (configuration) at time t, $\mathbf{Q} = \mathbf{Q}(\mathbf{r})$ is the point representative of the current system configuration in the chosen LE subspace, and $\mathbf{M}$ is the memory vector of the biasing potential. To entirely define the LEUS scheme, the above equations must be complemented by a representation of the biasing potential $\mathcal{U}_{bias}$ in terms of its memory $\mathbf{M}$ and by an updating scheme for this memory during the build-up phase, as discussed below.

The representation of the memory-based biasing potential relies on a discretization of the LE subspace (N_{LE} dimensions) by means of N_G grid points $\{\mathbf{Q}_n | n = 1, ..., N_G\}$ defining the centers of nonoverlapping grid cells tiling this subspace. The memory $\mathbf{M}$ consists of an N_G-dimensional integer vector accounting for the number of visits to a given grid cell (up to a time t during the build-up phase, or based on the total time t_{LE} of the build-up phase during the sampling phase). In its most general form, the biasing potential is then written

$$\mathcal{U}_{bias}(\mathbf{Q};\mathbf{M}) = \sum_{n=1}^{N_G} k_{LE,n} M_n F_n(\mathbf{Q}) \quad (A.3)$$

where F_n is a "unit" (i.e., normalized by the condition $F(0) = 1$) local repulsive function associated with grid point n and $k_{LE,n}$ is a corresponding force constant, while the updating scheme during the build-up phase is typically written

$$M_n(t + \Delta t) = M_n(t) + h_n(\mathbf{Q}) \quad (A.4)$$

where Δt is the simulation time step and h_n evaluates to one for the (single) grid cell encompassing $\mathbf{Q}$ and to zero otherwise. The grid-cell shapes and sizes, as well as the force constants $k_{LE,n}$ and functions F_n in eq A.3 can in principle be chosen differently for all grid points. However, unless such a high extent of flexibility is desired, it is common to assume a rectangular grid (spacing d_i along dimension i), a common force constant k_{LE}, and an unique functional form for F_n, the latter defined as a product of one-dimensional functions of the displacements relative to the grid-cell center along each dimension. In this case eq A.3 can be rewritten as

$$\mathcal{U}_{bias}(\mathbf{Q};\mathbf{M}) = k_{LE} \sum_{n=1}^{N_G} M_n F(\mathbf{Q} - \mathbf{Q}_n;\mathbf{d}) \quad (A.5)$$

with

$$F(\mathbf{Q};\mathbf{d}) = \prod_{i=1}^{N_{LE}} f(Q_i;d_i) \quad (A.6)$$

with the additional condition $f(0; d) = 1$ for any finite d.

The original LE method[56] relied on one-dimensional truncated Gaussian functions, that is,

$$f(x;d) = \exp\left[-\frac{x^2}{2\sigma^2}\right]H\left(1 - \frac{|x|}{d}\right) \quad \text{(A.7)}$$

where H is the Heaviside step function (evaluating to one when its argument is positive, to zero otherwise), and σ is a Gaussian width ($\sigma = d$ was used in the original article[56]). The use of eq A.7 is not recommended in practice because it leads to a biasing potential that is (i) generally not continuous and (ii) nonperiodic along possible periodic coordinates (e.g., angles).

The original LEUS method[15] suggested the use of one-dimensional minimum-image Gaussian functions instead, that is,

$$f(x) = \exp\left[-\frac{\mathrm{MI}(x)^2}{2\sigma^2}\right] \quad \text{(A.8)}$$

where MI is the minimum-image function, selecting values within the period centered at zero (i.e., for which $|x|$ is minimal) for periodic coordinates, and σ is a Gaussian width ($\sigma = d$ was used in the original article[15]). Note that k_{LE} in eq A.5 is related to c_{LE} in ref 15 (see eqs 1 and 2 therein) as $k_{\mathrm{LE}} = (2\pi\sigma^2)^{N_{\mathrm{LE}}/2c_{\mathrm{LE}}}$ (k_{LE} is a more convenient parameter, with units of energy). The use of eq A.8 represents an improvement, leading to a biasing potential that is (i) continuous and differentiable and (ii) periodic along possible periodic coordinates. However, because the Gaussian function is infinite-ranged, the formal range of the biasing potential contribution associated with a given grid point has been extended to the entire LE subspace (even if the corresponding effective range is much shorter when σ is chosen close to d). This leads to an increased computational cost or to the requirement of introducing a suitable cutoff distance (similar to eq A.7 but involving longer distances than d). Note that an alternative solution involves the multiplication of the Gaussian function by a polynomial that switches it to zero at finite range.[44]

Clearly, neither eq A.7 nor eq A.8 is entirely satisfactory. In addition to the shortcomings mentioned above, grid-based Gaussian functions do not form an appropriate basis set for the representation of a constant function. This means that in addition to "flattening" the free-energy hypersurface in the relevant conformational subspace, the corresponding biasing potential will introduce spurious oscillations on a length scale equal to the grid spacing (these oscillations are clearly visible in Figure 2 of ref 56). The magnitude of these oscillations will increase with the extent of build-up and may ultimately result in a strong artificial bias of the system toward grid-cell boundaries, and the appearance of very high artificial forces preventing an accurate integration of the equations of motion. Finally, the calculation of a Gaussian-based biasing potential involves the evaluation of (computationally expensive) exponential functions. In the present study, truncated polynomial functions (similar to the assignment functions described in ref 91) are used instead, as detailed below.

The new formulation of the biasing potential employed in the present study relies on a local function of the form

$$f(x;d) = \left[1 - 3\left(\frac{\mathrm{MI}(x)}{d}\right)^2 + 2\left(\frac{|\mathrm{MI}(x)|}{d}\right)^3\right]H\left(1 - \frac{|\mathrm{MI}(x)|}{d}\right) \quad \text{(A.9)}$$

This local function satisfies the following desirable properties:

1. It is finite ranged, that is, $f(x) = 0$ for $|x| \geq d$, allowing for an evaluation of the biasing potential based on a sum involving a limited number of grid points ($2^{N_{\mathrm{LE}}}$).
2. It is even, that is, $f(-x) = f(x)$, so that it does not induce any directional bias.
3. It is monotonic below and above $x = 0$, so that it does not induce artificial minima.
4. It is continuously differentiable (continuous first derivative), including at $x = \pm d$, leading to a continuously differentiable biasing potential.
5. It is periodic along periodic degrees of freedom, that is, $f(x + np) = f(x)$ where p is the period and n is a positive or negative integer (and assuming $p \geq d$).
6. It defines an appropriate grid-based basis set for the exact representation of the constant function, which follows from the property $f(x) + f(x - d) = 1$ for $0 < x < d$.
7. It evaluates to one at $x = 0$, that is, $f(0) = 1$, allowing for a direct interpretation of k_{LE} as the magnitude of the biasing energy at a grid point.
8. It is computationally inexpensive, since no exponential function is involved.

It is easily verified that if these one-dimensional functions define an appropriate grid-based basis set for the exact representation of the constant function, the same property holds for the function F of eq A.6 in the multidimensional LE subspace.

An alternative basis function satisfying properties 1−8 above and characterized in addition by a continuous second derivative would be

$$\tilde{f}(x;d) = \left[1 - 10\left(\frac{|\mathrm{MI}(x)|}{d}\right)^3 + 15\left(\frac{\mathrm{MI}(x)}{d}\right)^4 - 6\left(\frac{|\mathrm{MI}(x)|}{d}\right)^5\right]H\left(1 - \frac{|\mathrm{MI}(x)|}{d}\right) \quad \text{(A.10)}$$

This alternative function could be employed in situations where integration errors caused by second-order discontinuities are an issue.[92] Finally, it should be noted that an alternative approach involves the use of an interpolation scheme to obtain the bias energy between grid points.[48]

Appendix B. Grid-Based BF and VD Biasing Potentials

In the FB-LEUS scheme, the build-up phase of duration t_{LE} is performed at the fragment level (LE subspace of dimension N_{LE}) according to eqs A.1 and A.4, using a memory $\mathbf{M}$ of dimension N_{G}, resulting in an optimized biasing potential $\mathcal{U}_{\mathrm{bias}}(\mathbf{Q};\mathbf{M}(t_{\mathrm{LE}}))$. However, the sampling phase of duration t_{US} is performed at the system level (US subspace of dimension $N_{\mathrm{US}} = N_{\mathrm{F}}N_{\mathrm{LE}}$ where N_{F} is the number of fragments in the system), replacing eq A.2 by

$$\mathcal{U}_{\mathrm{US}}(\mathbf{r}, \mathcal{V}, t) = \mathcal{U}_{\mathrm{phys}}(\mathbf{r}, \mathcal{V}) + \sum_{m=1}^{N_{\mathrm{F}}} \mathcal{U}'_{\mathrm{bias}}(\mathbf{Q}^{(m)};h) \quad \text{(B.1)}$$

Here, $\mathbf{Q}^{(m)} = \mathbf{Q}^{(m)}(\mathbf{r})$ is the N_{LE}-dimensional point representative of fragment m in the LE subspace and $\mathcal{U}'_{\mathrm{bias}}$ is constructed according to the BF ($\mathcal{U}^{\mathrm{BF}}_{\mathrm{bias}}$) or VD ($\mathcal{U}^{\mathrm{VD}}_{\mathrm{bias}}$) procedure based on the given height parameter h (section 2.2). This construction requires the evaluation of the free energy $G(\mathbf{Q})$ from a sampling phase of duration t'_{US} at the fragment level, and stored as a N_{LE}-dimensional grid-based vector $\mathbf{G}$, that is,

$$G_n = -\beta^{-1} \ln\langle h_n(\mathbf{Q}) \exp[\beta \mathcal{U}_{\mathrm{bias}}(\mathbf{Q};\mathbf{M}(t_{\mathrm{LE}}))]\rangle_{t'_{\mathrm{US}}} + C \quad \text{(B.2)}$$

where C is chosen so that the lowest G_n value is zero (all others being positive). The procedure used to construct $\mathcal{U}^{\mathrm{BF}}_{\mathrm{bias}}$ and $\mathcal{U}^{\mathrm{VD}}_{\mathrm{bias}}$ based on $G(\mathbf{Q})$ is described qualitatively by eqs 1−5. A more precise (grid-based) description is provided below.

The grid-based expression corresponding to eq 1 is (in analogy to eq A.5)

$$\mathcal{U}^{\mathrm{BF}}_{\mathrm{bias}}(\{\phi,\psi\};h) = \sum_{n=1}^{N_{\mathrm{G}}} M_n^{\mathrm{BF}}(h) F_n(\{\phi - \phi_n, \psi - \psi_n\}; \{d_\phi, d_\psi\}) \quad \text{(B.3)}$$

where ϕ_n and ψ_n are the ϕ and ψ values associated with grid point n, d_ϕ and d_ψ are the grid spacings along ϕ and ψ, and

$$M_n^{\mathrm{BF}}(h) = \begin{cases} h - G_n & \text{if } G_n \le h \\ 0 & \text{otherwise} \end{cases} \quad \text{(B.4)}$$

The grid-based expression corresponding to eq 2 is (in analogy to eq A.5)

$$\mathcal{U}^{\mathrm{VD}}_{\mathrm{bias}}(\{\phi,\psi\};h) = \sum_{n=1}^{N_{\mathrm{G}}} M_n^{\mathrm{VD}}(h) F_n(\{\phi - \phi_n, \psi - \psi_n\}; \{d_\phi, d_\psi\}) \quad \text{(B.5)}$$

with

$$M_n^{\mathrm{VD}}(h) = \begin{cases} h - \Delta\tilde{G}_n & \text{if } \tilde{G}_n \le h \\ 0 & \text{otherwise} \end{cases} \quad \text{(B.6)}$$

where

$$\Delta\tilde{G}_n = \mathrm{MAX}[\Delta G_{i(n,k)} - fj(n,k), k = 1, ..., 8] \quad \text{(B.7)}$$

In the above equations, MAX returns the maximum value of a set, $i(n, k)$ is the grid point defined by the projection of grid point n onto line k, $j(n, k)$ is the number of grid points separating n from $i(n, k)$ and f is a force threshold in units of energy per grid spacing. For a grid point belonging to a line k, $i(n, k) = n$ and $j(n, k) = 0$, so that eqs B.5−B.7 are equivalent to eq 2. For a grid point not belonging to a line k, these equations ensure that the force component transverse to a line will never exceed f. In practice, this modification will only affect lines parallel to a given line k and distant by one (or a few) grid spacing units, and predominantly in the high free energy region of the line. In the present work $d^{-1}f$ was set to 1 kJ mol^{-1} deg^{-1}. Note that, as was the case for $\mathcal{U}_{\mathrm{bias}}$ in Appendix A, the biasing potentials $\mathcal{U}^{\mathrm{BF}}_{\mathrm{bias}}$ and $\mathcal{U}^{\mathrm{VD}}_{\mathrm{bias}}$ are weighted sums of continuously differentiable functions and are thus themselves continuously differentiable (despite the discontinuously differentiable truncation of the weights in eqs B.4 and B.6).

Appendix C. Significance of the Statistical Efficiency Factor

The statistical efficiency factor F introduced in eqs 8 and 9 characterizes the mutual relationship between the physical potential energy $\mathcal{U}_{\mathrm{phys}}(\mathbf{r})$ and the biasing potential $\mathcal{U}_{\mathrm{bias}}(\mathbf{r})$. This is most easily seen by considering the infinite-sampling limit of these equations. The configurational probability distributions within the physical and biased ensembles can be written

$$\rho_{\mathrm{phys}}(\mathbf{r}) = Z_{\mathrm{phys}}{}^{-1} \exp[-\beta \mathcal{U}_{\mathrm{phys}}(\mathbf{r})] \quad \text{(C.1)}$$

with

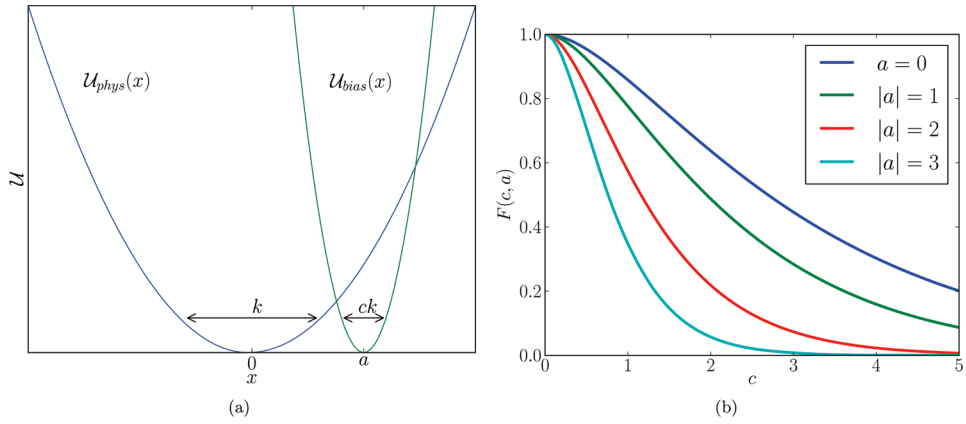

Figure 11. Illustration of the effect of the biasing potential on the statistical efficiency considering a simple one-dimensional harmonic situation: (a) physical potential energy $\mathcal{U}_{\mathrm{phys}}(x)$, harmonic with force constant k, and biasing potential energy $\mathcal{U}_{\mathrm{bias}}(x)$, harmonic with force constant ck and coordinate offset a (eqs C.10 and C.11); (b) statistical efficiency $F(c, a)$ as a function of c for different values of a (eq C.12).

$$Z_{\text{phys}} = \int \mathrm{d}\mathbf{r} \exp[-\beta \mathscr{U}_{\text{phys}}(\mathbf{r})] \quad \text{(C.2)}$$

and

$$\rho_{\text{biased}}(\mathbf{r}) = {Z_{\text{biased}}}^{-1} \exp\{-\beta[\mathscr{U}_{\text{phys}}(\mathbf{r}) + \mathscr{U}_{\text{bias}}(\mathbf{r})]\} \quad \text{(C.3)}$$

with

$$Z_{\text{biased}} = \int \mathrm{d}\mathbf{r} \exp\{-\beta[\mathscr{U}_{\text{phys}}(\mathbf{r}) + \mathscr{U}_{\text{bias}}(\mathbf{r})]\} \quad \text{(C.4)}$$

Given a trajectory of N_f frames and in the limit $N_f \rightarrow \infty$, the density of frames in the biased ensemble is given by $N_f \rho_{\text{biased}}(\mathbf{r})$. Equations 8 and 9 can thus be rewritten in this limit as

$$F_n = {N_f}^{-1} \exp[-N_f \int \mathrm{d}\mathbf{r}\, \rho_{\text{biased}}(\mathbf{r}) p(\mathbf{r}) \ln p(\mathbf{r})] \quad \text{(C.5)}$$

with

$$p(\mathbf{r}) = \{N_f \int \mathrm{d}\mathbf{r}\, \rho_{\text{biased}}(\mathbf{r}) \exp[\beta \mathscr{U}_{\text{bias}}(\mathbf{r})]\}^{-1} \exp[\beta \mathscr{U}_{\text{bias}}(\mathbf{r})] = \frac{Z_{\text{biased}}}{N_f Z_{\text{phys}}} \exp[\beta \mathscr{U}_{\text{bias}}(\mathbf{r})] \quad \text{(C.6)}$$

Using eqs C.1, C.4, and C.6, one can easily rearrange eq C.5 to

$$F = \frac{Z_{\text{phys}}}{Z_{\text{biased}}} \exp[-{Z_{\text{phys}}}^{-1} \int \mathrm{d}\mathbf{r}\, \beta \mathscr{U}_{\text{bias}}(\mathbf{r}) \exp[-\beta \mathscr{U}_{\text{phys}}(\mathbf{r})]] \quad \text{(C.7)}$$

or, in terms of ensemble averages $\langle ... \rangle$ over the physical ensemble,

$$F = \frac{\exp[-\langle \beta \mathscr{U}_{\text{bias}}(\mathbf{r}) \rangle]}{\langle \exp[-\beta \mathscr{U}_{\text{bias}}(\mathbf{r})] \rangle} \quad \text{(C.8)}$$

By applying the inequality of arithmetic and geometric means,[93] which states that

$$N^{-1} \sum_{i=1}^{N} x_i \geq (\prod_{i=1}^{N} x_i)^{1/N} \quad \text{(C.9)}$$

when all $x_i > 0$ (the equality holding only when all x_i are identical), it is easily seen that the quantity F is always positive and reaches a maximum value of one for $\mathscr{U}_{\text{bias}}(\mathbf{r}) = cst$. This observation is important because it shows that the application of a biasing potential always decreases the statistical efficiency compared with a sampling relying on $\mathscr{U}_{\text{phys}}(\mathbf{r})$ only.

This behavior is illustrated qualitatively in Figure 11, considering a simple one-dimensional harmonic situation with

$$\mathscr{U}_{\text{phys}}(x) = \frac{1}{2} k x^2 \quad \text{(C.10)}$$

and

$$\mathscr{U}_{\text{bias}}(x, c, a) = \frac{1}{2} c k (x - a)^2 \quad \text{(C.11)}$$

In this case, eq C.8 can be evaluated analytically, leading to

$$F(c, a) = (c + 1)^{1/2} \exp\left[-\frac{c(c + 1 + \beta c k a^2)}{2(c + 1)}\right] \quad \text{(C.12)}$$

In the absence of biasing potential ($c = 0$), F evaluates to one as expected. However, as the value of c is increased (narrowing of the biasing potential), the statistical efficiency monotonically decreases toward a limiting value of zero. The decrease is more rapid when $|a|$ is large (decentering of the biasing potential), because the narrowing of the biasing potential focuses the sampling on high-energy regions in terms of the physical potential energy, but is also observed for $a = 0$ (centered biasing potential).

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.* **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.; Hünenberger, 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) Daura, X.; van Gunsteren, W. F.; Mark, A. E. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 269–280.

(17) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.

(18) Snow, C. D.; Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 43–69.

(19) Eaton, W. A.; Muñoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J. *Acc. Chem. Res.* **1998**, *31*, 745–753.

(20) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102–106.

(21) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691–697.

(22) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220–7225.

(23) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196–199.

(24) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.

(25) Hünenberger, P. H.; Granwehr, J. K.; Aebischer, J.-N.; Ghoneim, N.; Haselbach, E.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 7533–7544.

(26) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925–932.

(27) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.

(28) Valleau, J. P.; Torrie, G. M. In *Modern Theoretical Chemistry*; Berne, B. J., Ed.; Plenum Press: New York, 1977; Vol. 16, pp 9−194.

(29) Beutler, T. C.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 1492–1497.

(30) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–1350.

(31) Heinz, T. N.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Chem. Phys.* **2001**, *115*, 1125–1136.

(32) Piccinini, E.; Ceccarelli, M.; Affinito, F.; Brunetti, R.; Jacoboni, C. *J. Chem. Theory Comput.* **2008**, *4*, 173–183.

(33) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *12*, 1195–1198.

(34) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(35) Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450–1462.

(36) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 26–41.

(37) Kästner, J.; Thiel, W. *J. Chem. Phys.* **2005**, *123*, 144104.

(38) Kästner, J.; Thiel, W. *J. Chem. Phys.* **2006**, *124*, 234106.

(39) Paine, G. H.; Scheraga, H. A. *Biopolymers* **1985**, *24*, 1391–1436.

(40) Mezei, M. *J. Comput. Phys.* **1987**, *68*, 237–248.

(41) Hooft, R. W. W.; van Eijck, B. P.; Kroon, J. *J. Chem. Phys.* **1992**, *97*, 6690–6694.

(42) Friedman, R. A.; Mezei, M. *J. Chem. Phys.* **1995**, *102*, 419–426.

(43) Wang, J.; Gu, Y.; Liu, H. *J. Chem. Phys.* **2006**, *125*, 094907.

(44) Babin, V.; Roland, C.; Darden, T. A.; Sagui, C. *J. Chem. Phys.* **2006**, *125*, 204909.

(45) Marsili, S.; Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2006**, *110*, 14011–14013.

(46) Lelièvre, T.; Rousset, M.; Stoltz, *J. Chem. Phys.* **2007**, *126*, 134111.

(47) van der Vaart, A.; Karplus, M. *J. Chem. Phys.* **2007**, *126*, 164106.

(48) Babin, V.; Roland, C.; Sagui, C. *J. Chem. Phys.* **2008**, *128*, 134101.

(49) Barnett, C. B.; Naidoo, K. J. *Mol. Phys.* **2009**, *107*, 1243–1250.

(50) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

(51) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.

(52) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120.

(53) Crippen, G. M.; Scheraga, H. A. *Chemistry* **1969**, *64*, 42–49.

(54) Levy, A. V.; Montalvo, A. *SIAM J. Sci. Stat. Comput.* **1985**, *6*, 15–29.

(55) Glover, F. *ORSA J. Comput.* **1989**, *1*, 190–206.

(56) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695–708.

(57) Grubmüller, H. *Phys. Rev. E* **1995**, *52*, 2893–2906.

(58) Engkvist, O.; Karlström, G. *Chem. Phys.* **1996**, *213*, 63–76.

(59) Fukunishi, Y.; Mikami, Y.; Nakamura, H. *J. Phys. Chem. B* **2003**, *107*, 13201–13210.

(60) 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.

(61) 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.

(62) Perić-Hassler, L.; Hansen, H. S.; Baron, R.; Hünenberger, P. H. *Carbohydr. Res.* **2010**, *345*, 1781–1801.

(63) Leitgeb, M.; Schröder, C.; Boresch, S. *J. Chem. Phys.* **2005**, *122*, 084109.

(64) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169–9183.

(65) Darve, E.; Wilson, M. A.; Pohorille, A. *Mol. Simul.* **2002**, *28*, 113–144.

(66) Naidoo, K. J.; Brady, J. W. *J. Am. Chem. Soc.* **1999**, *121*, 2244–2252.

(67) Kuttel, M. M.; Naidoo, K. J. *J. Phys. Chem. B* **2005**, *109*, 7468–7474.

(68) Strümpfer, J.; Naidoo, K. J. *J. Comput. Chem.* **2010**, *31*, 308–316.

(69) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. *J. Phys. Chem. B* **2005**, *109*, 6676–6687.

(70) Li, H.; Min, D.; Liu, Y.; Yang, W. *J. Chem. Phys.* **2007**, *127*, 094101.

(71) Hansen, H. S.; Hünenberger, P. H. *J. Comput. Chem.* 2010, submitted for publication.

(72) Kannan, S.; Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 697–706.

(73) Xu, C.; Wang, J.; Liu, H. *J. Chem. Theory Comput.* **2008**, *4*, 1348–1359.

(74) 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.

(75) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(76) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Eds.; Reidel: Dordrecht, The Netherlands, 1981; Vol. 33, pp 1−342.

(77) Feynman, R. P.; Leighton, R. B.; Sands, M. *The Feynman Lectures on Physics*; Addison-Wesley: Boston, MA, 1963.

(78) Hockney, R. W. *Methods Comput. Phys.* **1970**, *9*, 136–211.

(79) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(80) 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.

(81) Berendsen, H. J. C.; van Gunsteren, W. F.; Zwinderman, H. R. J.; Geurtsen, R. G. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 269–285.

(82) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

(83) Weber, W.; Hünenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668–3675.

(84) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 054103.

(85) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 084109.

(86) Shen, T.; Hamelberg, D. *J. Chem. Phys.* **2008**, *129*, 034103.

(87) Shell, M. S. *J. Chem. Phys.* **2008**, *129*, 144108.

(88) Zwanzig, R.; Szabo, A.; Bagchi, B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 20–22.

(89) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.

(90) Hansen, H. S.; Hünenberger, P. H. *J. Chem. Theory Comput.*, DOI: 10.1021/ct1003065.

(91) Hünenberger, P. H. In *Simulation and Theory of Electrostatic Interactions in Solution: Computational Chemistry, Biophysics, and Aqueous Solution*; Hummer, G., Pratt, L. R., Eds.; American Institute of Physics: New York, 1999; Vol. 1, pp 7−83.

(92) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.

(93) Jensen, J. L. W. V. *Acta Math.* **1906**, *30*, 175–193.

CT1003059