Computer Physics Communications 183 (2012) 203–211

Contents lists available at SciVerse ScienceDirect

Computer Physics Communications

www.elsevier.com/locate/cpc

METAGUI. A VMD interface for analyzing metadynamics and molecular dynamics simulations ☆

Xevi Biarnés [a,d,*], Fabio Pietrucci [b], Fabrizio Marinelli [c], Alessandro Laio [d]

[a] Institut Químic de Sarrià (IQS), Laboratory of Biochemistry, Via Augusta, 390, Barcelona, ES 08017, Spain
[b] Centre Européen de Calcul Atomique et Moléculaire (CECAM), Ecole Polytechnique Fédérale de Lausanne, BCH 4102, Lausanne, CH 1015, Switzerland
[c] Max Planck Institute of Biophysics (MPIBP), Theoretical Molecular Biophysics, Max-von-Laue-Straße, 3, Frankfurt am Main, DE 60438, Germany
[d] Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea, 265, Trieste, IT 34136, Italy

ARTICLE INFO

Article history:
Received 12 July 2011
Received in revised form 22 August 2011
Accepted 24 August 2011
Available online 21 September 2011

Keywords:
Metadynamics
Bias exchange
VMD
PLUMED
Molecular dynamics simulation
Thermodynamics
Kinetics

ABSTRACT

We present a new computational tool, METAGUI, which extends the VMD program with a graphical user interface that allows constructing a thermodynamic and kinetic model of a given process simulated by large-scale molecular dynamics. The tool is specially designed for analyzing metadynamics based simulations. The huge amount of diverse structures generated during such a simulation is partitioned into a set of microstates (i.e. structures with similar values of the collective variables). Their relative free energies are then computed by a weighted-histogram procedure and the most relevant free energy wells are identified by diagonalization of the rate matrix followed by a commitor analysis. All this procedure leads to a convenient representation of the metastable states and long-time kinetics of the system which can be compared with experimental data. The tool allows to seamlessly switch between a collective variables space representation of microstates and their atomic structure representation, which greatly facilitates the set-up and analysis of molecular dynamics simulations. METAGUI is based on the output format of the PLUMED plugin, making it compatible with a number of different molecular dynamics packages like AMBER, NAMD, GROMACS and several others. The METAGUI source files can be downloaded from the PLUMED web site (http://www.plumed-code.org).

Program summary

Program title: METAGUI
Catalogue identifier: AEKH_v1_0
Program summary URL: http://cpc.cs.qub.ac.uk/summaries/AEKH_v1_0.html
Program obtainable from: CPC Program Library, Queen's University, Belfast, N. Ireland
Licensing provisions: GNU General Public License version 3
No. of lines in distributed program, including test data, etc.: 117 545
No. of bytes in distributed program, including test data, etc.: 8 516 203
Distribution format: tar.gz
Programming language: TK/TCL, Fortran
Computer: Any computer with a VMD installation and capable of running an executable produced by a gfortran compiler
Operating system: Linux, Unix OS-es
RAM: 1 073 741 824 bytes
Classification: 23
External routines: A VMD installation (http://www.ks.uiuc.edu/Research/vmd/)
Nature of problem: Extract thermodynamic data and build a kinetic model of a given process simulated by metadynamics or molecular dynamics simulations, and provide this information on a dual representation that allows navigating and exploring the molecular structures corresponding to each point along the multi-dimensional free energy hypersurface.
Solution method: Graphical-user interface linked to VMD that
1. clusterizes the simulation trajectories in the space of a set of collective variables and assigns each frame to a given microstate,

☆ This paper and its associated computer program are available via the Computer Physics Communications homepage on ScienceDirect (http://www.sciencedirect.com/science/journal/00104655).

* Corresponding author at: Institut Químic de Sarrià (IQS), Laboratory of Biochemistry, Via Augusta, 390, Barcelona, ES 08017, Spain. Tel.: +34 932672000.
E-mail address: xevi.biarnes@iqs.edu (X. Biarnés).

0010-4655/$ – see front matter
doi:10.1016/j.cpc.2011.08.020

2. determines the free energy of each microstate by a weighted histogram analysis method, and
3. identifies the most relevant free energy wells (kinetic basins) by diagonalization of the rate matrix followed by a commitor analysis.

Restrictions: Input format files compatible with PLUMED and all the MD engines supported by PLUMED and VMD.
Running time: A few minutes.

1. Introduction

In these last years it has become possible to predict by accurate atomistic simulations the thermodynamics and kinetics of very complex processes, such as the folding of small proteins [1–7]. A very important and often overlooked role in these achievements has been played by visualization tools, such as VMD [8]. These tools allow handling and visualizing the huge amount of structures generated by molecular dynamics, allowing a direct grasp on the process that is investigated. Visualization tools have grown in complexity, speed and reliability together with computational power: the possibility of generating more and more structures very different from each other leads to the problem of ordering and visualizing them in a manner that helps human understanding.

Visualization becomes even more important if advanced sampling algorithms are used for enhancing the exploration of the configuration space. These approaches allow generating, in a relatively short time, much more independent structures than normal molecular dynamics. An additional "visualization" problem is posed by free energy methods and, in particular, by metadynamics [9]. This approach requires choosing a set of complex functions of the coordinates, the collective variables (CVs), that have the role of driving the simulation over the relevant barriers. If this set does not include all the relevant, "slow", degrees of freedom, any free energy method fails, as the estimate of the relevant thermodynamic potentials is affected by systematic errors. Checking if the CVs that have been chosen are appropriate and the simulation can be trusted is extremely challenging and is normally done exploiting chemical understanding. Visualization tools provide an essential support to this task.

We here present a new tool, METAGUI, based on VMD [8], specifically designed for analyzing and visualizing metadynamics-based simulations in biomolecules [10–12,7] and other systems. The tool can also be used to analyze unbiased molecular dynamics simulations. METAGUI works as follows: first, the trajectory frames are grouped into microstates in a possibly high-dimensional space of CVs. The tool allows checking the structural consistency of the microstates by visualizing the corresponding molecular structures that, for a well chosen set of CVs, have to be similar for configurations belonging to the same microstate. Next, the equilibrium free energy of each microstate and the kinetic transition matrix is computed by a weighted histogram technique. Finally, the microstates are grouped together into kinetic basins by analyzing the spectrum of the kinetic matrix. At the end of the process, all this information is represented in a graphical interface that allows an easy toggling between two visualization modes: one in which the free energy is represented in a suitable projection in CV space (by necessity at most three-dimensional), and one in which the molecular structures assigned to a specific location on this free energy surface are displayed. An example of the two visualizations is provided in Fig. 4b. In the left panel, the spheres represent "microstates". Different colors correspond to different kinetic basins, namely, at least qualitatively, to different free energy wells. By clicking on one of the spheres, the structures belonging to the corresponding microstate are visualized like in the right panel. As several structures are normally assigned to the same microstate (as they have similar values of CVs) one sees many superimposed configurations. If the CV set used for the analysis and the simulation is appropriate, the structures will be similar, like in the example in Fig. 4b. If the structures are not similar, one can add other CVs or change CVs in order to make the set more uniform. One can easily toggle between the two representations, and eventually choose another microstate, repeating the check in all the relevant locations of the free energy surface. Moreover, if the free energy is constructed in more than three dimensions, one can easily change the variables used for the projection in the left panel, for example choosing the combination in which the free energy minima are better separated. In this manner one can

- navigate through a multi-dimensional free energy surface, easily finding the relevant free energy minima and visualizing which structures belong to each minimum;
- check the consistency of structures that are assigned to the same microstate. This allows reversibly verifying if the collective variables that are used for the metadynamics simulation and the analysis are appropriate, or "hidden" degrees of freedom are present.

Arriving to the graphic representation in Fig. 4 requires performing a series of tasks: (i) Loading the trajectory in atomic-coordinates space and in CV space. (ii) Finding the microstates, namely grouping together the structures according to their closeness in CV space. (iii) Computing the free energies of the microstates considering the effect of the history-dependent bias of metadynamics (if present). (iv) Clustering together the microstates in kinetic basins, the different "free energy wells" in Fig. 4. These tasks are performed using a graphic user interface (GUI) whose typical layout is represented in Fig. 1.

In the next sections, we will illustrate in detail the tool on a simple example, a bias exchange metadynamics [13] simulation of the Ala–Ala–Ala peptide (ALA3). The zwitterionic form is simulated at $T = 300$ K in vacuum ($\epsilon_r = 80$), using the amber03 force field [14] and the GROMACS package [15].

2. Plugin usage

2.1. Loading the trajectory

First of all, one must load a trajectory containing the coordinates of the system under investigation, and the corresponding trajectory of the collective variables. The name of these two files can be directly typed or browsed in the text boxes "trajectory file" and "COLVAR file" respectively in the "SETTINGS – Trajectories:" section. One can load as many trajectories and collective variable files as needed by clicking on "Add new line". This is particularly useful if the tool is used for analyzing multiple replica runs such as replica exchange metadynamics [16], multiple walkers metadynamics [12] or bias exchange [13]. The COLVAR files must be synchronized with the corresponding trajectory files. This means that each line in the COLVAR file should contain the value of the CVs in the configuration corresponding to each frame in the trajectory file. Importantly, one can include also variables that are not biased but are putatively interesting, as they possibly describe a relevant

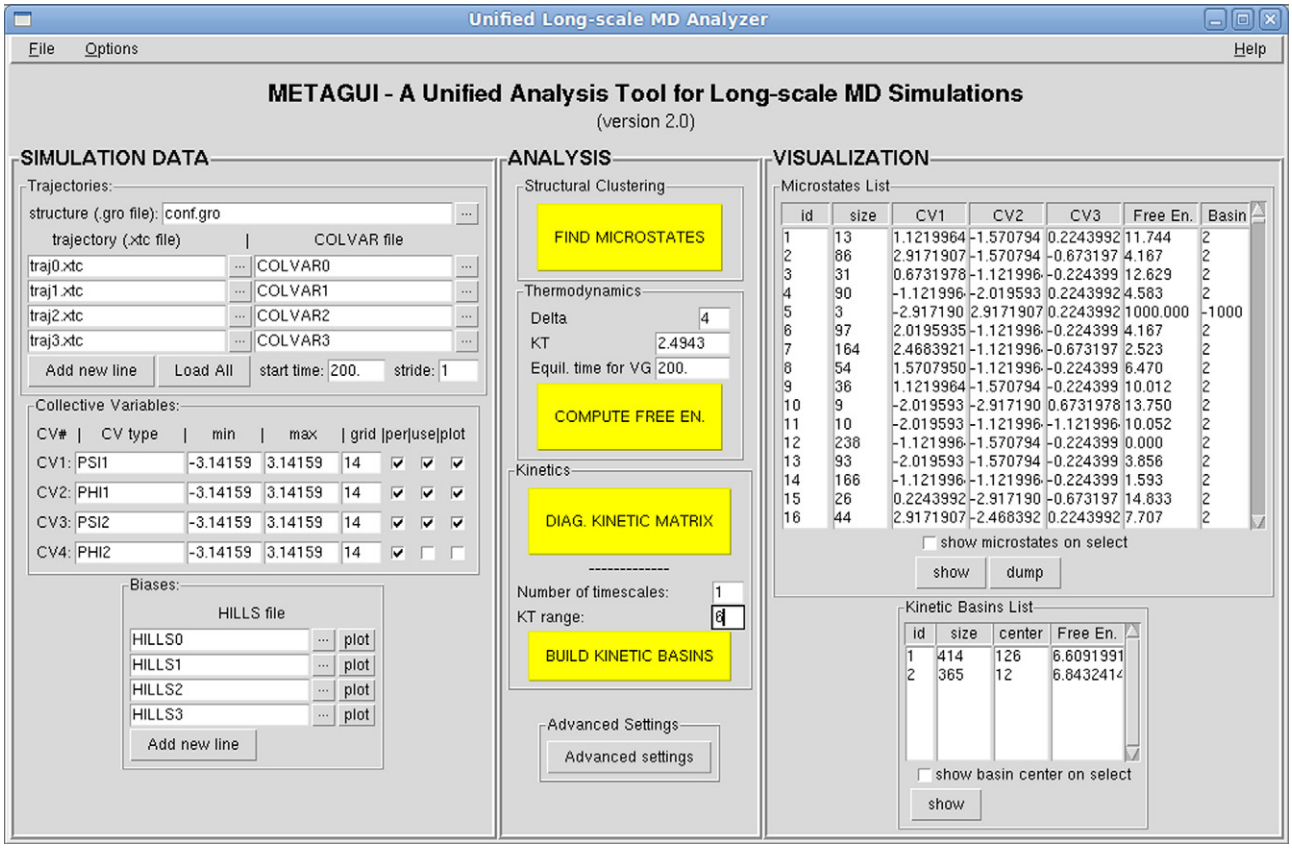

Fig. 1. The graphical user interface of METAGUI is splitted in three main blocks: (1) SIMULATION DATA: where all files coming out from the simulation are specified and loaded; (2) ANALYSIS: where the analysis tasks are executed (Structural Clustering, Free Energy computation, Diagonalization of the Rate Matrix and Representation of Kinetic Basins; (3) VISUALIZATION: where the output results of the analysis tasks are stored, mainly the list of microstates and kinetic basins with their respective populations and energies. (To distinguish between different colors in this figure, the reader is referred to the web version of this article.)

degree of freedom. In the case of analyzing metadynamics-based simulations, one must explicitly include at least the collective variables that are biased. In bias exchange [13], one typically uses N replicas, each biasing only one or two CVs. However, in each replica one monitors the value of all the N variables. As we will see, this information is essential for reconstructing the free energy in many dimensions. Thus, each COLVAR file will contain at least these N CVs, but possibly more putatively interesting variables. The format of this file is described in Appendix A.

The trajectory and COLVAR file/s are loaded directly into VMD by clicking on the "Load All" button or by typing the commands "load_data; load_trajs; " in the VMD shell. The starting time from which to load the data can be specified in the "start time:" text-box. Further, if graphical memory resources are limited, structures from the trajectory file can be loaded skipping the number of frames specified in the "stride:" text-box. Even in this case all the frames in the COLVAR files are loaded in memory for the analysis. After the data load is successful, the "SETTINGS – Collective Variables" is filled with the information extracted from the COLVAR files. The minimum and maximum values of each CV found in the COLVAR files is shown. One can specify here the name of each CV for reference ("CV type" box). One must also specify here if any of the CVs is periodic ("per" checkbox). The "use" and "plot" checkboxes are discussed below.

2.2. Specifying the bias

If the tool is used for analyzing metadynamics simulations, the history-dependent potentials must also be specified. These potentials are stored in the HILLS files. The name of each HILLS file can be typed or browsed in the text box "HILLS file". The HILLS files

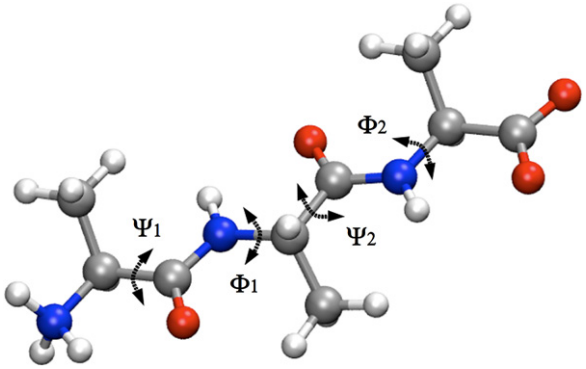

Fig. 2. The collective variables employed to simulate the ALA3 system. Each variable is biased on a different replica in a bias-exchange metadynamics simulation. (To distinguish between different colors in this figure, the reader is referred to the web version of this article.)

can be many, as in bias exchange metadynamics different replicas are biased on different CVs. In the case of the ALA3 example, we use 4 replicas, each biasing one CV, the value of the backbone dihedral angles of the peptide (see Fig. 2): Ψ_1 (N^1–C^1_α–C^1–N^2), Φ_1 (C^1–N^2–C^2_α–C^2), Ψ_2 (N^2–C^2_α–C^2–N^3), Φ_2 (C^2–N^3–C^3_α–C^3). Gaussian hills of height 1 kJ/mol and width $0.314 = \pi/10$ rad are deposited every 4 ps, while exchanges of the biases are attempted every 5 ps.

2.3. Finding the microstates

The next step that has to be performed is grouping together configurations in microstates, sets in which the value of the CVs

are similar. The GUI allows selecting easily the CVs used for this analysis from the set that is read from the COLVAR files as described above. This is done by checking the checkboxes "use" close to each collective variable in the "SETTINGS – Collective variables:" section. If a checkbox is on, the CV is used for the analysis. The configurations are grouped together in microstates simply by dividing the N-dimensional CV-space into a grid of small N-dimensional hypercubes. The size of the hypercube is defined by its side in each direction: $(ds_1, ds_2, \ldots, ds_N)$. This determines directly how far the hypercubes centers are, and, as a consequence, the typical difference between neighboring microstates. Each frame of the trajectory is assigned to the hypercube to which it belongs and the set of frames contained in a hypercube defines a microstate. The range and grid subdivision ds_i of each CV can be changed using the text-boxes "min", "max", and "grid" respectively. For ALA3, we use the backbone dihedral variables Ψ_1, Φ_1, Ψ_2, in Fig. 2 with a grid spacing of $2\pi/14 = 0.449$ over a range of 2π. The "per" checkbox is activated for all CVs, due to their periodicity.

By clicking on the "FIND MICROSTATES" button in the "ANALYSIS – Structural Clustering" section, or by typing the command "do_clusters; " in the VMD shell, the subdivision is performed and the first output file is written: MICROSTATES (see Appendix A). The result is automatically displayed in the "Microstates List" box in the "VISUALIZATION" section. The first column is the unique identifier of that microstate; the second column shows the population of that microstate, namely the total number of frames in the COLVAR file that are assigned to it; the next columns are the CV center of that microstate. One can sort the microstates according to their population, or the value of one CV by clicking on the appropriate column label.

The structures of the microstates can now be visualized directly on the VMD Graphic Display by clicking on the desired element in the "Microstates List". The structures of multiple microstates can be visualized at the same time by selecting multiple lines in the "Microstates List". Although not necessary for the analysis, the structures of any set of microstates can be written independently into PDB files for analysis/visualization with other software. This can be done by clicking on the "dump" button below the "Microstates List".

In a typical case, one should repeat the analysis using several combinations of CVs, in order to find the most appropriate combination in terms of structural consistency in the microstates (by visualizing the structures in the VMD Display as described) and in terms of statistical reliability (see below).

At this point one only knows the population of the microstates. This information would be enough in case of an unbiased molecular dynamics simulation in which no external potential is applied. Relative free energies of microstates could be directly computed from these populations. In the case of metadynamics these have to be corrected considering the effect of the bias.

2.4. Computing the free energy of the microstates

The tool is specifically designed for computing the free energy of microstates embedded in a high-dimensional collective variable space. In metadynamics the history-dependent potential provides an estimate of the low-dimensional projections of the free energy. In order to pass from low-dimensional projections to the free energy of multi-dimensional microstates we use the weighted-histogram technique (WHAM) described in Ref. [17]. Following the WHAM approach [18], the equilibrium probability of the microstate α is estimated as:

$$p^i_\alpha = n^i_\alpha e^{\beta(V^i_\alpha - f^i)} \tag{1}$$

where i is the replica index, f_i is a shift constant fixing the normalization, n^i_α is the number of times state α is observed in replica i and $V^i_\alpha = V^i(s_\alpha)$ is the bias potential evaluated on the microstate α. V^i is estimated from the time-average between the equilibration time of the bias potential t_{eq} (see Ref. [17]) and the total simulation time t_{tot} [17,19,20]:

$$V^i(s) = \frac{1}{t_{tot} - t_{eq}} \int_{t_{eq}}^{t_{tot}} dt\, V^i_G(s,t) \tag{2}$$

where $V^i_G(s,t)$ is the metadynamics bias at time t, computed from the sum of Gaussians specified in the appropriate HILLS file. In Eq. (1) we have not considered corrections due to the variations of $V^i(s)$ within structures assigned to the same microstate. These corrections are described in Ref. [17], and are used in the METAGUI, but lead to a formally identical expression for the free energy and are not introduced here in order to simplify the notation.

In estimating the error on p^i_α we here consider not only the standard statistical error deriving from the finite number of observations [18], but also the error $\sigma^2(V^i)$ on V^i, that in metadynamics is a fluctuating quantity. $\sigma^2(V^i)$ is here assumed to be equal to the mean square difference of two different time averages of $V^i_G(s,t)$ in the two intervals $(t_{eq}, \frac{t_{eq}+t_{tot}}{2})$ and $(\frac{t_{eq}+t_{tot}}{2}, t_{tot})$. Using error propagation on Eq. (1), we have

$$\sigma^2(p^i_\alpha) = \left[\sigma^2(n^i_\alpha) + \beta^2\sigma^2(V^i)(n^i_\alpha)^2\right] e^{2\beta(V^i_\alpha - f^i)}$$

Using $\sigma^2(n^i_\alpha) = g n^i_\alpha$, where g is a constant that takes into account the correlation time [17], we find

$$\sigma^2(p^i_\alpha) = \gamma^i_\alpha p^i_\alpha e^{\beta(\bar{V}^i_\alpha - f^i)} \cong \gamma^i_\alpha p_\alpha e^{\beta(\bar{V}^i_\alpha - f^i)} \tag{3}$$

where

$$\gamma^i_\alpha = g + \beta^2 n^i_\alpha \sigma^2(V^i)$$

The p^i_α-s are then combined in a single estimate of the probability following the standard WHAM procedure. This leads to

$$p_\alpha = C \frac{\sum_i \frac{1}{\sigma^2(p^i_\alpha)} p^i_\alpha}{\sum_i \frac{1}{\sigma^2(p^i_\alpha)}} = C \frac{\sum_i \frac{1}{\gamma^i_\alpha} n^i_\alpha}{\sum_i \frac{1}{\gamma^i_\alpha} e^{\beta(f^i - \bar{V}^i_\alpha)}} \tag{4}$$

where C is a normalization constant. The constants f^i are determined self-consistently like in Ref. [17]. Finally, the free energy of microstate α is given by the usual formula $F_\alpha = -k_B T \log p_\alpha$.

Clicking on the button "COMPUTE FREE EN.", or by typing the command "run_wham; " in the VMD shell, the free-energy of each microstate is estimated as explained above and the result is automatically updated in the MICROSTATES file and in the "ANALYSIS – Microstates List" section.

At this point, there are several checks that should be performed in order to assess the reliability of the results. In metadynamics, the accuracy of the free energy estimates is mainly affected by the reliability of the bias potentials V^i in Eq. (2). If the simulation is converged, the bias potential $V^i_G(s,t)$ is an unbiased estimator of the free energy, namely it fluctuates around an average value, $-F(s)$, with ripples whose size is determined by the metadynamics parameters. In these conditions the time average in Eq. (2) is meaningful, as $V^i_G(s,t)$ is stationary. The value of the parameter t_{eq} in Eq. (2) is specified in the text-box `EQUIL.time`. In order to check if the average in Eq. (2) is converged, one uses the button "plot" next to each HILLS file name in the "SETTINGS – Biases" category. A typical graphic output is presented in Fig. 3 for ALA3. The red and the blue curve represent respectively the time average of $V^i_G(s,t)$ in the two intervals $(t_{eq}, \frac{t_{eq}+t_{tot}}{2})$ and $(\frac{t_{eq}+t_{tot}}{2}, t_{tot})$. Ideally, if the free energy estimator is stationary, these two curves

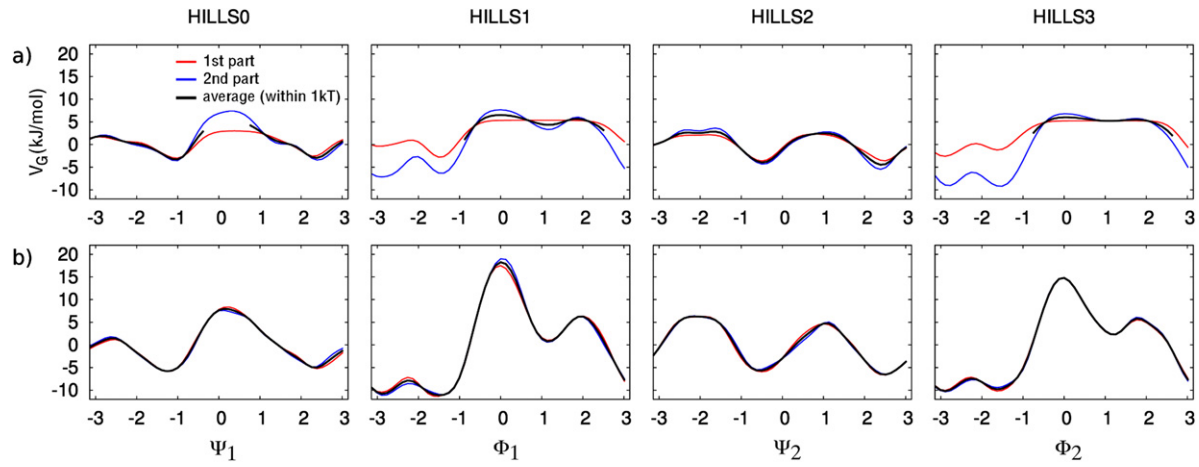

Fig. 3. Convergence of the bias profiles $V_G^i(s,t)$ for ALA3, as a function of the backbone dihedral angles. The red profile represents the time average within $(t_{eq}, \frac{t_{eq}+t_{tot}}{2})$, the blue one within $(\frac{t_{eq}+t_{tot}}{2}, t_{tot})$, and the black thick line, averaging over red and blue profiles which are consistent within $1k_BT$, is the best estimate for the free energy. Two cases are shown: a) filling time $t_{eq} = 10$ ps, total time $t_{tot} = 300$ ps: the statistics is poor and the profiles are not converged; b) $t_{eq} = 200$ ps, $t_{tot} = 8000$ ps: a well-converged simulation. (To distinguish between different colors in this figure, the reader is referred to the web version of this article.)

should be similar. The parameter `DELTA` ($= 1k_BT$ in Fig. 3) allows specifying the maximum allowed difference that two different free energy estimates can take in order to be considered reliable. The code automatically attempts aligning the two profiles, maximizing the size of the region in which the two profiles differ by less than `DELTA`. The error on V_G, namely the average standard deviation of the two profiles, will be by construction smaller than `DELTA` (normally significantly smaller). As a further constraint, the region on which the profiles are aligned must be continuous, as free energy estimators are reliable only within connected regions. The result of this procedure is represented as a black thick line in Fig. 3. Only the frames falling within the black region are retained for further analysis. For these frames, the equilibrium population is estimated using Eq. (1) with the bias potential given by Eq. (2). The two parameters `EQUIL.time` and `DELTA` should be varied until one is able to obtain, for all the relevant biased collective variables, a free energy estimate on a sufficiently wide region and with a sufficiently small difference between the blue and the red curve. Fig. 3a shows an example in which the bias potential is badly converged, due to insufficient statistics (too short t_{eq} and t_{tot}). We also show a case in which convergence is excellent (Fig. 3b). In this case the red and blue profiles of each HILLS file are very similar to each other (within less than DELTA) over all the range of each CVs.

2.5. Finding the kinetic basins

In the next step METAGUI allows performing a kinetic clustering of the hundreds of microstates whose free energy is estimated according to the procedure described above. The microstates space is subdivided in metastable sets ("kinetic basins"), normally corresponding to significant local free-energy minima. This is accomplished by constructing an approximate rate matrix among these microstates and analyzing its spectrum [21]. By construction, the typical transition time between two microstates will be much smaller when they belong to the same kinetic basin than when they belong to different basins. The transition rate between microstate α and β is assumed to be of the form [22]

$$k_{\alpha\beta} = \chi_{\alpha\beta} k^0_{\alpha\beta} e^{-\frac{1}{2}(F_\beta - F_\alpha)/k_BT} \tag{5}$$

where $k^0_{\alpha\beta} = k^0_{\beta\alpha}$ are the rates associated to simple diffusion on a flat free energy surface and $\chi_{\alpha\beta} = 1$ only if the microstates are neighbors, 0 otherwise. In d dimensions the microstates are labeled by d integers $(i_1, i_2, \ldots, i_d)$. If D is diagonal, the only rates differing from zero are those in which one of the components of $(i_1, i_2, \ldots, i_d)$ vary by one:

$$k^0_{(\ldots,i_k,\ldots)(\ldots,i_k\pm1,\ldots)} = \frac{D_{kk}}{ds_k^2} \tag{6}$$

where ds_k is the grid spacing. If D is non-diagonal, the expression for k^0 is more complicated and is reported in Ref. [17]. As a default, the tool simply takes $D_{ij} = \delta_{ij}\, ds_i\, ds_j$. If the diffusion matrix of the system is known one can insert its values in the corresponding text-boxes accessible by clicking the "Advanced settings" button.

By clicking on the "DIAG. KINETIC MATRIX" button, or by typing the command "run_basins; " in the VMD shell, the calculation starts. First, the code finds the largest subset of microstates that are connected in the multi-dimensional CV-space. Two microstates are connected if they are neighbors in the grid and their relative free energy difference is known. This preliminary step is necessary as one can look for a decomposition in kinetic basins only on a connected region. As outlined in the previous Section, the biases are considered reliable only when converged in a connected region – black line in Fig. 3. However, the WHAM procedure can at times provide converged free energies in physically separated regions, that look superimposed upon projection on one variable.

As a second step, the spectrum of the rate matrix in Eq. (5) is computed by diagonalization. By default, only the first 15 eigenvectors are computed in order to allow a fast matrix diagonalization. The eigenvalues and eigenvectors of this matrix provide direct information about the subdivision of the system in kinetic basins [21]. The relaxation times τ_i of the system are the inverse of the eigenvalues λ_i of this matrix. One eigenvalue is by construction equal to zero, as the rate matrix Eq. (5) satisfies detailed balance. Once the calculation is completed, the relaxation times are plotted in decreasing order and in logarithmic scale in a graphic window that pops out when the program terminates. The user can then chose in how many relaxation times to represent the splitting of the microstates space considering the subdivision in kinetic basins ("Number of timescales" text box in the "ANALYSIS – Kinetics" section). As explained above, a reasonable choice would be to include preferentially the first τ_i relaxation times until a big separation (one or two orders of magnitude) is observed with the following relaxation time (τ_{i+1}).

After the user chooses the number of timescales, the program finds the attractor of each kinetic basin by looking at the components distribution in each eigenvector, like in Ref. [21]. Each

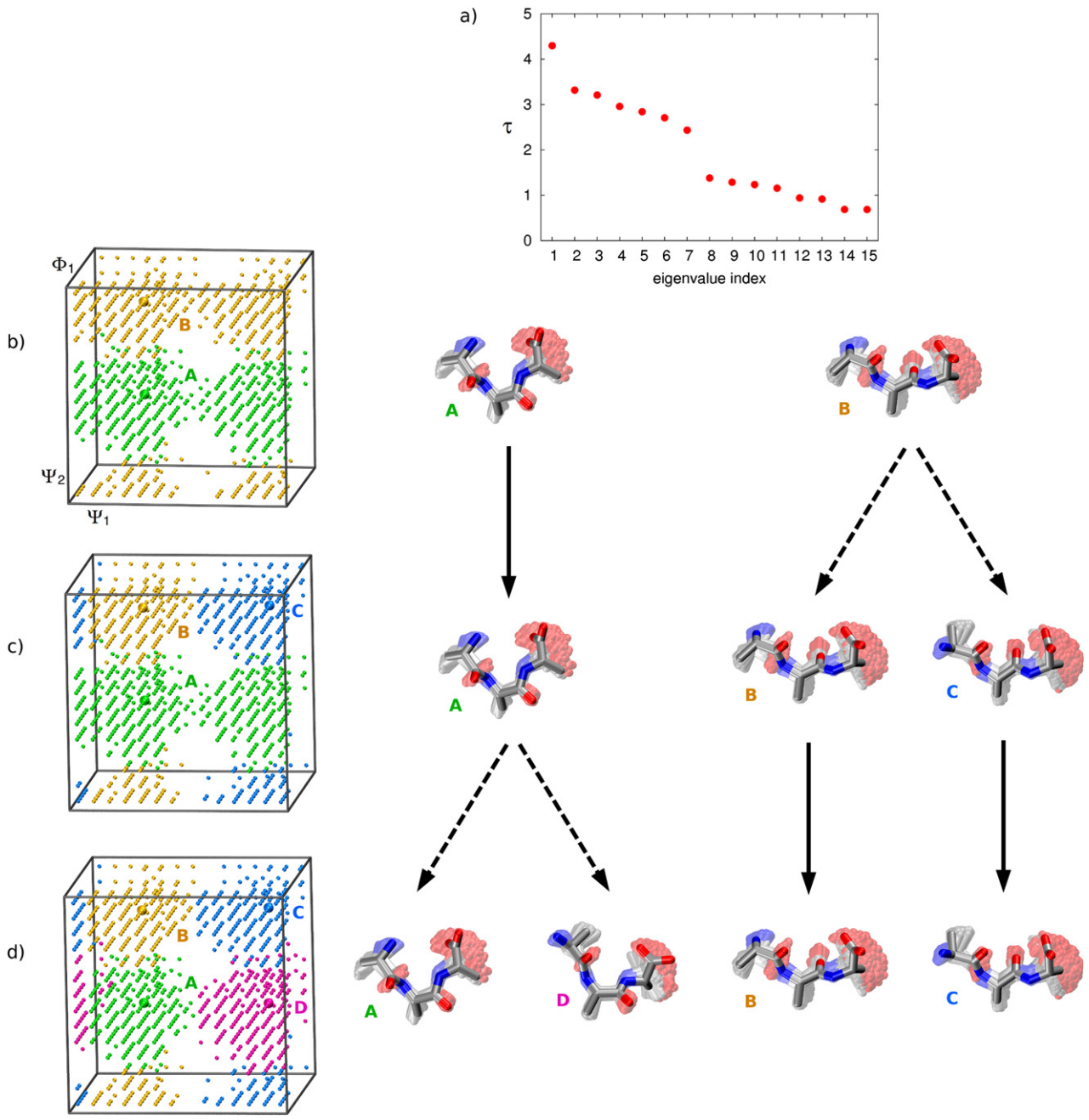

Fig. 4. Kinetics of the ALA3 system: a) relaxation times from the diagonalization of the kinetic matrix; b) kinetic basins corresponding to the largest relaxation time as a function of three CVs, in different colors, and atomic structures of the attractors (lowest free-energy microstates) of the basins; c) the same for the two largest relaxation times, and d) for the three largest. Basins A, B, C, and D have the same free energy within 1 kJ/mol. (To distinguish between different colors in this figure, the reader is referred to the web version of this article.)

microstate is then assigned to an attractor by performing a commitor analysis, namely running a few hundred trajectories starting from the microstate with a dynamics driven by the rates in Eq. (5). Each trajectory is ended when it reaches one of the attractors. The microstate is assigned to the kinetic basin of the attractor that has been reached more times. We found that this procedure is numerically more stable than the one proposed in Ref. [21], where the separation in kinetic basins is fully derived from the eigenvectors of the rate matrix.

In Fig. 4a the relaxation times for ALA3 are reported (obtained by restricting the CV-space to Ψ_1, Φ_1, and Ψ_2): the first τ_1 value is separated from the following ones by a gap of about one order of magnitude. This allows to identify two main kinetic basins. The kinetic basins subdivision of the system can be graphically represented as a projection on the CV space of the microstates. This is done by clicking the "Build basins" button. The microstates are plotted on the VMD Display window, represented as spheres, and their three-dimensional coordinates correspond to three CVs selected in the "plot" checkboxes next to each collective variable in the "SETTINGS – Collective Variables" section. The microstate spheres are colored according to the kinetic basin they belong to. The microstate of lowest free energy (attractor) in each kinetic basin is represented as a bigger sphere. In Fig. 4b the two main basins of ALA3 are displayed employing the CVs Ψ_1, Φ_1, and Ψ_2: the space is splitted along the Ψ_2 direction in a α-like basin (A) and a β-like basin (B). The following information appears in

the "Basins List" in the "VISUALIZATION – Kinetic" section of the METAGUI interface: a basin identifier (first column); the number of microstates belonging to that basin (second column); the identifier of the attractor, namely the microstate of minimum free energy of that basin (third column); the average free energy of the basin (fourth column). The plugin allows now an efficient navigation through the microstates space directly on the VMD Display window. By hitting the "1" key (atom picking mode) inside the VMD Display, one can select a state (sphere) from the displayed ones by left-clicking with the mouse. The atomic configurations corresponding to the selected microstate then appear. Fig. 4b shows the atomic structures of the attractor microstates of each of the two main kinetic basins of ALA3: it is clear that within each microstate the atomic configurations are structurally consistent among them. To return to the kinetic basins splitting in CV-space representation it is sufficient to hit the "F" key in the VMD Display window. Figs. 4c, d show the effect of considering the additional relaxation times τ_2 and τ_3 in the kinetic analysis of ALA3: the two main basins split in sub-basins along Ψ_1 in a hierarchical way, providing a more fine-grained kinetic picture of the system.

We remark that in practical applications it is important to assess the convergence of the kinetic model by checking its robustness with respect to variations of the relevant parameters, namely the length of the molecular dynamics trajectories (20 ns are sufficient for ALA3), the number and type of CVs employed to estimate the free-energy, the size of the grid in CV-space, the minimum population of each microstate.

3. Conclusions

We presented a new computer tool called METAGUI aimed at analyzing metadynamics and long-scale molecular dynamics simulations. The tool extends the popular molecular visualization program VMD [8] and relies on a graphical user interface (GUI) that allows building a microstates-based model of the system that includes equilibrium thermodynamics (free-energy in a multi-dimensional space) and kinetics (matrix of transition rates among microstates, and kinetic basins). Allowing the construction of a handy model of the system, METAGUI facilitates the analysis and interpretation of large-scale simulations on complex systems. The tool collects together several different functionalities which would otherwise force the user to employ distinct codes: clustering of trajectories, weighted-histogram analysis, construction and diagonalization of kinetic matrices, dual visualization of microstates in CV-space and molecular structures in Cartesian coordinates.

In this paper we illustrated the METAGUI tool using a very simple example of a conformational transition in ALA3, but potential applications range over the full spectrum of free-energy calculations including chemical reactions, phase transitions, molecular recognition events and others. As an example, we provide in Fig. 5a the output of the analysis of the Trp-cage folding simulation described in Ref. [17]. The results, in terms of number and nature of kinetic basins, are in full agreement with those reported in the original manuscript, but can now be derived in a straightforward and transparent manner with METAGUI.

The METAGUI can also be used as a preparative tool for setting-up free energy calculations, especially when defining the number and type of collective variables that will drive the dynamics of the system. By loading a large set of structures generated by whatever exploration method (e.g. molecular dynamics performed at high temperature) one can easily perform a preliminary microstates clustering of these structures in different sets of CVs. The Cartesian structures corresponding to these microstates can be directly visualized by the METAGUI on the VMD display as explained above.

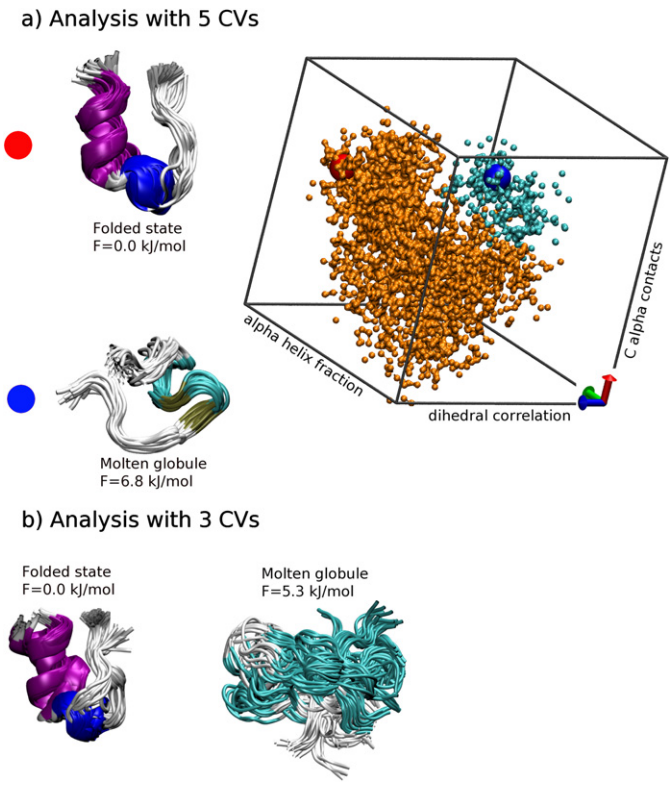

Fig. 5. a) Kinetics of the Trp-cage system. 4300 bins were obtained dividing the multi-dimensional space in 5 CVs. This result is in agreement with the kinetic analysis of Ref. [17] in which a molten globule structure was find to act as a kinetic trap. b) The same analysis performed using only 3 CVs produced structurally inhomogeneous microstates. (To distinguish between different colors in this figure, the reader is referred to the web version of this article.)

Ideally, for an efficient simulation, one should use a set of CVs in which the structures within a given microstate are very similar. As an example, in Fig. 5b we report for the Trp-cage example how the two relevant microstate would look like for an analysis performed with a "wrong" set of CVs, in which one of the relevant slow variables is not included. If an output of this form is obtained, it is necessary to analyze the structures assigned to the microstate in order to find out which additional variable can provide a splitting into structurally-consistent states.

The METAGUI currently supports the analysis of metadynamics [9,19], in single replica, multiple walker [12] and bias-exchange [13] variants. Of course the tool can also be used for analyzing single – and multiple replicas unbiased molecular dynamics simulations, since these can be considered as a metadynamics performed with no bias. Future developments of METAGUI will go in the direction of enlarging the spectrum of free-energy methods supported, including also umbrella sampling [23] and well-tempered metadynamics [24].

The tool is based on the output format of PLUMED [25], which makes it directly compatible with a large number of molecular dynamics codes like NAMD, GROMACS, SANDER, LAMMPS, DLPOLY, QUANTUM-ESPRESSO, and more. The METAGUI source files, together with the ALA3 example shown here, can be downloaded from the PLUMED website (http://www.plumed-code.org).

Acknowledgements

The authors acknowledge Davide Branduardi for useful discussions. X.B. acknowledges financial support from a Beatriu de Pinós fellowship (BP2007-A) granted by the Generalitat de Catalunya.

Appendix A

A.1. Input data files format

We here describe the format of the input files that have to be provided to the GUI, and of the output files that are created. Three different types of data files must be provided to the GUI:

- TRAJECTORY: containing the temporal evolution of the coordinates of the system.
- COLVAR: containing the temporal evolution of the CVs.
- HILLS: containing the metadynamics history-dependent potentials.

TRAJECTORY

METAGUI can read trajectories in any file format supported by VMD.

COLVAR

METAGUI expects to read COLVAR files compatible with the output format of PLUMED v1.3. Since the format of these files may change from one version of PLUMED to another, it is recommended to adjust the file format of the COLVAR files to the one described here. An allowed input format of this file is as follows:

```
#! FIELDS time cv1 cv2 cv3 cv4 vbias
#! ACTIVE 1 2 B
0.0  2.263 -2.742 -0.099  2.991 0.0000
2.0 -1.090 -2.790 -0.683 -1.556 0.0000
4.0 -1.313 -2.933 -0.511 -1.788 1.0000
 ..................
```

The first line describes the content of each field in the COLVAR file. This line contains information on the number of CVs. It is assumed that the number of CVs is equal to the number of times the string "cv" appears. In the case of the example, there are four CVs, cv1, cv2, cv3 and cv4. If several COLVAR files are loaded, they must contain the same CVs in the same order. The second line describes the metadynamics bias potential (if any) acting on the following frames. The bias is identified by the number of active CVs and their index (starting from 1): in the example 1 and 2, respectively, together with a text label (B in the example). Therefore the bias is acting on cv2. The tool is constructed in order to allow analyzing simulations in which different parts of the trajectories are biased by different history-dependent potentials. Whenever the history-dependent potential changes, this line (#! ACTIVE) is repeated again with the information on the new bias. IMPORTANT: the COLVAR files must be synchronized with the corresponding trajectory files, this means that each line in the COLVAR file should contain the value of the CVs in the configuration corresponding to each frame in the trajectory file.

HILLS

The METAGUI reads history-dependent potentials files compatible with the output format of PLUMED v1.3. Since the format of these files may change from one version of PLUMED to another, it is recommended to adjust the file format of the HILLS files to the one described here. An allowed input format of this file is as follows:

```
 #! ACTIVE 1 2 B
4.000  -2.933 0.314 1.0000
8.000  -2.791 0.314 1.0000
12.000 -2.206 0.314 1.0000
 ..................
```

The header of the HILLS files informs on which CVs the bias is acting on, with the same format as in the COLVAR files above. The bias is identified by the last text label in the line.

A.2. Output files format

MICROSTATES

This file contains a list of microstates. The format of this file is as follows:

```
1 1  2.945 -0.982 -0.589 -1.374    6.666     2
2 3  0.196 -1.374 -0.982 -1.767   12.375 -1000
3 3 -1.374 -2.945  0.196 -2.945    4.396     2
4 2  2.160 -1.374  0.196 -2.160    4.045     2
5 1  1.374 -2.945 -0.589 -1.374    8.436     2
6 1  2.553 -1.767 -0.589 -1.767 1000.000 -1000
7 1 -0.982 -1.767 -0.589  2.553    5.548     2
 ..................
```

The first column is the unique identifier of that microstate; the second column shows the population of the microstate, namely the total number of trajectory frames assigned to it; the next columns are the CV values of the center of the microstate; the second last column contains the free energy of the microstate. Note that some states may have a free energy of 1000: this means that due to poor statistics the free energy could not be assigned. The last column is the identifier of the kinetic basin to which the microstate belongs. Note that some states my have not been assigned to any basin: in that case the identifier is -1000

BASINS

This file lists the kinetic basins in which the space has been divided. The format of this file is as follows:

```
1   15 12.574 4931  2.945 -0.982 2.160 -2.160  5.868
2 3407  6.403 1582 -1.374 -1.374 2.553 -1.374  0.000
3   84 16.589 1866 -1.374  0.982 0.196 -2.945 11.500
```

The first column contains the basin identifier. The second column contains the number of microstates belonging to that basin. The third column shows the free energy of the whole kinetic basin. The fourth column shows the identifier of the microstate of minimum free energy belonging to that basin: this is the kinetic basin attractor. The following columns show the CV coordinates of the attractor. The last column stores the free energy of the attractor.

References

[1] M. Karplus, J. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct. Biol. 9 (2002) 788.

[2] R. Zhou, Trp-cage: Folding free energy landscape in explicit water, Proc. Natl. Acad. Sci. USA 100 (2003) 13280–13285.

[3] G. Jayachandran, V. Vishal, V. Pande, Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece, J. Chem. Phys. 124 (2006).

[4] J. Juraszek, P.G. Bolhuis, Sampling the multiple folding mechanisms of Trp-cage in explicit solvent, Proc. Natl. Acad. Sci. USA 103 (2006) 15859–15864.

[5] J.L. Klepeis, K. Lindorff-Larsen, R.O. Dror, D.E. Shaw, Long-timescale molecular dynamics simulations of protein structure and function, Curr. Opin. Struct. Biol. 19 (2009) 120–127.

[6] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers, Atomic-level characterization of the structural dynamics of proteins, Science 330 (2010) 341–346.

[7] P. Cossio, F. Marinelli, A. Laio, F. Pietrucci, Optimizing the performance of bias-exchange metadynamics: folding a 48-residue lysm domain using a coarse-grained model, J. Phys. Chem. B 114 (2010) 3259.

[8] W. Humphrey, A. Dalke, K. Schulten, VMD – visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.

[9] A. Laio, M. Parrinello, Escaping free energy minima, Proc. Natl. Acad. Sci. USA 99 (2002) 12562–12566.

[10] S. Piana, A. Laio, F. Marinelli, M.V. Troys, D. Bourry, C. Ampe, J.C. Martins, Predicting the effect of a point mutation on a protein fold: The villin and advillin headpieces and their Pro62Ala mutants, J. Mol. Biol. 375 (2008) 460–470.
[11] F. Pietrucci, A. Laio, A collective variable for the efficient exploration of protein beta-sheet structures: application to sh3 and gb1, J. Chem. Theory Comput. 5 (2009) 2197.
[12] P. Raiteri, A. Laio, F. Gervasio, C. Micheletti, M. Parrinello, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, J. Phys. Chem. B 110 (2006) 3533–3539.
[13] S. Piana, A. Laio, A bias-exchange approach to protein folding, J. Phys. Chem. B 111 (2007) 4553–4559.
[14] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.
[15] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: a package for molecular simulation and trajectory analysis, J. Mol. Mod. 7 (2001) 306–317.
[16] G. Bussi, F.L. Gervasio, A. Laio, M. Parrinello, Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics, J. Am. Chem. Soc. 128 (2006) 13435–13441.
[17] F. Marinelli, F. Pietrucci, A. Laio, S. Piana, A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations, PLoS Comput. Biol. 5 (2009) e100045.
[18] S. Kumar, D. Bouzida, R. Swendsen, P.A. Kollman, J. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method, J. Comput. Chem. 13 (1992) 1011–1021.
[19] A. Laio, F.L. Gervasio, Metadynamics: a model to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, Rep. Progr. Phys. 71 (2008) 126601.
[20] Y. Crespo, F. Marinelli, F. Pietrucci, A. Laio, Metadynamics convergence law in a multidimensional system, Phys. Rev. E 81 (2010) 055701(R).
[21] F. Noe, I. Horenko, C. Schuette, J.C. Smith, Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states, J. Chem. Phys. 126 (2007) 155102.
[22] G. Hummer, Position-dependent diffusion coefficients and free energies from bayesian analysis of equilibrium and replica molecular dynamics simulations, New J. Phys. 7 (2005) 34.
[23] S. Kumar, P.W. Payne, M. Vásquez, Method for free-energy calculations using iterative techniques, J. Comput. Chem. 17 (1996) 1269–1275.
[24] A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method, Phys. Rev. Lett. 1 (2008) 020603.
[25] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, M. Parrinello, Plumed: a portable plugin for free-energy calculations with molecular dynamics, Comp. Phys. Comm. 180 (2009) 1961.