Characterizing Energy Landscapes of Peptides using a Combination of Stochastic Algorithms

Didier Devaurs*†§, Kevin Molloy*†, Marc Vaisset*†, Amarda Shehu‡, Thierry Siméon*†, Juan Cortés*†

*CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France
†Univ de Toulouse, LAAS, F-31400 Toulouse, France
‡Department of Computer Science, George Mason University, Fairfax, VA 22030, USA
§Current address: Department of Computer Science, Rice University, Houston, TX 77005, USA

Abstract—Obtaining accurate representations of energy landscapes of biomolecules such as proteins and peptides is central to the study of their physicochemical properties and biological functions. Peptides are particularly interesting, as they exploit structural flexibility to modulate their biological function. Despite their small size, peptide modeling remains challenging due to the complexity of the energy landscape of such highly-flexible dynamic systems. Currently, only stochastic sampling-based methods can efficiently explore the conformational space of a peptide. In this paper, we suggest to combine two such methods to obtain a full characterization of energy landscapes of small yet flexible peptides. First, we propose a simplified version of the classical Basin Hopping algorithm to reveal low-energy regions in the landscape, and thus to identify the corresponding meta-stable structural states of a peptide. Then, we present several variants of a robotics-inspired algorithm, the Transition-based Rapidly-exploring Random Tree, to quickly determine transition path ensembles, as well as transition probabilities between meta-stable states. We demonstrate this combined approach on met-enkephalin.

Index Terms—energy landscape; peptides; stochastic algorithms

I. Introduction

Global thermodynamic and kinetic properties of molecules can be extracted from an analysis of their conformational energy landscapes [1]. In particular, obtaining an accurate representation of a molecule's energy landscape is a significant first step to conducting detailed structure-function studies for bio-molecules of central importance in the cell, such as proteins and peptides [2].

In this work, we focus on small peptides. Despite their modest size, they represent in many ways a more challenging setting than larger proteins. Peptides exhibit high structural flexibility, which enables them to recognize different molecular partners in the cell, and thus to modulate their biological function [3]. Contrary to proteins, that are often characterized by a unique native state and a funnel-shaped energy landscape, peptides are characterized by several meta-stable structural states; their energy landscape may contain a multitude of competitive low-energy basins.

The existence of multiple local minima in a molecule's energy landscape makes it particularly challenging to map this landscape and reconstruct all the functionally-important regions in it. Experimental methods, such as X-ray crystallography or nuclear magnetic resonance (NMR), cannot reveal

such maps, as they can uncover only few structures at best [4]. It is therefore the task of computational techniques to obtain detailed representations of energy landscapes.

Currently, only sample-based representations of the energy landscape can be afforded. Even for small peptides, the space of possible conformations is vast, and the effective degrees of freedom needed to represent a conformation are numerous. The high dimensionality of the space is accompanied by a complex (non-linear, non-convex) expression for the conformational energy, which is the result of competing local and non-local inter-atomic interactions. Probing this landscape is therefore very computationally-costly. Currently, only stochastic optimization techniques provide the right balance between accuracy and computational efficiency [1], [2].

Obtaining a representation of an energy landscape can be divided into two sub-problems: (1) determining meta-stable structural states (i.e. local energy minima); (2) computing transition paths between the identified states. Both can be addressed by achieving an effective sampling of the conformational space. In this paper, we propose to combine two sampling-based techniques to obtain a full characterization of energy landscapes of small yet highly-flexible peptides. First, we present a simple variant of the Basin Hopping algorithm [5] to sample local minima in a peptide's energy landscape. Local minima are then organized via density-based clustering to reveal meta-stable structural states. Second, we present several variants of a robotics-inspired algorithm, the Transition-based Rapidly-exploring Random Tree (T-RRT) [6], to map out the connectivity between these states, thus completing the reconstruction of the peptide's energy landscape. In particular, we propose a new variant of T-RRT allowing the computation of transition path ensembles and transition probabilities between meta-stable structural states. In a preliminary version of this paper [7], we presented proof-of-concept results on a minimalist peptide: the terminally-blocked alanine. Here we present results on a larger and much more challenging system: met-enkephalin.

II. Methods

The work presented in this paper is motivated by recent studies showing that robotics-inspired sampling-based algorithms provide a good basis for efficient conformational exploration in computational structural biology [8], [9]. The

Transition-based RRT (T-RRT) algorithm is an example of such algorithms [6], [10]. It is based on the Rapidly-exploring Random Tree (RRT) [11], a popular path planning algorithm that can tackle complex problems in high-dimensional spaces. RRT has been successfully used in various disciplines, such as robotics, aerospace, computer animation, and computational structural biology. T-RRT is an extension of RRT involving a probabilistic transition test based on the Metropolis criterion. Like Metropolis Monte Carlo (MC) methods [12], it applies small moves to the molecular system; but, instead of generating a single path over the search space, it constructs a tree, providing a more efficient exploration. Moreover, the tree construction is intrinsically biased toward unexplored regions of the space, and favors expansions in low-energy areas.

This work is partly based on an existing variant of T-RRT, called Multi-T-RRT [13], and on a new extension to it, called Multi-T-RRT with Cycles. These algorithms are combined with an in-house variant of the Basin Hopping algorithm involving a simplified minimization process. Together they are used to obtain a complete representation of the energy landscape of highly-flexible peptides. Details of these methods are presented in the next sections.

A. Basin Hopping

The Basin Hopping (BH) algorithm is a popular method for sampling local minima of an energy landscape. It was originally introduced to obtain the Lennard-Jones minima of small atom clusters [5]. Recently, BH has gained new attention to predict protein structure [14], and to find intermediate structures of chemical reactions [15]. The method consists of repeatedly applying a structural perturbation followed by an energy minimization, which yields a trajectory of minima. The result is a (discrete) coarse-grained representation of the energy landscape that can be seen as a collection of interpenetrating staircases.

Our implementation of BH (presented in Algorithm 1) differs from the classical one in that it does not involve local, gradient-based minimizations, but relies on simple Monte-Carlo-based (MC-based) minimizations. The algorithm follows a random restart procedure performing several rounds, each one starting from a conformation randomly sampled in the search space. Every round builds a trajectory of minima by performing a succession of structural perturbations followed by MC-based minimizations. Every MC-based minimization starts from a conformation obtained by performing a large-amplitude perturbation of the minimum reached at the previous step, or from the random sample, in the first step. An MC-based minimization is an iterative succession of small-amplitude perturbations. At each iteration, the perturbed conformation replaces the previous one if the Metropolis criterion is satisfied. More precisely, a downhill move in the energy landscape is always accepted. An uphill move is accepted or rejected based on the probability $e^{-(E_j - E_i)/(K \cdot T)}$ (where K is the Boltzmann constant), which decreases exponentially with the energy variation $E_j - E_i$ for a given temperature T, where E_i and E_j are the energies of the old and the perturbed state, respectively. Every MC-based minimization produces a

---

Algorithm 1: Basin Hopping

> **input** : the conformational space $\mathcal{C}$
> the number of rounds $nbRounds$
> the number of Monte Carlo minimizations $nbMC$
> **output**: the list of trajectories of minima $\mathcal{L}$

1   $\mathcal{L} \leftarrow \phi$
2   **for** $r = 1..nbRounds$ **do**
3     $\mathcal{T} \leftarrow \phi$
4     $q \leftarrow$ sampleRandomConformation$(\mathcal{C})$
5     $q_p \leftarrow q$
6   **for** $m = 1..nbMC$ **do**
7   **if** $m > 1$ **then**
8         $q \leftarrow$ doLargeAmplitudePerturbation(q_p)
9       $q \leftarrow$ doMonteCarloBasedMinimization(q)
10   **if** MetropolisTest(q, q_p) **then**
11   addMinimum$(\mathcal{T}, q)$
12        $q_p \leftarrow q$
13   addTrajectory$(\mathcal{L}, \mathcal{T})$
14   **return** $\mathcal{L}$

---

low-energy conformation that we call a *minimum* in a minor abuse of language. It is compared to the minimum obtained in the previous iteration, and accepted or rejected also based on the result of a Metropolis-like transition test. Different temperatures can be used for the transition test inside the local energy minimization procedure and for the one at each iteration of BH (usually, the former being much lower than the latter). Every round produces what we call a *milestone*: the minimum (along the trajectory) having the lowest energy. Note that, in order to speedup computation, a round can be stopped based on a consecutive number of rejections, similarly to the MC-based minimization procedure.

All the milestones (or the minima) produced by BH have to be grouped to provide a comprehensible list of meta-stable structural states. This clustering can be done in several ways. In this work, we have applied a density-based clustering technique that has been shown to provide good results for small peptides [16].

B. Multi-T-RRT with Cycle-Addition

The Rapidly-exploring Random Tree (RRT) algorithm [11] is a well-known path planning method in robotics. It can deal with complex problems by performing an efficient exploration, even in high-dimensional search spaces. Starting from an initial conformation q_{init}, RRT iteratively constructs a tree $\mathcal{T}$ that tends to rapidly expand over the conformational space $\mathcal{C}$. The nodes and edges of $\mathcal{T}$ correspond to states (i.e. molecular conformations) and small-amplitude moves between states, respectively. At each iteration of the tree construction, a conformation q_{rand} is randomly sampled in $\mathcal{C}$. Then, an extension toward q_{rand} is attempted, starting from its nearest neighbor q_{near}, in $\mathcal{T}$. This means performing a linear interpolation between q_{near} and q_{rand}, at a distance equal to the extension step-size δ, from q_{near}. If the extension succeeds, a new conformation q_{new} is added to $\mathcal{T}$ and an edge is built between q_{near} and q_{new}. The criteria on when to stop the exploration can be reaching a given target conformation q_{goal}, a given

---

Algorithm 2: transitionTest ($\mathcal{G}$, E_i, E_j)

---

input : the current temperature T; the temperature increase
rate T_rate; the Boltzmann constant K
output: *true* if the transition is accepted, *false* otherwise

1 **if** $E_\text{j} \leq E_\text{i}$ **then return** True
2 **if** $e^{-(E_\text{j}-E_\text{i})/(K \cdot T)} > 0.5$ **then**
3    $\lfloor$ $T \leftarrow T / 2^{(E_j-E_i)/\texttt{energyRange}(\mathcal{G})}$; **return** True
4 **else**
5    $\lfloor$ $T \leftarrow T \cdot 2^{T_\text{rate}}$; **return** False

---

number of nodes in the tree, a given number of iterations, or a given running time.

The Transition-based RRT (T-RRT) algorithm is a variant of RRT developed to explore a conformational space while taking the conformational energy into account [6], [10]. T-RRT extends RRT by integrating a stochastic transition test used to evaluate the local move from q_{near} to q_{new} based on their respective energies E_i and E_j, aiming to favor the exploration of low-energy regions of the space. This transition test is based on the Metropolis criterion, as explained for BH in the previous subsection.

The level of selectivity of this transition test is controlled by the *temperature* T: low temperatures limit the expansion to gentle slopes of the energy landscape, and high temperatures enable it to climb steep slopes. The basic MC method, as well as BH, consider constant temperature. In contrast, T is a self-adaptive parameter of the T-RRT algorithm. After each accepted uphill move, T is decreased to avoid over-exploring high-energy regions: it is divided by $2^{(E_j-E_i)/\texttt{energyRange}(\mathcal{G})}$, where $\texttt{energyRange}(\mathcal{G})$ is the energy difference between the highest-energy and the lowest-energy conformations in the graph $\mathcal{G}$ built so far. After each rejected uphill move, T is increased to facilitate the exploration and avoid being trapped in a local energy minimum: it is multiplied by 2^{T_rate}, where $T_\text{rate} \in (0,1]$ is the temperature increase rate. The pseudo-code of the T-RRT $\texttt{transitionTest}$ is presented in Algorithm 2.

The Multi-T-RRT algorithm is a multiple-tree variant of T-RRT [13]. Instead of building a single tree rooted at some initial conformation, the idea is to build n trees rooted at n given conformations q_init^k, $k = 1..n$. The pseudo-code of the Multi-T-RRT is presented in Algorithm 3. At each iteration, a tree $\mathcal{T}'$ is chosen for expansion in a round-robin fashion. Then, an extension is attempted toward a randomly sampled conformation q_rand, starting from its nearest neighbor q'_near, in $\mathcal{T}'$. If the extension succeeds, the new conformation q_new is added to $\mathcal{T}'$, and connected to q'_near. Then, we search for the conformation q''_near, which is the closest to q_new within all trees other than $\mathcal{T}'$. If the distance between q_new and q''_near is less than or equal to the extension step-size δ, $\mathcal{T}'$ is linked to and merged with $\mathcal{T}''$, the tree to which q''_near belongs. In that case, the number of trees is decreased by 1. The space exploration continues until all trees are merged into a single one or another stopping criterion (number of nodes, number of expansions, running time) is met.

A drawback of the Multi-T-RRT is that it returns a single

Algorithm 3: Multi-T-RRT

---

input : the conformational space $\mathcal{C}$; the extension step-size δ;
the energy function $E : \mathcal{C} \rightarrow \mathbb{R}$;
the initial conformations q_init^k, $k = 1..n$
output: the tree $\mathcal{T}$

1 **for** $k = 1..n$ **do**
2    $\lfloor$ $\mathcal{T}_\text{k} \leftarrow \texttt{initTree}(q_\text{init}^\text{k})$
3 **while not** $\texttt{stoppingCriteria}(\{\mathcal{T}_\text{k} \,|\, k = 1..n\})$ **do**
4    $\mathcal{T}' \leftarrow \texttt{chooseNextTreeToExpand}()$
5    $q_\text{rand} \leftarrow \texttt{sampleRandomConformation}(\mathcal{C})$
6    $q'_\text{near} \leftarrow \texttt{findNearestNeighbor}(\mathcal{T}', q_\text{rand})$
7    $q_\text{new} \leftarrow \texttt{extend}(q'_\text{near}, q_\text{rand}, \delta)$
8   **if** $q_\text{new} \neq null$ **and**
9    $\texttt{transitionTest}(\mathcal{T}', E(q'_\text{near}), E(q_\text{new}))$ **then**
10      $\texttt{addNewNode}(\mathcal{T}', q_\text{new})$
11      $\texttt{addNewEdge}(\mathcal{T}', q'_\text{near}, q_\text{new})$
12      $(\mathcal{T}'', q''_\text{near}) \leftarrow \texttt{findNearestTree}(q_\text{new})$
13   **if** $\texttt{distance}(q_\text{new}, q''_\text{near}) \leq \delta$ **then**
14        $\lfloor$ $\mathcal{T} \leftarrow \texttt{merge}(\mathcal{T}', q_\text{new}, \mathcal{T}'', q''_\text{near})$; $n \leftarrow n - 1$
15 **return** $\mathcal{T}$

Algorithm 4: T-RRT with Cycles

---

input : the conformational space $\mathcal{C}$; the extension step-size δ;
the energy function $E : \mathcal{C} \rightarrow \mathbb{R}$;
the tree built by the Multi-T-RRT $\mathcal{T}$
output: the graph $\mathcal{G}$

1 $\mathcal{G} \leftarrow \texttt{initGraph}(\mathcal{T})$
2 **while not** $\texttt{stoppingCriteria}(\mathcal{G})$ **do**
3    $q_\text{rand} \leftarrow \texttt{sampleRandomConformation}(\mathcal{C})$
4    $q_\text{near} \leftarrow \texttt{findNearestNeighbor}(\mathcal{G}, q_\text{rand})$
5    $q_\text{new} \leftarrow \texttt{extend}(q_\text{near}, q_\text{rand}, \delta)$
6   **if** $q_\text{new} \neq null$ **and**
7    $\texttt{transitionTest}(\mathcal{G}, E(q_\text{near}), E(q_\text{new}))$ **then**
8      $\texttt{addNewNode}(\mathcal{G}, q_\text{new})$
9      $\texttt{addNewEdge}(\mathcal{G}, q_\text{near}, q_\text{new})$
10   **for** $q_\text{can} \in \mathcal{G} \setminus \{q_\text{new}\}$ **do**
11   **if** $\texttt{distance}(q_\text{new}, q_\text{can}) \leq \delta$ **and**
12        $\texttt{noEdgeBetween}(q_\text{new}, q_\text{can})$ **then**
13          $\lfloor$ $\texttt{addNewEdge}(\mathcal{G}, q_\text{new}, q_\text{can})$
14 **return** $\mathcal{G}$

---

path connecting each pair of initial conformations, possibly not being the most likely transition path. To address this issue, we propose a new extension to T-RRT, based on a *cycle-addition* procedure. Starting from the tree produced by the Multi-T-RRT, the idea is to allow the space exploration to continue, and to add edges leading to the creation of cycles. This enables us to construct a graph from which several paths can be extracted between two given conformations. The pseudo-code of this *T-RRT with Cycles* is shown in Algorithm 4. It differs from T-RRT only in that, after every successful extension, an edge is added between q_new and each conformation in the graph $\mathcal{G}$ that is not already connected to q_new, and whose distance to q_new is less than or equal to the extension step-size δ. The stopping criteria can involve the number of iterations or the running time, as well as a convergence test.

As several paths may exist in the graph $\mathcal{G}$ between two given conformations, a quality criterion is required to compare paths.

This is achieved by associating weights with the edges of $\mathcal{G}$, based on the notion of mechanical work [10]. More precisely, the weight of the directed edge connecting q_i and q_j is equal to $\max\{\,0\,,\,E(q_j) - E(q_i)\,\}$, i.e. to the positive energy variation between q_i and q_j. This constitutes the amount of energy that has to be added to the molecule for the transition from q_i to q_j to happen. Note that using the mechanical work as a quality criterion requires to create two directed edges between q_i and q_j, instead of creating a single undirected edge. Finally, given two conformations in $\mathcal{G}$, the best (directed) path linking them is defined as the one minimizing the mechanical work, and is obtained using Dijkstra's algorithm.

In this work, we use the mechanical work as a path-quality criterion because it has been shown that T-RRT tends to generate paths having a low mechanical work [10]. Furthermore, in many situations, the mechanical work can assess the quality of a path better than a simple criterion such as the integral of the cost along the path [10]. Other criteria could be considered, such as the minimum resistance (linked to the MaxFlux algorithm [17], [18]) or the maximum flux [19]. However, finding which criterion is the most relevant is out of the scope of this paper.

III. RESULTS AND DISCUSSION

In a preliminary version of this paper [7], we presented results provided by the combination of BH and Multi-T-RRT to characterize the conformational energy landscape of the terminally-blocked alanine. The energy landscape of this minimalist peptide, which involves several local minima connected by multiple pathways, can be projected on a single and meaningful 2-dimensional map. Thus, it is a very good benchmark system to illustrate the performance of computational methods in physical chemistry. Results on the terminally-blocked alanine are not presented again here because of space limitations. For details on this illustrative test system, the interested reader is referred to [7].

Here, we present results on met-enkephalin, a pentapeptide with sequence: Tyr–Gly–Gly–Phe–Met. Despite its relatively small size, studying met-enkephalin is challenging from a structural point-of-view because of its great conformational variability. Met-enkephalin is still the focus of a large amount of work (e.g. [20], [21]). This endogenous opioid peptide is of significant interest due to its important roles in physiological processes, mostly related to pain and depression [22].

A. Met-Enkephalin Model and Settings

In this work, we consider an internal-coordinate representation of the molecule, assuming constant bond lengths and bond angles. Therefore, the conformational parameters are the $\{\omega, \phi, \psi\}$ dihedral angles of the backbone of each amino-acid residue plus the χ_i dihedral angles of the side-chains, which adds up to 23 degrees of freedom. We assume that all the angles but the ω angles can take values in the full angular range $[-180°, 180°)$. As the peptide bond torsions are known to undergo only small variations, the ω angles are allowed to vary only up to $10°$ from the planar trans conformation.

Measuring the distance between two conformations is required to select neighbor nodes in RRT-based algorithms and for clustering. Here we use a weighted l^2-norm in the space of the ϕ and ψ backbone dihedral angles (excluding the first and the last angle of the molecular chain), because they are the most relevant variables describing conformations of peptides. The weight of each dihedral angle within the norm is chosen as the Euclidian distance between its rotation axis and the farthest atom to it. In other words, greater weights are assigned to angles whose rotation results in larger conformational variation.

To illustrate the performance of the algorithms, we project their output on Ramachandran maps (i.e. ϕ-ψ maps) of the three middle residues (Gly2, Gly3, Phe4). Other representations are possible, such as projections on the first components provided by dimensionality reduction techniques (e.g. PCA, Isomap), or on pairs of structural descriptors (e.g. RMSD to a reference structure vs. end-to-end distance). However, we believe that a small set of ϕ-ψ maps is a clearer representation to identify conformational regions of a small peptide. Each 2-D map was generated using an exhaustive search procedure, by varying both dihedral angles with a $10°$ step-size and finding the lowest-energy conformation corresponding to each (ϕ, ψ) pair using an MC-based minimization procedure, such as the one involved in the BH algorithm. Note that such a computationally expensive procedure is only used here to visualize and validate our results.

To compute conformational energy values, we use an in-house implementation of the AMBER parm96 force-field [23] with an implicit representation of the solvent using the Generalized Born approximation. All programs have been run on a singe core of an Intel© Xeon© CPU E5-2650 at 2.00GHz.

B. Local Energy Minima of Met-Enkephalin

We have used the Basing Hopping (BH) algorithm, as described in Section II-A, to sample low-energy conformations of met-enkephalin. The parameters of BH were set as follows: 500 rounds were performed, with each round executing a maximum of 1000 MC-based minimizations. The number of consecutive rejections to estimate the convergence of each local minimization and of each BH round was set to 100. The large-amplitude perturbations in the inner loop of the BH algorithm affected only the ϕ and ψ dihedral angles, with a maximum step-size of 2 radians. The small-amplitude perturbations performed during an MC-based minimization affected all variables, with a maximum step-size of 0.2 radians. Only one angle was perturbed at each iteration. The temperatures T for the Metropolis-like transition tests at each iteration of BH and in the MC-based minimization method were set to 1000 K and 0.01 K, respectively.

The 232,440 *minima* and 500 *milestones* produced by the BH algorithm are projected in the three Ramachandran maps shown in Figure 1.[1] The algorithm required about 24 CPU hours. Note however that this computing time can be significantly reduced by using a more efficient implementation of the energy function, and by means of parallel computation

---

[1]Color figures are available in the online version of this manuscript.

(parallelization of the main loop in Algorithm 1 is straightforward). Computational efficiency issues are further discussed in Section IV. Note also that the use of a more sophisticated, gradient-based local minimization method would help speeding up the performance of BH.

The plots in Figure 1 show that the minima provide a good coverage of all the low-energy regions, the milestones being more concentrated in a few areas. The density-based clustering algorithm identified 8 clusters, some of which are not clearly separated. The clusters containing the lowest-energy minima correspond to folded conformations of met-enkephalin, whereas clusters of higher-energy minima contain stretched conformations of the peptide. Intermediate conformations are grouped into several medium-energy clusters. These results are consistent with those presented in other related work on met-enkephalin.

Here we take a closer look at results presented in some recent work providing a meaningful representation of the energy landscape of met-enkephalin through a Markov State Model (MSM) built from Molecular Dynamics simulations (MD) [21]. Even though this other approach applies a different energy model (GROMOS96 force field), a different conformational sampling method (MD) and a different clustering algorithm (hybrid kcenter-kmedoid), results show striking similarities to ours. Results in [21] suggest that four out of the eight backbone dihedral angles involved in our metric are sufficient to characterize the conformational space of met-enkephalin. These four angles are the ψ angles of Gly2, Gly3 and Phe4, and the ϕ angle of Gly3: $\{\psi_2, \psi_3, \psi_4, \phi_3\}$. The authors define a 4-sign code to characterize conformations with respect to the values (positive or negative) of these four angles. The two regions most frequently sampled by MD correspond to conformations of type $(----)$ and $(+--+)$. As can be seen in Figure 1, a large number of the BH milestones are found in these two regions. Interestingly, these two classes are correctly identified among the clusters of BH milestones, and they contain the lowest-energy conformations. The class $(+-++)$, which is also significantly sampled by MD, is well represented in our results. Only a few milestones appear in the region $(-++-)$, although results in [21] suggest a relatively more important population of this type of conformation. This is probably due to differences in the force-fields. The highest-energy cluster of BH milestones contains stretched conformations of both types $(++++)$ and $(-++-)$ merged together. We refer to this cluster as $(\star++\star)$.

To illustrate the application of the Multi-T-RRT, presented in the next subsection, we have selected the lowest energy conformation of each class $(+--+)$, $(----)$, $(+-++)$ and $(\star++\star)$. The first two correspond to folded conformations, the last one is a stretched/unfolded conformation and the third one is an intermediate conformation. Other intermediate states identified by clustering are not considered in the following analysis. Table I shows a structural representation of these four conformations (only the backbone is represented for clarity), together with the corresponding 4-sign code and the potential energy. A geometric symbol is associated to each conformation and used in the tables and Ramachandran maps to identify these conformations.

C. Transition Paths of Met-Enkephalin

We have applied the Multi-T-RRT algorithm to quickly discover many transition paths between the aforementioned meta-stable states of met-enkephalin, presented in Table **??**. We have used the following parameter settings: The extension step-size δ is set to 2.0, which means that, in the worst case, the maximum displacement of an atom between two conformations connected by an edge is of 2 Å. Note however that this distance is usually significantly smaller. At the beginning of the exploration, we impose the probability of accepting an energy increment of 0.1 kcal/mol to be approximately 50% by using the Boltzmann constant $(1.987 \cdot 10^{-3}$ kcal/mol/K) together with an initial temperature to 70 K. The temperature increase rate T_{rate} is set to 0.1.

Starting from four roots at the given states, Multi-T-RRT returns a single tree in approximately 2 minutes. Paths connecting the minima are extracted for analysis. To study the diversity of the transition paths of met-enkephalin, we run Multi-T-RRT 100 times. As an illustrative example, paths between the folded $(+--+)$ and unfolded $(\star++\star)$ states are visualized on the same Ramachandran maps in Figure 2. The average cost for the transition from $(+--+)$ to $(\star++\star)$ is 49.8 (MW) and in the reverse direction is 44.6 (MW), with a standard deviation across all the runs of 14.5. We observe that all of these paths avoid the high-energy regions in all projections. The paths are extremely diverse, although the standard deviation of their cost is not very high. This indicates that the energy landscape between these minima is relatively flat. One can see in the figure that a few paths go across saddle regions corresponding to values of $\phi_{2,3,4}$ close to zero. This shows that Multi-T-RRT is able to effectively explore vast portions of the conformational space.

D. Capturing Distinct Transitions of Met-Enkephalin

We have employed Multi-T-RRT with Cycles to compute transition probabilities between all pairs of energy minima for met-enkephalin. We have executed the method 100 times with a maximal running time of 10 minutes as stopping condition. Figure 3 shows the projection of one of the resulting graphs, which covers all the low-energy regions on the Ramachandran maps and constructs connections through transition regions. To estimate transition probabilities amongst the 100 executions, we count how many runs yield a graph from which a direct transition path can be extracted between a given pair of minima. More precisely, we consider that a run produces a direct transition path between two minima if the best path (with respect to the cost in terms of mechanical work) in the graph between them does not go through another minimum. Table I shows the probabilities of direct transitions between minima and the costs (average and standard deviation) associated to them. We can observe that direct transitions between all pairs of states are highly probable in most cases. Since the mechanical work is an additive function, we can also compute the cost of indirect paths going through a sequence of minima from the values in Table I. Table II presents the costs of different pathways from one of the two folded states $(+--+)$ and $(----)$ to the unfolded state $(\star++\star)$. In the
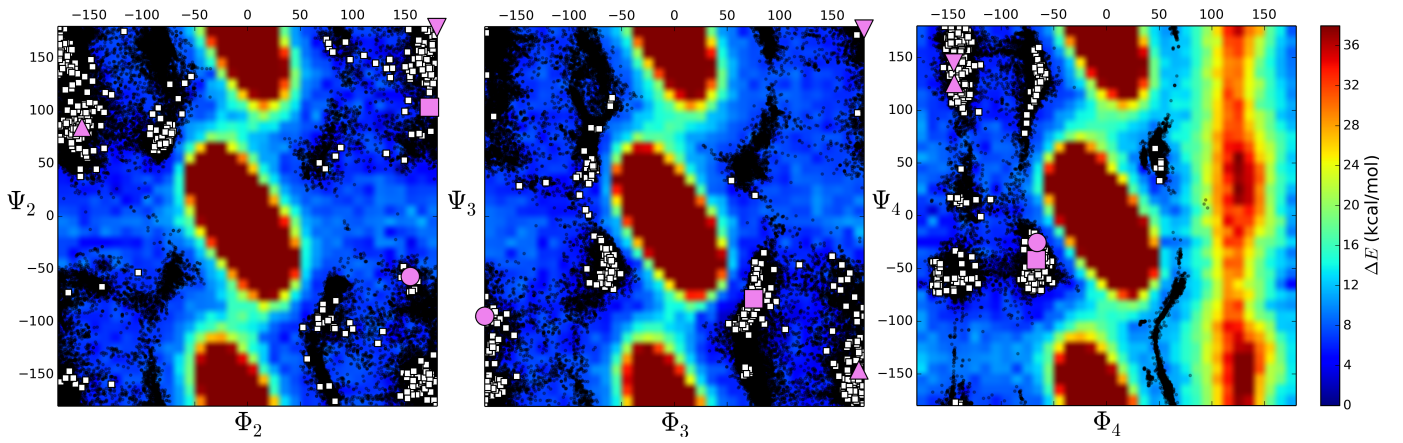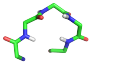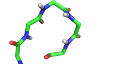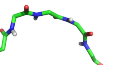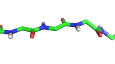
Fig. 1. Sampling the energy landscape of met-enkephalin using the Basin Hopping algorithm. The black markers represent the sampled *minima* and the white markers represent the *milestones*. The four large violet symbols indicate representative states identified by clustering (see Table I).

TABLE I
PROBABILITIES OF DIRECT TRANSITIONS AND THE MINIMAL-WORK PATHS BETWEEN THE FOUR IDENTIFIED MINIMA USING THE MULTI-T-RRT METHOD WITH CYCLES.

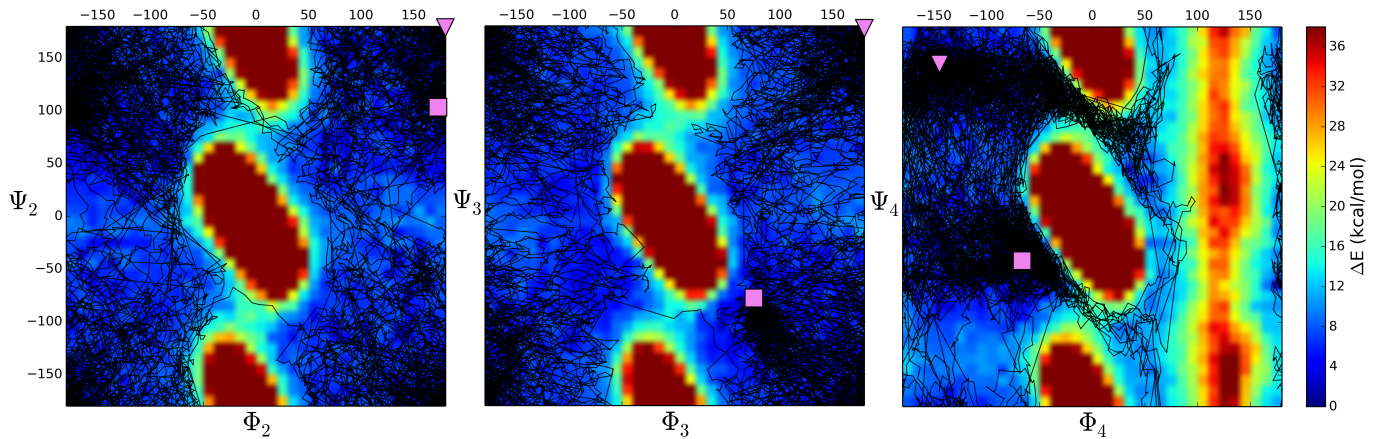| Symbol	Conformation	4-sign code	Energy (kcal/mol)	Transition To							
				■		●		▲		▼	
				Trans Prob	Path Cost (MW)	Trans Prob	Path Cost (MW)	Trans Prob	Path Cost (MW)	Trans Prob	Path Cost (MW)
■		$+--+$	-217.9	–	–	0.64	125.5 ± 34.7	0.63	105.5 ± 25.9	1.0	48.2 ± 8.9
●		$----$	-216.5	0.64	123.8 ± 34.7	–	–	1.0	55.3 ± 18.3	0.86	93.7 ± 27.0
▲		$+-++$	-215.9	0.63	103.5 ± 25.9	1.0	55.1 ± 18.3	–	–	0.89	72.6 ± 31.3
▼		$\star++\star$	-212.7	1.0	43.0 ± 8.9	0.86	90.2 ± 27.0	0.89	69.4 ± 31.3	–	–

Fig. 2. Projection of 100 path between a folded $(+--+)$ and unfolded $(\star++\star)$ states of met-enkephalin computed with the Multi-T-RRT algorithm.

aforementioned related work on met-enkephalin [21], different costs are also associated to alternative transition pathways from folded to unfolded states based on the Markov state model generated form MD. Although our cost rankings do not match perfectly to those in [21], most likely because of differences in the landscapes inferred by the different force-fields, the overall conclusion is the same: direct transitions from folded states to unfolded states are more probable than transitions going through intermediate states.
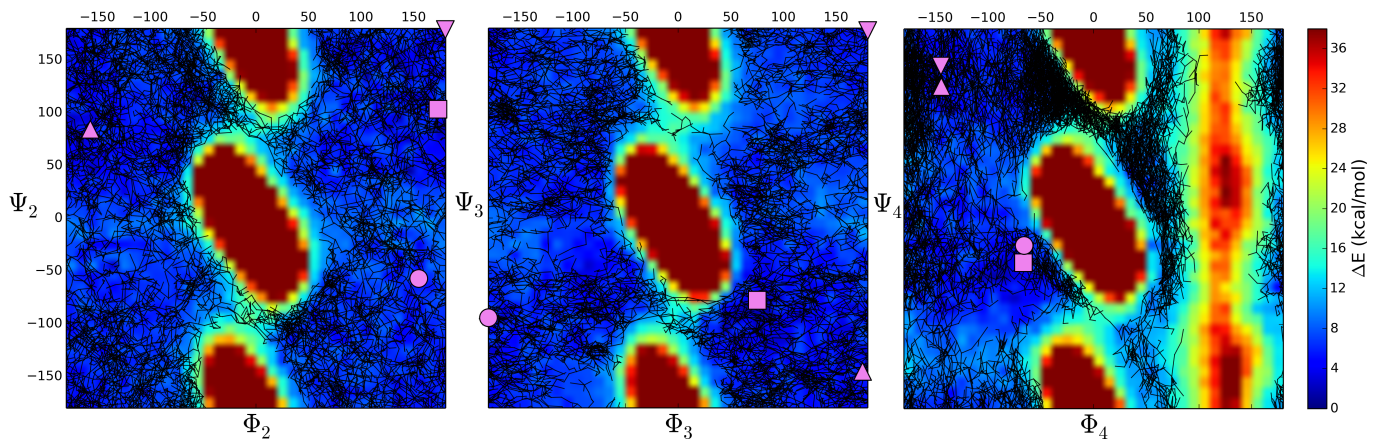
Fig. 3. Projection of a graph obtained by a single run of Multi-T-RRT with Cycles starting from four meta-stable states of met-enkephalin.

TABLE II
COSTS OF TRANSITION PATHWAYS FROM FOLDED TO UNFOLDED STATES

Pathway	Cost
■ → ▼	48.2
■ → ▲ → ▼	178.1
■ → ● → ▼	219.2
■ → ● → ▲ → ▼	253.4
● → ▼	93.7
● → ▲ → ▼	127.9
● → ■ → ▼	172.0
● → ■ → ▲ → ▼	301.9

IV. CONCLUSION

In this paper, we have presented a methodology to explore and characterize the energy landscape of flexible peptides. This methodology combines variants of two stochastic, sampling-based algorithms: Basin Hopping (BH) and Transition-based Rapidly-exploring Random Tree (T-RRT). A simplified version of the BH algorithm, where local, gradient-based energy minimization steps are replaced by simple Monte-Carlo-based minimization steps, achieves a relevant exploration of the energy landscape yielding numerous samples around energy minima. This leads to a quick determination of meta-stable structural states, which can be used as a starting point for the analysis of conformational transitions. The multiple-tree version of T-RRT is very fast at generating transition paths between a set of states. Running this algorithm several times produces a good description of the transition path ensembles. Finally, Multi-T-RRT with Cycles yields diverse transition paths in a single run of the algorithm.

Results on met-enkephalin presented in this paper, and results on the terminally-blocked alanine described in a preliminary version [7], show that the combination of BH and T-RRT quickly produces a meaningful representation of the energy landscape of small yet highly-flexible peptides. The identified meta-stable states and the transition pathways between them are comparable to those obtained with other, more expensive computational methods. Our results also illustrate the fact that stochastic algorithms can compete with MD-based approaches in providing accurate and insightful findings about flexible biomolecules. Nevertheless, this is an early work, and many

aspects still have to be improved and further investigated in terms of theory and implementation. Directions of future work include exploiting the graph produced by a single run of the Multi-T-RRT with Cycles to describe transition path and transition state ensembles. This will allow us to make better use of computational resources, as opposed to aggregating paths extracted from several runs of the Multi-T-RRT. In addition, Markov-based transition-step analysis can be conducted on the graph produced by one or more runs of T-RRT. This analysis allows one to estimate the stability of each computed state, and provides a rigorous basis for the designation of a state as stable or semi-stable.

We also aim to improve the implementation in order to efficiently deal with larger peptides and proteins. We have initiated some experiments with larger systems to identify the efficiency of our algorithms. One of our current test systems is chignolin, an artificial "mini-protein" composed of 10 residues. Chignolin presents an interesting energy landscape that has been investigated with other methods [24]. Our first tests with chignolin show that our methods are still applicable in practice to larger systems, although computing time increases significantly. When running the Multi-T-RRT using four conformations as input (one folded, one unfolded and two intermediate states identified by BH), computing a graph connecting them takes about 100 minutes on average. To reduce the computation time, it will be necessary to carefully implement the underlying methods in the algorithms, in particular the nearest neighbor search and energy computation. We are also working on the parallelization of Multi-T-RRT, building on our previous experience with the basic RRT algorithm [25].

References

[1] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge University Press, 2003.

[2] A. Shehu, "Probabilistic search and optimization for protein energy landscapes," in *Handbook of Computational Molecular Biology*, S. Aluru and A. Singh, Eds. Chapman & Hall/CRC Computer & Information Science Series, 2013.

[3] G. Paine and H. Scheraga, "Prediction of the native conformation of a polypeptide by a statistical-mechanical procedure. III. Probable and average conformations of enkephalin," *Biopolym.*, vol. 26, no. 7, pp. 1125–62, 1987.

[4] P. Amodeo, F. Naider, D. Picone, T. Tancredi, and P. Temussi, "Conformational sampling of bioactive conformers: A low-temperature NMR study of ^{15}N-Leu-Enkephalin," *J. Pept. Sci.*, vol. 4, no. 4, pp. 253–65, 1998.

[5] D. Wales and J. Doye, "Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms," *J. Phys. Chem. A*, vol. 101, no. 28, pp. 5111–16, 1997.

[6] L. Jaillet, F. Corcho, J.-J. Pérez, and J. Cortés, "Randomized tree construction algorithm to explore energy landscapes," *J. Comput. Chem.*, vol. 32, no. 16, pp. 3464–74, 2011.

[7] D. Devaurs, A. Shehu, T. Siméon, and J. Cortés, "Sampling-based methods for a full characterization of energy landscapes of small peptides," in *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on*, Nov 2014, pp. 37–44.

[8] B. Gipson, D. Hsu, L. Kavraki, and J.-C. Latombe, "Computational models of protein kinematics and dynamics: Beyond simulation," *Ann. Rev. Analyt. Chem.*, vol. 5, pp. 273–91, 2012.

[9] I. Al-Bluwi, T. Siméon, and J. Cortés, "Motion planning algorithms for molecular simulations: A survey," *Comput. Sci. Rev.*, vol. 6, no. 4, pp. 125–43, 2012.

[10] L. Jaillet, J. Cortés, and T. Siméon, "Sampling-based path planning on configuration-space costmaps," *IEEE Trans. Robotics*, vol. 26, no. 4, pp. 635–46, 2010.

[11] S. LaValle and J. Kuffner, "Rapidly-exploring random trees: progress and prospects," in *Algorithmic and Computational Robotics: New Directions*, 2001, pp. 293–308.

[12] D. Frenkel and B. Smit, *Understanding Molecular Simulations: From Algorithms to Applications*, 2nd ed. Academic Press, 2001.

[13] D. Devaurs, M. Vaisset, T. Siméon, and J. Cortés, "A multi-tree approach to compute transition paths on energy landscapes," in *Proc. AIRMCB Workshop*, 2013.

[14] B. Olson and A. Shehu, "Evolutionary-inspired probabilistic search for enhancing sampling of local minima in the protein energy surface," *Proteome Sci.*, vol. 10, no. Suppl 1, p. S5, 2012.

[15] Y. Kim, S. Choi, and W. Kim, "Efficient basin-hopping sampling of reaction intermediates through molecular fragmentation and graph theory," *J. Chem. Theory Comput.*, vol. 10, no. 6, pp. 2419–26, 2014.

[16] M. Kim, S.-H. Choi, J. Kim, K. Choi, J.-M. Shin, S.-K. Kang, Y.-J. Choi, and D. Jung, "Density-based clustering of small peptide conformations sampled from a molecular dynamics simulation," *J. Chem. Inform. Model.*, vol. 49, no. 11, pp. 2528–36, 2009.

[17] S. Huo and J. Straub, "The MaxFlux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature," *J. Chem. Phys.*, vol. 107, no. 13, pp. 5000–6, 1997.

[18] ——, "Direct computation of long time processes in peptides and proteins: reaction path study of the coil-to-helix transition in polyalanine," *Proteins: Structure, Function, Genetics*, vol. 36, no. 2, pp. 249–61, 1999.

[19] R. Zhao, J. Shen, and R. Skeel, "Maximum flux transition paths of conformational change," *J. Chem. Theory Comput.*, vol. 6, no. 8, pp. 2411–23, 2010.

[20] F. Sicard and P. Senet, "Reconstructing the free-energy landscape of met-enkephalin using dihedral principal component analysis and well-tempered metadynamics," *The Journal of Chemical Physics*, vol. 138, no. 23, 2013.

[21] R. Banerjee and R. I. Cukier, "Transition paths of met-enkephalin from markov state modeling of a molecular dynamics trajectory," *The Journal of Physical Chemistry B*, vol. 118, no. 11, pp. 2883–2895, 2014.

[22] D. Sauriyal, A. Jaggi, and N. Singh, "Extending pharmacological spectrum of opioids beyond analgesia: Multifunctional aspects in different pathophysiological states," *Neuropeptides*, vol. 45, pp. 175–88, 2011.

[23] P. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot, and A. Pohorille, "The development/application of a "minimalist" organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data," in *Computer Simulation of Biomolecular Systems*, ser. Computer Simulations of Biomolecular Systems, W. van Gunsteren, P. Weiner, and A. Wilkinson, Eds. Springer Netherlands, 1997, vol. 3, pp. 83–96.

[24] D. Satoh, K. Shimizu, S. Nakamura, and T. Terada, "Folding free-energy landscape of a 10-residue mini-protein, chignolin," *FEBS Letters*, vol. 580, no. 14, pp. 3422 – 3426, 2006.

[25] D. Devaurs, T. Siméon, and J. Cortés, "Parallelizing RRT on large-scale distributed-memory architectures," *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 571–579, 2013.