A randomized tree construction algorithm to explore energy landscapes

Léonard Jaillet, Francesc J Corcho, Juan J Pérez, Juan Cortés

A Randomized Tree Construction Algorithm to Explore Energy Landscapes

Léonard Jaillet[a)]
Institut de Robòtica i Informàtica Industrial, Consejo Superior de Investigaciones Científicas, Universitat Politecnica de Catalunya, Barcelona 08028, Spain
Francesc J. Corcho and Juan J. Pérez
Dep. d'Enginyeria Química, UPC. ETS d'Enginyeria Industrial, Av. Diagonal, 647, 08028 Barcelona, Spain
Juan Cortés[b)]
CNRS; LAAS; 7 avenue du colonel Roche, F-31077 Toulouse, France and Université de Toulouse; UPS, INSA, INP, ISAE; LAAS; F-31077 Toulouse, France

This paper presents a new method for exploring conformational energy landscapes. The method, called T-RRT, combines ideas from robotics path planning and statistical physics. A search tree is constructed on the conformational space starting from a given state. The tree expansion is driven by a double strategy: On the one hand, it is naturally biased toward yet unexplored regions of the space. On the other hand, a Monte Carlo-like transition test guides the expansion toward energetically favorable regions. The balance between these two strategies is automatically achieved thanks to a self-tuning mechanism. The method is able to efficiently find both energy minima and transition paths between them. As a proof of concept, the method is applied to several academic benchmarks and to the alanine dipeptide.

I. INTRODUCTION

Estimation of the physicochemical properties of molecules from their atomic structure requires the characterization of the stationary points on the underlying energy landscape. However, for complex systems, the states of interest may subsume an exponentially large number of local minima that require appropriate sampling. This is a hard problem that has attracted the interest of scientists from decades and that has been addressed using different approaches. These can be grouped into three categories: i) canonical sampling methods that produce a Boltzmann weighted set of configurations; ii) non-canonical sampling methods that aim to enhance sampling; iii) methods that characterize non-related stationary points on the potential energy surface[1,2].

Methods in the first category include the Metropolis Monte Carlo (MC) algorithm and Molecular Dynamics (MD)[3,4]. Even though both methods provide Boltzmann weighted distributions, the former allows only computing thermodynamic properties of the system, whereas the latter allows computing both thermodynamic and dynamic properties of the system though the analysis of the metastable states and the transitions that may occur between them. Unfortunately, both methods are inefficient for sampling conformations that are beyond high potential energy barriers.

Generalized-ensemble methods such as Replica Exchange and Simulated Tempering[5,6], which explore the conformational space at different temperatures, have become popular techniques to enhance coverage and convergence properties of MD and MC methods. In addition, different non-canonical methods have been developed to enhance the sampling of infrequent events or of transitions between conformational states[7–9], to focus the simulation on regions of interest[10], or to bias the evolution of the system[11,12] within MD simulations. Variants of MC methods involving sophisticated moves (e.g. the Activation-Relaxation Technique[13]) have also been developed for a more effective exploration of the energy landscape.

The third category of methods includes algorithms that combine conformational sampling and energy minimization. The stochastic method can be considered the simplest procedure, where the starting points for minimization are generated through random sampling. Algorithms that bias the search to characterize the global energy minimum also allow a sampling of the most relevant energy minima. The two most widely used methods are the genetic algorithms and the simulated annealing procedure using either MC or MD as drivers[14]. A different approach to find stable conformations of polypeptides is the building-up procedure by assembling the conformations of short fragments[15,16].

In addition to the aforementioned conformational sampling methods, various approaches have been proposed to extract ensemble properties from multiple trajectories (obtained by MD or MC methods), with the aim of providing a clearer representation of the conformational landscape, i.e. a set of stable conformations and the most probable transitions between them. The general idea is to analyze the set of trajectories by statistical mechanics tools (e.g. Markovian models) and transition rate theory in order to construct a transition network over the conformational space[17–21]. The main drawback

[a)] Electronic mail: ljaillet@iri.upc.edu
[b)] Electronic mail: juan.cortes@laas.fr

of such methods is that they require information provided by numerous and computationally expensive simulations. As an alternative, several methods[22,23] inspired by robotic path planning algorithms propose to build networks (called roadmaps) that capture the most relevant regions of the the energy landscape using random conformational sampling combined with local interpolation methods to connect neighbor samples. Such methods represent important seminal work on the application of robotic path planning to problems in computational biology, and have interesting applications such as the characterization of (un)folding pathways. However, these methods that rely on a simple sampling scheme over the entire conformational space are not adapted to accurately explore the high-dimensional energy landscape of biological molecules.

Besides global exploration methods, various local methods have been proposed to find transition paths between two given stable states. In general, the idea is to start from a trivial path and to deform it iteratively in order to improve its energy profile. Examples of these methods are Nudged Elastic Band[24–26], Zero Temperature String[27,28], and Transition Path Sampling[29,30]. A recent approach called Forward Flux Sampling[31] does not require an initial path, but needs to define series of interfaces between the initial and the goal state. Methods for computing conformational transitions have also been developed based on biased or targeted MD[32,33], and Normal Mode Analysis[34,35].

This paper presents a conformational exploration method called Transition-RRT (T-RRT)[36] which is inspired by robotic path planning algorithms and by methods in statistical physics. T-RRT can be seen as a non-canonical sampling method and/or as a method to compute energetically favorable conformational transition paths. Similarly to MC methods, T-RRT applies small moves and a transition test based on the Metropolis criterion. However, instead of generating a single path on the conformational space, it constructs a tree with better coverage properties. Such a data structure enables the simultaneous exploration of different regions of the space. Moreover, it avoids the ineffective behavior of MC algorithms, which tend to waste time getting back to regions of the space already explored. Finally, T-RRT is a reactive search method[37] that uses a self-tuning mechanism to improve its overall efficiency. Starting from a given conformation, the tree branches grow first on the more favorable regions (the valleys of the landscape), and tend to cover the whole search-space while the number of iterations increases. Such exploration enables to find the local minima and the saddle-points of the landscape. Besides, paths extracted from the tree can be directly exploited as a good approximation of transition paths between stable conformations.

II. METHOD

This section describes the Transition RRT algorithm (T-RRT), whose pseudo-code is sketched in Figure 1. T-RRT extends the Rapidly-exploring Random Tree (RRT) algorithm[38] by incorporating a stochastic state-transition test, similarly to MC methods. RRT is a randomized space-filling method that was initially developed for path planning in robotics. Its most interesting feature is the implicit bias of the tree expansion toward yet unexplored regions of the space (Subsection II A).

The proposed variant, T-RRT, also holds this interesting property. In addition, the integrated transition test rejects some of the generated states if they do not correspond to energetically acceptable moves (Subsection II B). Thus, the expansion is biased toward both unexplored and low energy regions. The appropriate balance between these two types of bias relies on a reactive scheme presented in Subsection II C. Also, a filtering procedure rejects new states if they are too close to nodes of the tree, which improves the space-covering property of the method (Subsection II D). Overall, T-RRT is an effective and general exploration method that can be used to find stable states, or to compute probable transition paths between given pairs of states (Subsection II E).

A. RRT Principle: Bias Toward Unexplored Regions

The core of the T-RRT algorithm (Figure 1) is inherited from the basic RRT[38]. RRT is an efficient path planning method able to tackle complex problems in high-dimensional spaces. It has been successively used in many disciplines such as robotics, computer animation, and structural biology[39,40]. The idea is to iteratively construct a tree data structure made of nodes and edges that correspond to states and small-amplitude motions between neighbor states, respectively. At each iteration, a state (i.e. a conformation in the context of structural biology) is randomly sampled (`SampleConf` function). The nearest state already contained in the search tree is then searched (`NearestNeighbor` function). Finally, a new node is created by extending the nearest neighbor toward the random sample (`Extend` function). Employing the simplest expansion strategy (called RRT-*Extend* in related literature[38]), the extension step-size δ remains constant for all the iterations. Following such a construction procedure (illustrated in Figure 2) the tree is implicitly biased toward yet unexplored regions. This behavior comes from the probability for a node to be extended, which is proportional to the volume of its Voronoi cell (i.e. the set of points closer to this node than to any other node). Note that this property does not require the explicit construction of the Voronoi cells, which would be computationally expensive.

Classically, RRT has been used to search paths in a continuous state-space composed of feasible-state and infeasible-state subsets. In this context, RRT is much

Algorithm 1: Transition-RRT

```
input  : the Conformational Space CS;
         the energy function E : CS → ℝ;
         the initial conformation q_init ;
output : the tree T;
begin
    T ← InitTree(q_init);
    while not StopCondition(T) do
        q_rand ← SampleConf(CS) ;
        q_near ← NearestNeighbor(q_rand, T);
        if ExploGuarantee(T, q_near, q_rand) then
            q_new ← Extend(T, q_rand, q_near);
            if TransitionTest(E(q_near), E(q_new)) then
                AddNewNode(T, q_new);
                AddNewEdge(T, q_near, q_new);
end
```

FIG. 1. Transition-RRT

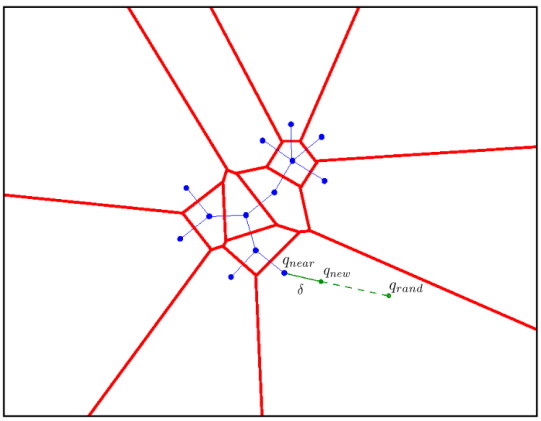

FIG. 2. RRT construction scheme. In blue/thin lines, the RRT tree. In red/bold lines, the Voronoi cells associated with the states contained in the tree. At each step, a state q_{rand} is randomly sampled, and its nearest neighbor in the search tree q_{near} is selected. It corresponds to the node in the Voronoi cell where q_{rand} has been sampled. A new node q_{new} is created by moving from q_{near} a distance δ in the direction of q_{rand}. The Voronoi bias favors the tree expansion toward unexplored regions of the space.

more efficient than a simple random walk method, since it avoids wandering around in already explored regions[41]. By inference, T-RRT is expected to be more effective than standard MC methods to explore molecular energy landscapes.

B. Transition Test: Hindering Steep Climbing

T-RRT extends RRT by integrating a transition test to hinder the tree expansion toward energetically unfavorable regions of the space (`Transition Test` function). Similarly to MC methods, the acceptance rule of a local move is defined by comparing the energy E_j of the new state with the energy E_i of the previous state (i.e. the parent node in the tree). This test is based on the Metropolis criterion, with a transition probability p_{ij} defined as follows:

$$p_{ij} = \begin{cases} exp(-\frac{\Delta E_{ij}}{kT}), & \text{if } \Delta E_{ij} > 0 \\ 1, & \text{otherwise} \end{cases} \quad (1)$$

where $\Delta E_{ij} = E_j - E_i$ is the energy variation between the two states, k is the Boltzmann constant, and T is the temperature. Note however that T-RRT is a non-canonical sampling method, which is not expected to produce a Boltzmann weighted set of conformations, but to efficiently find energy minima and probable conformational transition paths. Therefore, T does not necessarily carry any physical meaning. Indeed, T is simply a parameter of the algorithm.

Within search methods involving the Metropolis criterion, the temperature is usually kept constant (e.g. MC simulation) or is subject to predefined variations (e.g. heating and cooling phases in simulated annealing). In the case of T-RRT, the `Transition Test` function incorporates a reactive scheme to dynamically tune this parameter. It allows to control the level of difficulty of the transition test, according to the information acquired during the exploration.

C. Automatic Temperature Tuning

During the construction of the search tree, the number of attempts necessary to add a new node is a good indicator to measure the evolution of the exploration process. A large number of consecutive failures means that the exploration is stuck because the tree cannot be further expanded toward favorable regions. Within T-RRT, this information is used to regulate the temperature that determines the difficulty of the transition test.

At the initialization, T is set to a low value in order to only permit the tree expansion on very easy positive slopes (and negative ones). Then, during the exploration, the number of consecutive times the Metropolis criterion discards a state is recorded and used for temperature tuning. When the T-RRT search reaches a maximum number of consecutive rejections $Fail_{max}$, the temperature increases by a factor λ. Contrarily, each time an uphill transition test succeeds, the temperature decreases by the same factor λ. Thus, the temperature automatically adapts itself in such a way that an extension corresponding to a positive energy variation is performed every $Fail_{max}$ times in average. This temperature regulation strategy is a way to balance the search between unexplored regions and low energy regions.

D. Exploration Guarantee

The adaptive temperature tuning introduced earlier may however lead to bottleneck situations. The temperature T may be stabilized by the insertion of new states very similar to the ones already contained in the tree, whereas the expansion toward new regions of the space would require an increment of T. The insertion of such states only contributes to the refinement of the exploration in regions already reached by the tree. This situation is illustrated on a 2D fictive energy landscape in Figure 3-a.

To overcome this drawback, the selected state q_{near} is not extended if the the distance to the random state q_{rand} is smaller than the extension step-size δ (`ExploGuarantee` function). Such a simple filtering avoids an excessive refinement of low-energy regions, therefore facilitating the tree expansion toward new regions of the space. Furthermore, the size of the tree (in number of nodes) is limited, which reduces the computational cost of operations such as neighbor search. The improvement provided by this filtering process is illustrated in Figure 3-b.

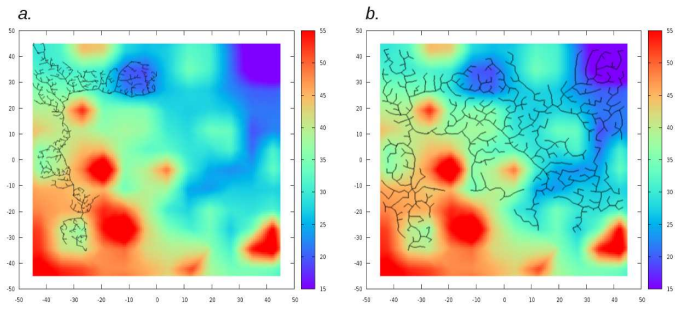

FIG. 3. Impact of the Exploration Guarantee on the T-RRT algorithm. The trees in both pictures have the same size (800 nodes), and are rooted at the same coordinate (-30, -30). Without this filtering process (a), the insertion of nodes very close to existing ones tends to slow down the exploration by decreasing the temperature. With filtering (b), the algorithm continues to explore new regions of the space.

E. Applicability of T-RRT

1. Main Minima Search Method

The proposed method can be used to find the main minima of a conformational landscape. Hereafter, this type of operating mode is called T-RRT$_{min}$. Starting from a given conformation, the method explores the landscape until a stop condition is reached. This condition can be defined by an amount of computing time, a maximum number of created conformations, or from an estimation of the space coverage. Once the search is stopped, a minima extraction method can be applied to the conformations contained in the tree. In the current implementation, we apply a method based on the the root mean square deviation (RMSD) between conformations: the main minima are the conformations whose energy is lower than the energy of all its neighbors for a given RMSD threshold.

2. Transition Path Search Method

T-RRT can also be applied to find low-energy paths between a given pair of stable conformations, and the transitions states associated with. This operating mode is called T-RRT$_{trans}$. For such a search, the tree is rooted at one of the stable conformations, and the algorithm is iterated until one of the tree leaves reaches the target conformation (i.e. the distance between both conformations is less than the extension step-size δ). The transition path is then extracted from the tree structure, by following the branches from the leaf to the root. The quality of the computed path relies on two points. First, the Voronoi bias avoids backtrack motions, contrarily to basic MC techniques that propagate a single state. Second, the temperature is regulated all over the tree construction process, so that heating phases only occur when necessary for passing through higher energy barriers in order to reach other conformational regions. Consequently, paths computed by T-RRT tend to minimize both the total amount of positive energy variation and the length (empirical proofs have been provided by Jaillet *et al.*[42]). Therefore, such paths are good candidates to represent transitions between pairs of stable conformations.

III. RESULTS

As a proof of concept, this section first presents results on two academic benchmarks, for which the energy landscape is represented by a two-parameter analytic function. Then, the method is applied to study the energy landscape of the alanine dipeptide using an implicit description of the solvent.

For each problem, T-RRT is first used to find the main energy minima (T-RRT$_{min}$ search), and then to find the transition paths between these states (T-RRT$_{trans}$ search). The algorithm parameters are the following. The Boltzmann constant k being $3.297 \cdot 10^{-27}$ kcal/K, the initial temperature is set to $T = 70$ K. These values make that, at the initialization of the algorithm, the probability of accepting an energy increment of 0.1 kcal/mol is around 50%. The maximum number of consecutive expansion failures before a temperature increase is set to $Fail_{max} = 10$ and $Fail_{max} = 100$ for T-RRT$_{min}$ and T-RRT$_{trans}$, respectively. With these settings, T-RRT$_{min}$ covers the space more rapidly than T-RRT$_{trans}$, while T-RRT$_{trans}$ finds the saddle-points more accurately. The temperature variation factor is $\lambda = 0.1$ in all the cases.

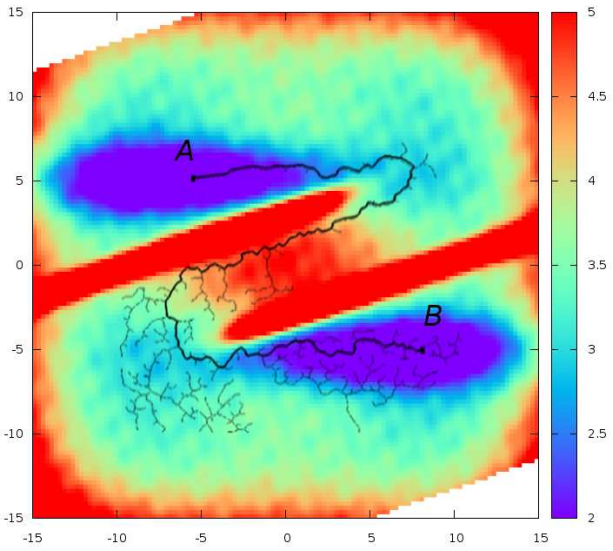

FIG. 4. *Zorro* potential. In bold black, a T-RRT transition path found between minima A and B. It circumvents two energy barriers and passes through a higher energy saddle-point. In thin black, other branches of the associated search tree.

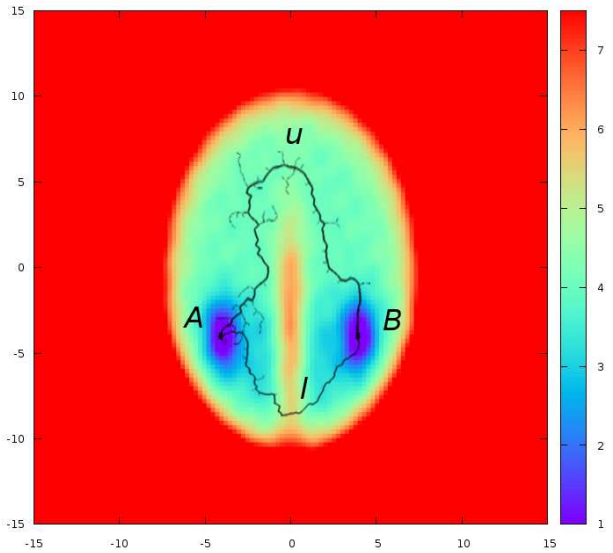

FIG. 5. *Alien* potential. Minima A and B can be connected through two pathways, u and l. In bold black, two paths found with T-RRT. In thin black, branches of the associated search trees.

A. 2D Academic Benchmarks

The apparently simple landscapes described below[43] represent tricky test systems for benchmarking methods that search for conformational transition pathways.

The *Zorro* potential, represented in Figure 4, involves two low-energy regions with respective minima A and B. The pathway connecting these minima needs to circumvent two energy barriers and passes through a saddle-point located in the middle part of the landscape. The analytic expression of this energy landscape is:

$$\begin{aligned} E(x,y) = {} & 0.2((x/8)^4 + (y/8)^4) \\ & - 3e^{-0.2(0.05(x+5)^2+(y+5)^2)} - 3e^{-0.2(0.05(x-5)^2+(y-5)^2)} \\ & + 5e^{-0.2(x+3(y-3))^2}/(1+e^{-x-3}) \\ & + 5e^{-0.2(x+3(y+3))^2}/(1+e^{x-3}) + 3e^{-0.01(x^2+y^2)} \\ & + 0.06 * (\sin(5x+\sqrt{2}y) + \cos(\sqrt{5}x+\sqrt{3}y) \\ & + \sin(3*y-\sqrt{2}x) + \cos(3*x-\sqrt{5}y)). \end{aligned}$$

The variation of parameters x and y is limited to the interval $[-15, 15]$. Within these bounds, the energy varies from 0.5 up to 9.7 (in arbitrary units).

The *Alien* potential is represented in Figure 5. Like the previous benchmark, it involves two low-energy basins with respective minima A and B. These regions are connected through two main pathways, which we refer to as lower and upper (l and u in Figure 5). The energy value of the transition states for these two pathways is very similar. However, the upper pathway is much larger than the lower one. The analytic expression of this energy landscape is given by:

$$\begin{aligned} E(x,y) = {} & 3 + \frac{3}{e^{5(\frac{x^2}{4}+\frac{(2+\frac{y}{2})^2}{10})}} - \frac{3}{e^{5((-2+\frac{x}{2})^2+(2+\frac{y}{2})^2)}} \\ & - \frac{3}{e^{5((2+\frac{x}{2})^2+(2+\frac{y}{2})^2)}} + \frac{(\frac{x^2}{4}+\frac{y^2}{8})^4}{10000} + \frac{1+erf(1+\frac{y}{2})}{2} \\ & + \frac{3}{50}\left(\cos(\sqrt{5}x+\sqrt{3}y) + \cos(3x-\sqrt{5}y))\right) \\ & + \frac{3}{50}\left(\sin(5x+\sqrt{2}y) - \sin(\sqrt{2}x-3y)\right). \end{aligned}$$

Like in the previous example, the variation of parameters x and y is limited to the interval $[-15, 15]$. Within these bounds, the energy varies from 0.1 up to more than 5000 (in arbitrary units).

The low-dimensionality of these benchmarks enables comparison of results with those obtained by exhaustive search. A 128×128 grid discretizing the search-space was used to perform such search. In order to analyze the variability of T-RRT results (due to the randomized exploration), the algorithm was run several times on each problem. Results presented below show the average and the standard deviation over 100 runs.

1. T-RRT$_{min}$ Search

Table I shows results of the T-RRT$_{min}$ search for the *Zorro* and the *Alien* benchmarks. In both cases, the minima found are very similar (in position as well as in energy) to those extracted from an exhaustive grid search method. Moreover, the low values of the standard devi-

Zorro

	T-RRT$_{min}$		Grid	
	A	B	A	B
x	7.6 ± 0.2	-5.6 ± 0.1	7.7	-5.6
y	-5.1 ± 0.2	5.1 ± 0.1	-5.1	5.1
E	0.59 ± 0.03	0.60 ± 0.02	0.55	0.57

Alien

	T-RRT$_{min}$		Grid	
	A	B	A	B
x	-4.0 ± 0.2	4.1 ± 0.2	-3.9	4.1
y	-4.0 ± 0.2	-4.1 ± 0.2	-4.1	-3.9
E	0.30 ± 0.18	0.51 ± 0.18	0.14	0.34

TABLE I. Energy minima for the academic benchmarks.

Zorro

	T-RRT	Grid
	A → B	A → B
x	-0.9 ± 0.6	0.5
y	0.7 ± 0.6	-0.6
E	4.96 ± 0.02	4.95

Alien

	T-RRT		Grid	
	A $\xrightarrow{l}$ B	A $\xrightarrow{u}$ B	A $\xrightarrow{l}$ B	A $\xrightarrow{u}$ B
x	0.2 ± 0.1	0.8 ± 2.3	0.1	-0.9
y	-8.6 ± 0.2	3.3 ± 2.4	-8.2	0.5
E	3.98 ± 0.03	4.13 ± 0.08	3.96	4.07

TABLE II. Transition states for the academic benchmarks.

ation confirms the robustness of T-RRT despite the random nature of the search process. For these experiments, the T-RRT iteration was stopped after the insertion of 1000 nodes. The extension step-size (i.e. the euclidean distance between two connected states in the tree) was set to $\delta = 0.5$. The time performance for such a search ranged from 20 to 30 seconds[44]. Then, the extraction of the main minima from the T-RRT tree, required less than 0.1 seconds.

2. T-RRT$_{trans}$ Search

The localization of the minima being worked out, T-RRT was used to find transition paths between them. Table II shows the characteristics of the associated transition states found for the two benchmarks. In both cases, the transition states computed with T-RRT are very close to the ones found by exhaustive search. Moreover, in the case of the *Alien*, both the lower and the upper pathways were found with T-RRT. Over the 100 computed paths, 66 passed through the upper region part whereas only 34 passed through the lower one. In other words for energy barriers of similar height, the T-RRT exhibits a higher probability to go through wider pass transition regions than through narrow pass pathways. This result illustrates that the preferred pathway chosen by the T-RRT search is affected by the width of the pass pathway. Accordingly, the T-RRT solutions do not only depend on the potential energy of the explored states, but also on the number of possible equivalent paths to reach the other minium, and consequently the procedure searches for the lowest (free) energy route.

B. Alanine Dipeptide in Implicit Solvent

The alanine dipeptide is a very frequent test-model molecule for theoretical studies[45–50]. In this work, we have studied the alanine residue acetylated in its N-terminus and methylamidated in its C-terminus (see Figure 6). Despite its small size, alanine dipeptide shares some structural features with larger peptides and proteins. In particular, due to the flexibility of the ϕ and ψ angles, the molecule is able of forming internal hydrogen bonds.

For facilitating the analysis of results obtained with T-RRT, an energy map on the $\{\phi, \psi\}$ coordinates of the peptide was generated using a systematic procedure. The two dihedral angles were varied with constant 10° step-size. For each $\{\phi, \psi\}$ value, the conformation was energy-minimized using a steepest descent method. In order to fix the $\{\phi, \psi\}$ angles during the minimization, we used an additional $\{\phi, \psi\}$-harmonic potential whose minimum was equal to the desired values of the two angles. The optimization was stopped when the RMSD for consecutive iterations reached $1 \cdot 10^{-3}$Å. The computed energy map appears in background in Figure 7. For the construction of the energy map, as well as for the energy evaluation within T-RRT, we used the parm96 AMBER force field together with an implicit representation of the solvent with the Generalized Born approximation. The values of the internal and external dielectric constants were set to 1.0 and 78.5 respectively.

The conformational exploration with T-RRT was performed on an internal coordinate representation of alanine dipeptide with constant bond lengths and bond angles. Thus, the conformational parameters are the seven bond torsions associated to the dihedral angles ϕ, ψ, $\omega_{1,2}$, and $\chi_{1,2,3}$ represented in Figure 6. Note that, since the peptide bond torsions $\omega_{1,2}$ are known to only vary slightly, they were bound to the interval $[170°, 190°]$.

1. T-RRT$_{min}$ Search

The energy landscape exploration yielded six minima that correspond to the P_{II}, α_R, α_L, C_7^{ax}, α_P and C_5 stable states of the alanine dipeptide. Their position and energy are presented in Table III (for reference, the energy of the minimum-energy conformation is set to zero). Figure 7.a shows these minima superimposed on

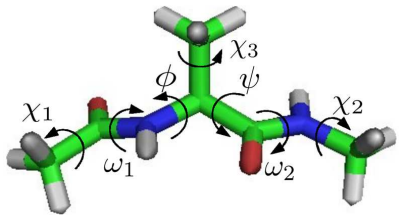

FIG. 6. Alanine dipeptide and the seven conformational parameters used for the exploration.

	P_{II}	α_R	α_L	C_7^{ax}	α_P	C_5
ϕ	-66.7	-62.7	46.8	49.6	-148.4	-146.5
ψ	144.3	-44.0	51.3	-137.7	-70.3	161.6
E	0.3	1.1	4.4	4.2	1.7	0.0

TABLE III. Energy minima found by T-RRT for the alanine dipeptide within a Generalized Born solvent.

the computed energy map of the $\{\phi, \psi\}$ internal coordinates. It appears that T-RRT solutions fit very well the minimum energy region of the map. This result shows the capacity of the method to find multiple minima in high-dimensional landscapes. For obtaining these results, T-RRT$_{min}$ was iterated until the insertion of 8000 nodes in the tree. The exploration step-size δ was set such that the maximal angular variation was of 5°. The search took 15 minutes. Finally the six minima were identified by applying the minima extracting method based on RSMD. For improving quality of results, these minima were locally optimized by a steepest descent method with the same stop criterion than for the construction of the $\{\phi, \psi\}$ energy map. Overall, this rapid optimization process took about 1 second.

2. T-RRT$_{trans}$ Search

T-RRT was used to find transitions paths between several pairs of minima. Figure 7.b-d, shows paths for the transitions $\alpha_L \rightarrow C_7^{ax}$, $\alpha_R \rightarrow P_{II}$ and $C_7^{ax} \rightarrow C_5$, respectively. The algorithm was run 50 times for each pair. The solutions are all represented in the figure. The computing time for each transition path ranged from 1 to 5 minutes. The projection of the computed paths on the $\{\phi, \psi\}$ energy map underlines the ability of T-RRT to follow the valleys of the energy landscape (remind that the figure shows a two-dimensional projection of the results, while the exploration takes place in a seven-dimensional space). Table IV shows the distribution of the solutions according to the different pathways connecting the pairs of minima. For the $\alpha_L \rightarrow C_7^{ax}$ transition, two main pathways appear (I and II). The higher probability for class-I paths can be explained by an energy barrier slightly lower than for class-II paths. The variability of the solutions within each pathway is low, as the corridors that connect the two minima are quite narrow. Looking at the

	$\alpha_L \rightarrow C_7^{ax}$		$\alpha_R \rightarrow CP_{II}$		$C_7^{ax} \rightarrow C_5$				
Pathway	I	II	I	II	I	II	III	IV	V
Distrib.	31	19	33	17	22	10	9	7	2

TABLE IV. Distribution of T-RRT solutions into path classes for three conformational transitions of alanine dipeptide.

$\alpha_R \rightarrow P_{II}$ transition, T-RRT also detect two possible pathways (I and II). Class-I paths are also more probable in this case. The variability of class-II paths is due to the wideness and the flatness of the the saddle region that connects the two minima. Finally, five pathways (I to V) were found for the $C_7^{ax} \rightarrow C_5$ transition, with very different probabilities. Note that the highest-energy path found by T-RRT corresponds to an energy barrier of approximately 6 kcal·mol^{-1}. This example illustrates the capacity of the method to capture a large variety of possible transition paths when necessary. In particular, it can be advantageous compared to many path sampling methods based on an initial trajectory, or requiring a reaction coordinate that bias the search. Regarding the saddle-points that appear in Figure 7-d, 29 paths over 50 pass through S_1, whereas only 10 pass through S_2. Such a distribution is in agreement with results of a restricted perturbation-targeted MD method[51].

IV. CONCLUSION

We have proposed a novel method, called T-RRT, to explore conformational energy landscapes. The method combines recent robotic path planning algorithms with basic concepts of statistical physics. The T-RRT algorithm can be applied to find reachable energy minima from an arbitrary conformation. The same algorithm (possibly with a different parameter setting for improving performance) can also be applied to compute conformational transition paths between pairs of minima. Simple benchmarks have been used in this work to validate the approach, and to facilitate the interpretation of the main features of the method.

This paper aims to provide a basic algorithmic framework that could be extended for treating more complex systems. Similarly to MC-based methods, more sophisticated sampling schemes could be devised to enhance the efficacy of the exploration. Additionally, when a target conformation is specified, a biased scheme could be used to drive more quickly the tree toward the final conformation. In the short future, we expect to investigate extensions of T-RRT to yield a more efficient exploration of the conformational space of longer polypeptides.

ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Ministry of Science and Innovation under project

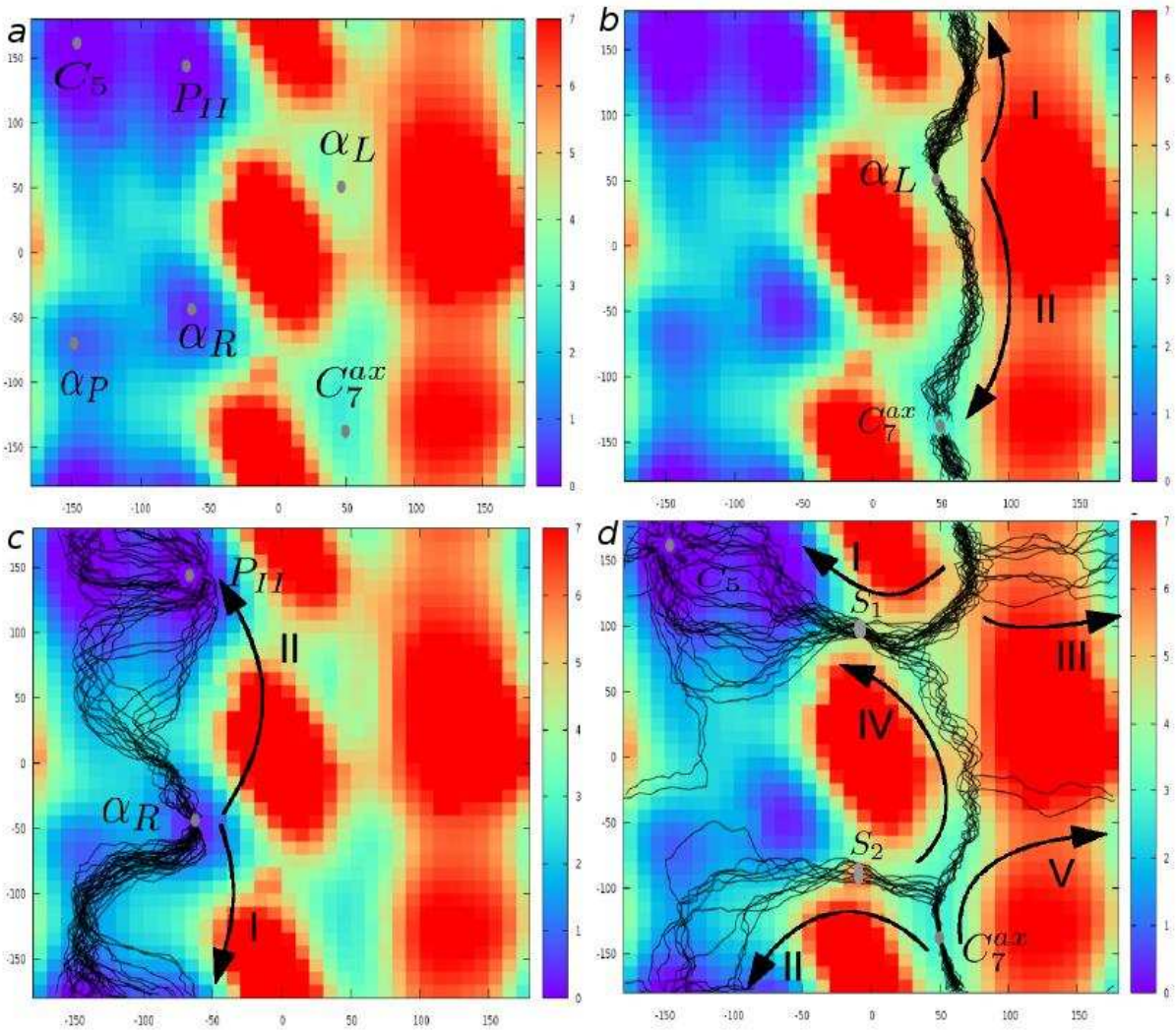

FIG. 7. T-RRT minima and transition paths for the alanine dipeptide projected on the ϕ, ψ energy map. (a) the 6 minima found correctly match the minima of the energy map. (b) The $\alpha_L \rightarrow C_7^{ax}$ transition, (c) $\alpha_R \rightarrow C_5$ transition, (d) $C_7^{ax} \rightarrow C_5$ transition.

DPI2007-60858. Léonard Jaillet was supported by CSIC under JAE-Doc fellowship.

[1] S. O. A. Liwo, C. Czaplewski and H. A. Scheraga, Curr. Opin. Struct. Biol. **18**, 134 (2008).
[2] M. Christen and W. F. van Gusteren, J. Comp. Chem. **29**, 157 (2008).
[3] A. Leach, *Molecular Modelling: Principles and Applications* (Prentice Hall, 2001), 2nd ed.
[4] D. Frenkel and B. Smit, Computational Science Series **1** (2001).
[5] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).
[6] S. Rauscher, C. Neale, and R. Pomès, J. Chem. Theory Comput. **5**, 2640 (2009).
[7] A. F. Voter, J. Chem. Phys. **106**, 4665 (1997).
[8] R. R. Sørensen and A. F. Voter, J. Comput. Phys. **112**, 9599 (2000).
[9] D. Hamelberg, J. Morgan, and J. A. McCammon, J. Chem. Phys. **120**, 11919 (2004).
[10] A. Laio and M. Parrinello, Proc Natl. Acad. Sci. pp. 12562–12566 (2002).
[11] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten, in *Computational Molecular Dynamics: Challenges, Methods, Ideas* (Springer-Verlag, 1998), pp. 39–65.
[12] S. K. Luedemann, V. Lounnas, and R. C. Wade, J. Mol. Biol. **303**, 797 (2000).
[13] G.-H. Wei, N. Mousseau, and P. Derreumaux, J. Chem. Phys. **117**, 11379 (2002).
[14] S. R. Wilson, W. Cui, J. W. Moskowitz, and K. E. Schmidt, Journal of Computational Chemistry **12(3)**, 342 (1990).
[15] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, J. Mol. Biol. **268**, 209 (2008).
[16] J. A. Hegler, J. Lätzer, A. Shehu, C. Clementi, and P. G. Wolynes, Proc. Nat. Acad. Sci. USA **106**, 15302 (2009).
[17] F. Rao and A. Caflisch, J. Mol Biol. **342**, 299 (2004).
[18] F. Noé, D. Krachtus, J. C. Smith, and S. Fischer, J. Chem. Theory Comput. **2**, 840 (2006).
[19] F. Noé and S. Fischer, Current Opinion in Structural Biology **18 (2)**, 154 (2008).
[20] J. D. Chodera N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, J. of Chem. Phys. **126**, 155101 (2007).

[21] D. Prada-Gracia, J. G.-G. nes, P. Echenique, and F. Falo, PLoS Computational Biology **5**, e1000415 (2009).
[22] M. S. Apaydin, A. P. Singh, D. L. Brutlag, and J. C. Latombe, Proc. IEEE Int. Conf. on Robotics and Automation pp. 932–939 (2001).
[23] N. M. Amato, K. A. Dill, and G. Song, J. Comput. Biol. **10(3-4)**, 239 (2003).
[24] G. Mills and H. Jónsson, Phys. Rev. Lett. **72**, 1124 (1994).
[25] G. Henkelman, B. P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).
[26] P. Maragakisa, S. A. Andreev, Y. Brumer, D. R. Reichman, and E. Kaxiras, J. Chem. Phys. **117**, 4651 (2002).
[27] W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B. **66**, 052301 (2002).
[28] W. E, W. Ren, and E. Vanden-Eijnden, J. Chem. Phys. **126**, 164103 (2007).
[29] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[30] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).
[31] R. J. Allen, C. Valeriani, and P. R. ten Wolde, Journal of Physics: Condensed Matter **21** (2009).
[32] E. Paci and M. Karplus, J. Mol. Biol. **288**, 441 (1999).
[33] J. Schlitter, J. Engels, P. Krüger, E. Jacoby, and A. Wollmer, Mol. Sim. **10**, 291 (1993).
[34] L. Mouawad and D. Perahia, J. Mol. Biol. **258**, 393 (1996).
[35] S. Kirillova, J. Cortés, A. Stefaniu, and T. Siméon, Proteins **70**, 131 (2008).
[36] T-RRT was initially introduced as a method to compute good-quality paths of robot systems[42].
[37] R. Battiti, M. Brunato, and F. Mascia, *Reactive Search and Intelligent Optimization*, vol. 45 of *Operations research/Computer Science Interfaces* (Springer Verlag, 2008).
[38] S. M. LaValle and J. J. Kuffner, *Algorithmic and Computational Robotics: New Directions* pp. 293–308 (2000).
[39] J. Cortés, T. Siméon, V. Ruiz de Angulo, D. Guieysse, M. Remaud-Siméon, and V. Tran, Bioinformatics **21(1)**, 116 (2005).
[40] B. Raveh, A. Enosh, O. Schueler-Furman, and D. Halperin, PLoS Comput. Biol. **5**, e1000295 (2009).
[41] S. M. LaValle, *Planning Algorithms* (Cambridge University Press, New York, 2006).
[42] L. Jaillet, J. Cortés, and T. Siméon, IEEE Transactions on Robotics (2009), in press.
[43] These benchmarks were first proposed at the 2005 Workshop on Conformational Dynamics in Complex Systems.
[44] The computing times mentioned in this paper correspond to tests based on a non-optimized code and performed on a single PC with an Intel Core 2 processor at 3.0 GHz.
[45] C. Brooks, D. A.Case, Chem. Rev. **93 (7)**, 2487 (1993).
[46] J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, Multiscale Model. Simul. **5 (4)**, 1214 (2006).
[47] D. S. Chekmarev, T. Ishida, and R. M. Levy, J. Phys. Chem. B **108 (50)**, 19487 (2004).
[48] H. Okumura and Y. Okamoto, J. Phys. Chem. B **112 (38)**, 12038 (2008).
[49] P. G. Bolhuis, C. Dellago, and D. Chandler, Proc. of the National Academy of Sciences of the USA **97**, 5877 (2000).
[50] B. Strodel and D. J. Wales, Chemical Physics Letters **466**, 105 (2008).
[51] A. van der Vaart and M. Karplus, J. Chem. Phys **122**, 114903 (2005).