

# Exploring the Energy Landscapes of Flexible Molecular Loops using Higher-Dimensional Continuation

J. M. Porta and L. Jaillet

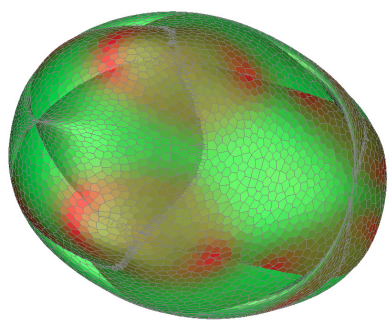
Institut de Robòtica i Informàtica Industrial, UPC-CSIC

Llorens Artigas 4-6, 08028 Barcelona, Spain

## Abstract

The conformational space of a flexible molecular loop includes the set of conformations fulfilling the geometric loop-closure constraints and its energy landscape can be seen as a scalar field defined on this implicit set. Higher-dimensional continuation tools, recently developed in Dynamical Systems and also applied to Robotics, provide efficient algorithms to trace out implicitly defined sets. This paper describes these tools and applies them to obtain full descriptions of the energy landscapes of short molecular loops that, otherwise, can only be partially explored, mainly via sampling. Moreover, to deal with larger loops, this paper exploits the higher-dimensional continuation tools to find local minima and minimum energy transition paths between them, without deviating from the loop-closure constraints. The proposed techniques are applied to previously studied molecules revealing the intricate structure of their energy landscapes.

**Keywords:** Molecular loop modeling, Energy landscapes, Differential geometry, Higher-dimensional continuation, Manifolds, Bifurcations.



This paper introduces the use of higher-dimensional continuation tools to explore the energy landscapes of the implicitly-defined conformational spaces of flexible molecular loops. These tools produce an atlas composed by coordinated charts that parametrize the conformational space. The atlas captures the structure of this space, which determines the motion of the loop and the transition paths between conformations, something that is hard to obtain using existing approaches.

# 1 Introduction

In Biochemistry, the loop modeling problem consist in determining the energetically favorable conformations of a molecular chain that fulfill a set of geometric constraints on its extremes. This problem appears in many relevant situations such as when determining the valid conformations of cyclic molecules, when analyzing the movements of an atom chain with fixed ends, or when trying to complete the missing fragments in protein modeling [26, 62, 1, 51, 55, 36]. In general, the loops are not rigid, but flexible, since they can undergo internal motions. Actually, the function of many biomolecules depends on this flexibility. Unfortunately, the internal motions induce noisy data when trying to characterize the loops using X-ray crystallography and similar techniques. Therefore, loops are typically modeled using computational methods.

Approaches focusing either on the energetic or on the geometric aspects of the loop modeling problem provide complementary insights for such a challenging problem. Geometry-based approaches typically use coarse-grained representations of the problem to provide information about the coordinated motions of the loop in idealized or average cases, an aspect that is hard to apprehend using energy-based techniques. The mainstream of geometry-based approaches is based on sampling the set of valid conformations implicitly given by the loop-closure constraints. Actually, the situation where a solution set is implicitly defined by some constraints appears in many fields and, formally, such set forms a variety, possibly involving several manifolds. This fact and its possible effects on the algorithms for computational chemistry have been largely disregarded until recent works [6, 39]. The immediate consequence is that since the implicit variety is usually intricate and non-parametric, sampling methods have difficulties in fully covering it. Moreover, a collection of samples does not provide direct information about the structure of the variety, which is fundamental to determine the likely transitions between conformations. For instance, conformations with similar potential energy and low root mean square distance can be actually far away when moving on the variety and, thus, the transition between them can be unlikely, despite its apparent proximity.

From Differential Geometry, it is well known that an implicit variety can be described by an atlas containing a collection of charts, each chart providing a local parametrization of the variety [19]. Higher-dimensional continuation techniques provide tools to generate the atlas of the compact subset of a variety reachable from a given point. This paper applies recently developed higher-dimensional continuation techniques to describe the conformational space of molecules (or fragments of molecules) defining one or more loops. For low dimensional problems, we describe a systematic procedure to define an atlas that can be used to estimate the energy landscape readily identifying all the minima and the valid transition paths between them. This procedure can address problems with moderate dimensions by incorporating additional constraints like the ones derived from the Ramachandran plots or from the steric clashes. For larger dimensional problems, the continuation tools can be used to enhance the completeness and the efficiency of existing local exploration techniques. The proposed procedures add new tools to the arsenal of methods for molecular loop analysis that complement existing techniques, with the distinctive feature that the atlas directly reveals the actual underlying structure of the conformation space.

The rest of this paper is organized as follows. Section 2 characterizes the proposed contributions in the context of existing work. Then, Section 3 outlines a technique to built an atlas over an implicit set that is mainly adequate to describe short loops or strictly constrained problems. Section 4 describes the extensions of this technique to deal with higher-dimensional solution sets, and Section 5 applies the proposed methods to several biomolecules. Finally, Section 6 summarizes the contributions and limitations of the proposed approach.

## 2 Related work

The loop modeling problem has been studied in computational chemistry at least from the 60s [30], specially in the context of protein completion. Next, we give a brief review of the techniques that have been used to address this problem and that are related with the approach proposed herein. For a more detailed survey on this subject we refer the reader to [56].

The loop closure problem can be formalized as a search problem and, thus, existing methods can be analyzed considering three key aspects: the representation used for the search space, the function used to guide the search, and the technique used to explore the search space.

The most popular representation of the search space is probably the one that encodes the 3D coordinates for each atom, but other representations are possible [12] such as the one using distances between atoms, or the one using the internal coordinates, which only considers the dihedral angles. This last representation relies on the rigid geometry hypothesis [26, 44, 25] in which bonds lengths and angles are fixed to their equilibrium values. Each representation has its advantages and drawbacks. The 3D coordinates allow for a detailed representation of the conformations, but they are redundant and prone to abrupt changes in the potential energy. In contrast, the dihedral angle representation is significantly more compact, and factors out high-frequency fluctuations of the potential energy due to small-amplitude atom vibrations, concentrating on the fluctuations caused by the coordinated motion of atoms.

As mentioned, in problems involving loops, the search should focus on the energetically and geometrically valid conformations. Since the energy somehow accounts for the geometry too, some methods guide the search considering only energetic aspects. In contrast, other methods focus first on the geometric aspects of the problem, to avoid considering conformations that are for sure invalid. Different representations fit with particular functions to guide the search. For instance, atom-based representations are more adequate for all-atom potential energy functions and dihedral angle representations are typically preferred for geometric-based approaches.

Finally, both the guiding function and the representation have influence in the choice of the technique used to explore the loop conformational space. For instance, in the context of energy-guided approaches, close-and-relax [66] or Molecular Dynamics methods [7, 61] rely on atom-based representation, but Monte Carlo methods [54] use either atom-based [8] or dihedral angle representations [10, 37]. The later is used, for instance, in the concerted motion approaches where the Jacobian of the geometric conditions plays a central role in the transformation necessary to evaluate the Metropolis acceptance criterion [18]. Dihedral angle representations are also used in systematic searches [2, 42]. These methods can easily incorporate additional constraints [48] and provide exhaustive information on the conformational space.

In contrast to the diversity of representations used in the energy-guided methods, geometric-guided approaches typically rely on the dihedral angle representation. Some of these approaches exploit the parallelism between molecules and robots. In this context, closed-form inverse kinematics have been applied to this problem [26, 38, 15]. However, these techniques are limited to a single loops with at most six degrees of freedom, and they have problems at singularities [41]. To address problems with more degrees of freedom, we can resort to numerical techniques like the random tweak [22] or the Jacobian pseudo-inverse method [59], both using the Jacobian and, thus, also with problems at singularities. An alternative is the cyclic coordinate descent method [9], which is not affected by them. Although these techniques can be iterated, they are designed to obtain one solution at a time. Numerical methods able to isolate the whole conformational space and that are not affected by singularities also exist [46]. Finally, problems with more degrees of freedom can also be addressed with local perturbation techniques [58, 35], with approaches where some of the variables are sampled and the rest is solved via inverse kinematics [15, 14] or combining inverse kinematics, fragment assembly, and local optimization methods [26, 36].

Following the previous review, we can say that the techniques presented in this paper represents the loop closure problem using dihedral angles, and that they are essentially geometric. For strictly constrained problems, we present a systematic method based on higher-dimensional continuation. This technique continues the research line initiated in [46] which aims at developing tools to isolate the full conformational space of molecular loops. However, the method presented in [46] is complete, while the method introduced here defines an atlas of the connected component of the conformational space reachable from a given point. This sacrifice in completeness, though, results in a significant increment in computational efficiency. Thus, whereas some of the existing methods are local in nature [7, 61], the proposed procedure provides a global description (of part) of the loop-closure variety and, thus, it gives detailed information on its structure, something only offered by few of the existing methods [55, 13, 6, 39]. This information is obtained directly from the geometric equations, without the need of generating dense sets of points from them [6], with the consequent gain in efficiency. The difference of the proposed approach with respect to other systematic methods using dihedral angles [2, 42] is that the approach introduced here does not explore the space of dihedral angles, but directly the variety of loop-closed conformations. This difference allows us to consider more complex problems, since we exploit the fact that the geometric constraints reduce the dimensionality of the search space. Like existing systematic approaches [2, 42], though, the proposed technique can incorporate additional constraints so that the exploration is focused on the biologically feasible conformations given by the Ramachandra plots [48], or by the steric-clashes, both aspects difficult to incorporate in existing approaches such as the ones relying on closed-form inverse kinematics [26, 38, 15]. A search even more focused on the feasible conformations would require the use of database [21] or fragment assembly methods [55, 36], something not considered in this work.

Both the proposed technique and the one introduced in [46] were initially developed as numerical inverse kinematics approaches, but designed to identify all the possible solutions and not just few of them, as in existing works [22, 59, 9]. Like some of these methods, the procedure presented here is strongly based on the use of Jacobians to generate linear spaces tangent to the loop-closure variety, but in contrast with existing Jacobian-based techniques [22, 18, 59, 36], the proposed procedure is aware of the presence of bifurcations, the only type of singularities that, to the best of our knowledge, has been shown to play a role in the context of Biochemistry [39]. Other types of singularities might affect the performance of the presented system, although their analysis is out of the scope of this paper.

Despite the efficiency of the proposed method, the generation of the atlas for a large dimensional variety is computationally expensive. Therefore, we propose to trade off exhaustiveness by efficiency, exploiting the higher-dimensional continuation tools to perform either a local optimization on the space of closed conformations or a tree-based Monte Carlo like [54, 40, 10, 37] randomized exploration of this space. The advantage of the proposed local optimization technique is that the higher-dimensional continuation techniques provide a parametrization that changes when necessary and, thus, that is valid all over the conformational space except at singularities, which are partially considered by detecting bifurcations. This aspect is not treated in existing approaches [36]. In the tree-based exploration the higher-dimensional continuation tools allow addressing problems with geometric constraints, while the original approach is only valid for unconstrained systems [31].

All the techniques introduced in this paper can deal with multi-loop problems, something only considered in few of the previous works [58, 35, 13]. Finally, note that, like in these works, the proposed techniques require an initial sample from where to start the exploration of the conformational variety. Any of the existing sampling-based methods can be used to generate this initial point [22, 59, 9].

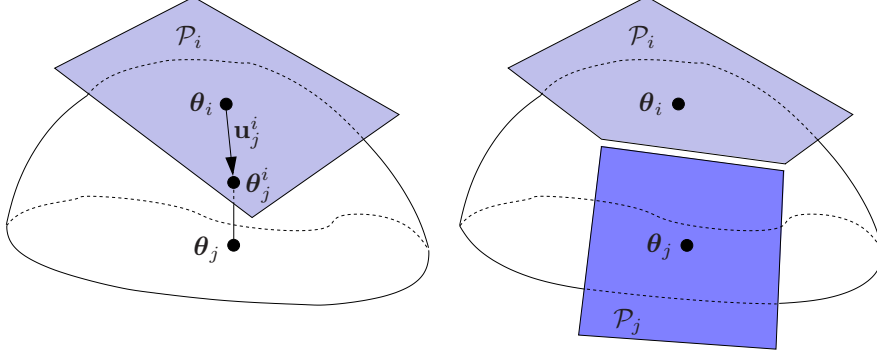


Figure 1: Higher-dimensional continuation method applied to a 2D manifold embedded in a 3D dihedral space. **Left** A chart at a given point,  $\theta_i$ , is defined using the tangent space at this point. The area of applicability in the tangent space of the chart is denoted as  $\mathcal{P}_i$ . A point,  $\theta_j^i$ , is defined using the tangent space and is orthogonally projected on the manifold to determine  $\theta_j$ , the center of a neighboring chart. **Right** The areas of applicability of the first chart,  $\mathcal{P}_i$ , and the new one,  $\mathcal{P}_j$ , are coordinated so that their projections on the manifold slightly overlap.

### 3 Higher-dimensional Continuation

Each molecular loop defines six constraints fixing the relative position and orientation between the extremes of the loop. Thus, the loop constraints determine a system of equations

$$\mathbf{F}(\theta) = 0, \quad (1)$$

implicitly defining a variety of valid conformations, with  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^{n-k}$ ,  $n > k > 0$ ,  $n$  the number of dihedral angles and  $k$  the dimension of the variety. In a non-redundant formulation, the number of equations in  $\mathbf{F}$  is  $m = n - k = 6l$ , with  $l$  the number of independent loops in the problem.

Next, we briefly summarize the main algorithmic tools to generate atlas of implicitly defined varieties introduced by [28, 29, 5] in the context of Dynamical Systems and also applied to Robotics [45, 32]. To simplify the description, we will assume that the variety is actually a manifold until Section 3.3, where we consider the presence of bifurcations.

#### 3.1 Defining a Chart

A chart,  $\mathcal{C}_i$ , locally parametrizes the  $k$ -dimensional manifold around a given valid conformation  $\theta_i$  with a bijective map,  $\theta_j = \psi_i(\mathbf{u}_j^i)$ , between parameters  $\mathbf{u}_j^i$  in  $\mathbb{R}^k$  and conformations  $\theta_j$  on the manifold, with  $\psi_i(\mathbf{0}) = \theta_i$ . Such a map can be implemented using the  $k$ -dimensional tangent space at  $\theta_i$  [49]. An orthonormal basis for this tangent space is given by the  $n \times k$  matrix,  $\Phi_i$ , satisfying

$$\begin{bmatrix} \mathbf{J}(\theta_i) \\ \Phi_i^\top \end{bmatrix} \Phi_i = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}, \quad (2)$$

with  $\mathbf{J}(\theta_i)$  the Jacobian of  $\mathbf{F}$  evaluated at  $\theta_i$ , and  $\mathbf{I}$ , the identity matrix. Using this basis, the mapping  $\psi_i$  is computed by first computing the dihedral angles corresponding to a given vector of parameters in the tangent space

$$\theta_j^i = \theta_i + \Phi_i \mathbf{u}_j^i, \quad (3)$$

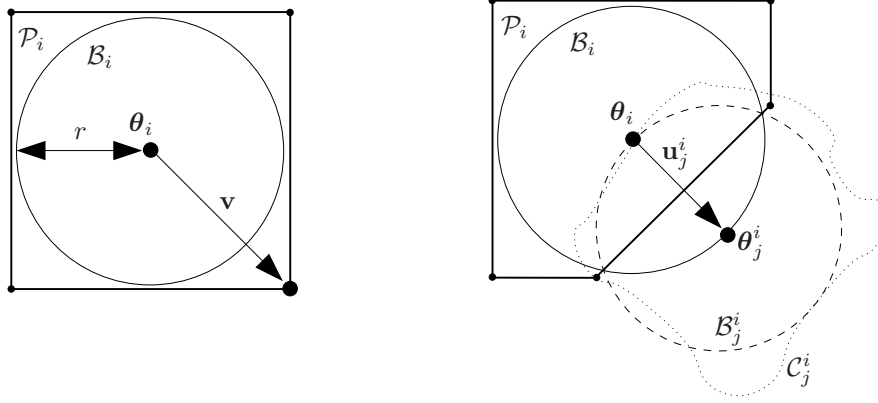


Figure 2: Progressive refinement of the applicability area of a chart for a 2D manifold. **Left** The applicability area,  $\mathcal{P}_i$ , of chart  $\mathcal{C}_i$  is initialized as a box in tangent space including a ball of radius  $r$  around  $\theta_i$ . **Right** The polytope is refined using a ball  $\mathcal{B}_j^i$  that approximates  $\mathcal{C}_j^i$ , the projection on the current chart of the area of the manifold covered by the neighboring chart.

and then, orthogonally projecting  $\theta_j^i$  on the manifold to obtain  $\theta_j$  (see Fig. 1-left). This projection can be computed by solving

$$\begin{cases} \mathbf{F}(\theta_j) = \mathbf{0}, \\ \Phi_i^\top (\theta_j - \theta_j^i) = \mathbf{0}, \end{cases} \quad (4)$$

using a Newton procedure where  $\theta_j$  is initialized to  $\theta_j^i$ . If the process does not converge, the input parameters  $\mathbf{u}_j^i$  are out of  $\mathcal{P}_i$ , the applicability area for chart  $\mathcal{C}_i$ . The areas out of the scope of the chart are to be parametrized by other charts.

### 3.2 Defining an Atlas

Since the applicability area for each chart is limited, the full description of the set of valid conformations requires to define an atlas, i.e., a collection of charts properly coordinated (see Fig. 1-right). The algorithm proposed in [28] gives a systematic way to add new charts to the atlas and to bound their associated applicability areas. In this work,  $\mathcal{P}_i$  is represented as a polytope which is initialized as an hypercube enclosing a ball,  $\mathcal{B}_i$ , of radius  $r$ , defined in the tangent space associated with the chart  $\mathcal{C}_i$ , as shown in Fig. 2-left. A vector,  $\mathbf{v}$ , pointing toward a vertex of  $\mathcal{P}_i$  external to  $\mathcal{B}_i$  is used to generate a new chart. From  $\mathbf{v}$ , a vector of local parameters giving a point on  $\mathcal{B}_i$  is computed as

$$\mathbf{u}_j^i = \frac{\rho}{\|\mathbf{v}\|} \mathbf{v}, \quad (5)$$

with  $\rho$  initialized to  $r$ . A new neighboring chart,  $\mathcal{C}_j$ , is then defined on the point  $\theta_j = \psi_i(\mathbf{u}_j^i)$ . If  $\theta_j$  is too far away from  $\mathcal{C}_i$  or if the curvature between  $\mathcal{C}_i$  and the new chart  $\mathcal{C}_j$  is too large, i.e., if

$$\|\theta_j - \theta_j^i\| > \epsilon, \quad (6)$$

or if

$$\|\Phi_i^\top \Phi_j\| < \cos(\alpha), \quad (7)$$

---

**Algorithm 1:** Computation of the atlas of a loop-closure variety.

---

**AtlasComputation**( $\mathbf{F}, \boldsymbol{\theta}_1, r, \epsilon, \alpha$ )  
**input** : The functions defining the loop’s geometric constraints,  $\mathbf{F}$ , the initial conformation  $\boldsymbol{\theta}_1$ , and the parameters  $r$ ,  $\epsilon$ , and  $\alpha$  used to build the atlas.  
**output**: An atlas  $\mathcal{A}$  of the connected component of the variety including  $\boldsymbol{\theta}_1$ .

```

1  $\mathcal{A} \leftarrow \{\text{NEWCHART}(\mathbf{F}, \boldsymbol{\theta}_1, r)\}$ 
2 while not BOUNDED( $\mathcal{A}$ ) do
3    $\mathcal{C}_i \leftarrow \text{NOTBOUNDEDCHART}(\mathcal{A})$ 
4    $\rho \leftarrow r$ 
5    $\mathbf{v} \leftarrow \text{VERTEX}(\mathcal{P}_i)$  such that  $\mathbf{v} \notin \mathcal{B}_i$ 
6   repeat
7      $\mathbf{u}_j^i \leftarrow \frac{\rho}{\|\mathbf{v}\|} \mathbf{v}$ 
8      $\boldsymbol{\theta}_j^i = \boldsymbol{\theta}_i + \boldsymbol{\Phi}_i \mathbf{u}_j^i$ 
9      $\boldsymbol{\theta}_j = \boldsymbol{\psi}_i(\mathbf{u}_j^i)$ 
10    if  $\|\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^i\| \leq \epsilon$  then
11       $\mathcal{C}_j \leftarrow \text{NEWCHART}(\mathbf{F}, \boldsymbol{\theta}_j, r)$ 
12     $\rho \leftarrow \rho \cdot 0.9$ 
13  until  $\|\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^i\| \leq \epsilon$  and  $\|\boldsymbol{\Phi}_i^\top \boldsymbol{\Phi}_j\| \geq \cos(\alpha)$ 
14   $\mathcal{A} \leftarrow \mathcal{A} \cup \{\mathcal{C}_j\}$ 
15 RETURN( $\mathcal{A}$ )
```

---

for given parameters  $\epsilon$  and  $\alpha$ ,  $\mathcal{C}_j$  is discarded and a new chart is generated from parameters  $\mathbf{u}_j^i$  closer to  $\boldsymbol{\theta}_i$ , i.e., with a smaller  $\rho$  in Eq. (5). This adapts the distribution of the charts to the local curvature of the manifold.

The applicability area of each new chart,  $\mathcal{C}_j$ , has to be properly coordinated with the applicability areas of the neighboring charts already in the atlas. Note that to determine the neighboring charts, the topology of each variable has to be taken into account and, thus, charts defined around  $\pm\pi$  in the dihedral angles are also considered neighbors. In the example in Fig. 2-right, a neighboring chart  $\mathcal{C}_j$  is used to refine  $\mathcal{P}_i$  from the intersection between  $\mathcal{B}_i$  and  $\mathcal{B}_j^i$ , a ball of radius  $r$  that approximates  $\mathcal{C}_j^i$ , the projection on the tangent space of  $\mathcal{C}_i$  of the part of the manifold covered by  $\mathcal{C}_j$ . The hyperplane defined by the intersection of  $\mathcal{B}_i$  and  $\mathcal{B}_j^i$  defines a new face of  $\mathcal{P}_i$  that eliminates some of its vertexes (in particular the one giving  $\mathbf{v}$ ) and generates new ones. Similarly,  $\mathcal{P}_j$ , the polytope associated to  $\mathcal{B}_j$ , is cropped using an approximation of the projection of  $\mathcal{C}_i$  on  $\mathcal{C}_j$ . The applicability areas of the two neighboring charts are not necessarily continuous, but under mild conditions [28] their projections smoothly cover the manifold.

When a chart is fully surrounded by other charts, all the vertexes of its polytope are inside the associated ball and the chart is not further expanded since its applicability area is considered bounded. This process of chart expansion continues as far as there are non-bounded charts in the atlas. At the end of the process, the part of the conformational space containing the initial conformation is fully covered by a set of charts forming an atlas.

Algorithm 1 summarizes the steps to build an atlas. The atlas is initialized with a chart at the given conformation,  $\boldsymbol{\theta}_1$ . Each new chart  $\mathcal{C}_i$  includes the center of the chart,  $\boldsymbol{\theta}_i$ , the basis of the tangent space at this point,  $\boldsymbol{\Phi}_i$ , the approximation of the validity area,  $\mathcal{P}_i$ , and a ball of radius  $r$ ,  $\mathcal{B}_i$ , initially included in  $\mathcal{P}_i$ . The algorithm iterates while there are non-bounded charts, i.e., charts with vertexes of  $\mathcal{P}_i$  out of  $\mathcal{B}_i$ . In the loop (lines 2 to 14) we select one of those non-bounded charts (line 3), take one of its vertexes that are out of the ball (line 5), and generate a



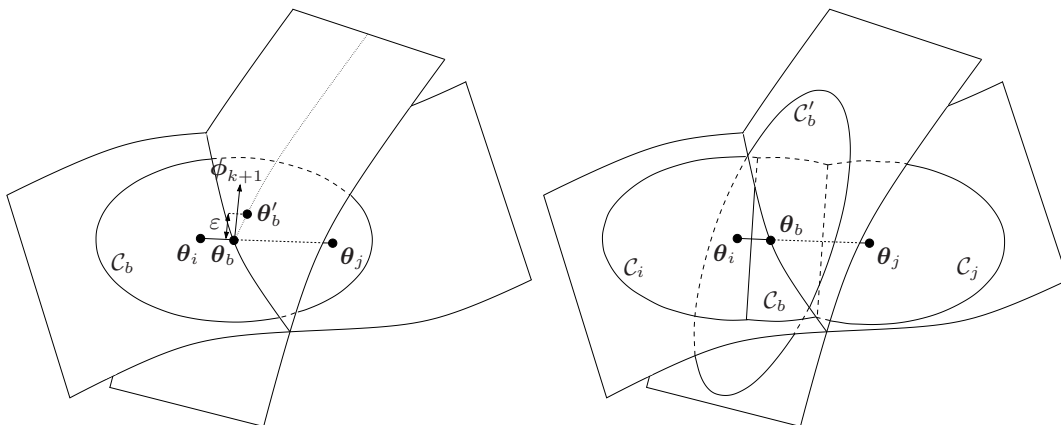


Figure 3: A bifurcation in the conformational space. **Left** Once a bifurcation point,  $\theta_b$ , is located using the indicator function  $\chi$ , a point,  $\theta'_b$ , on the new branch is obtained to estimate the tangent space of this new branch at  $\theta_b$ . **Right** Two new charts,  $C_b$  and  $C'_b$ , are defined on  $\theta_b$  and coordinated with the charts already in the atlas. In this example,  $C_b$  is coordinated with  $C_i$  and  $C_j$ , whereas  $C'_b$  is the first chart on the new branch.

new chart in the direction of this vertex (lines 6 to 13). If necessary, the vector of parameters,  $\mathbf{u}_j^i$  used to generate the center of the new chart is scaled so that the transition between charts is smooth, as defined by Eqs. (6) and (7). Finally, the new chart is added to the atlas (line 14) coordinating it with the charts already forming it.

The number of charts in an atlas scales exponentially with the dimensionality of the set of valid conformations. Thus, the cost of the algorithm at each step is dominated by the cost of two searches among the set of charts: one to find a non-bounded chart (line 3) and another to find the potential neighbors when adding a chart to the atlas (line 14). The first search can be saved keeping the non-bounded charts in a list. The cost of the second search can be reduced using a kd-tree [65].

### 3.3 Dealing with Bifurcations

Typically, the conformational spaces of loops appearing in molecular systems were assumed to be manifold, but it has been shown recently that this is not always the case [6, 39]. If the conformational space is not a manifold everywhere, we need to consider the presence of singularities. Mathematically, a singularity occurs at those points where the Jacobian of  $\mathbf{F}$  is not full rank. Although many types of singularities exist [3], we only consider situations where the variety bifurcates in two branches, that are the only type of singularities that have been shown to play a role in the context of Biochemistry [39]. In these cases, the atlas construction process evolving in one of the branches needs to detect the presence of the other branch and to generate a chart on it from where to extend the atlas. Otherwise, the connected component of the conformational space reachable from the initial point might not be completely represented.

Points on the bifurcation can be located by monitoring a scalar indicator function,  $\chi(\theta)$ , whose value is different for two points  $\theta_i$  and  $\theta_j$  at opposite sides of the bifurcation, and that



vanishes at the bifurcation. One possible indicator function is

$$\chi(\boldsymbol{\theta}) = \text{sign} \left| \begin{array}{c} \mathbf{J}(\boldsymbol{\theta}) \\ \boldsymbol{\Phi}_b \end{array} \right|, \quad (8)$$

where  $\boldsymbol{\Phi}_b$  is an approximation of the tangent space at the bifurcation for the branch including both  $\boldsymbol{\theta}_i$  and  $\boldsymbol{\theta}_j$ . This approximation can be computed, for instance, interpolating between  $\boldsymbol{\Phi}_i$  and  $\boldsymbol{\Phi}_j$ . Thus, if when adding a chart to the atlas (line 14 of Algorithm 1),  $\chi$  varies, the transition between the new chart,  $\mathcal{C}_j$ , and its parent,  $\mathcal{C}_i$ , crosses a bifurcation. In this case and since chart  $\mathcal{C}_j$  is generated from chart  $\mathcal{C}_i$  using  $\boldsymbol{\theta}_j = \boldsymbol{\psi}_i(\mathbf{u}_j^i)$ , a point on the bifurcation,  $\boldsymbol{\theta}_b$ , is located using a dichotomic search along vector  $\mathbf{u}_j^i$ . At a bifurcation point, the kernel of the Jacobian of  $\mathbf{F}$  includes  $k + 1$  vectors. The first  $k$  vectors,  $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_k$ , correspond to those in  $\boldsymbol{\Phi}_b$ . The last vector,  $\boldsymbol{\phi}_{k+1}$ , provides a direction to expand the new branch. Thus, the tangent space for the new branch,  $\boldsymbol{\Phi}'_b$ , can be approximated using the tangent space at a point  $\boldsymbol{\theta}'_b$  on the other branch close to  $\boldsymbol{\theta}_b$ , that can be computed solving

$$\left. \begin{array}{l} \mathbf{F}(\boldsymbol{\theta}) = \mathbf{0} \\ \boldsymbol{\phi}_{k+1}^\top (\boldsymbol{\theta} - \boldsymbol{\theta}_b) = \varepsilon \end{array} \right\} \quad (9)$$

with  $\varepsilon$  small ( $10^{-3}$  in our implementation) and using a Newton process initialized at  $\boldsymbol{\theta}_b + \varepsilon \boldsymbol{\phi}_{k+1}$ , as illustrated in Fig. 3-left. This procedure works well in practice, but it might fail for bifurcations where the two tangent spaces are almost coplanar. In this case, more sophisticated procedures can be used [29, 5]. Finally, the two charts defined at  $\boldsymbol{\theta}_b$  using  $\boldsymbol{\Phi}_b$  and  $\boldsymbol{\Phi}'_b$  are added to the atlas and coordinated with previously added charts, as shown in Fig. 3-right.

## 4 Local Exploration of the Conformational Manifold

The tools presented in the previous section are only practical to trace low dimensional conformational spaces. Nevertheless, these tools can be used to adapt existing algorithm able to deal with high-dimensional problems so that they can also consider the loop-closure constraints.

### 4.1 Local Optimization on the Energy Landscape

The potential energy optimization of a conformation is typically addressed with gradient-based techniques. However, the potential energy functions are usually expressed in terms of the atoms coordinates and following their gradient might lead to an undesired deformation of the loop. When dealing with molecular loops from a geometric perspective, the gradient should be restricted to the variety of valid conformations. Analytically, this can be formulated using the chain rule as

$$\frac{\partial e}{\partial \mathbf{u}} = \frac{\partial e}{\partial \mathbf{a}} \frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{u}}, \quad (10)$$

with  $e$  the scalar atom-based potential energy function,  $\mathbf{u}$  the internal coordinates of the variety,  $\mathbf{a}$  the atom positions, and  $\boldsymbol{\theta}$  the dihedral angles defining the conformation. Eq. (10) projects the gradient from the space of atom coordinates to the space of dihedral angles and, then, to the tangent space defining the current chart. The projection gives a displacement on this tangent space that has to be projected on the variety with the corresponding  $\boldsymbol{\psi}$  mapping to obtain a valid conformation. This conformation can be used to define a new chart from where to continue the minimization process, until it reaches a point where the gradient vanishes. Note that, in practice, the gradient can be approximated numerically considering small displacements on the tangent space.

This local minimization technique is similar, for instance, to the one used in [36]. However, in [36] a fixed parametrization is used, whereas here the parametrization is changed as the minimization process moves on the variety. Using a fix parametrization, any degeneracy of the sub-matrix of the Jacobian formed by the columns corresponding to the variables used as parameters induces a singularity and, thus, restricts the minimization process. When the parametrization changes, only points where the full Jacobian is rank-deficient might cause problems. Some of these points corresponds to bifurcations. When the minimization path crosses a bifurcation, two gradients are possible, one for each intersecting branch of the conformational space and, thus, the minimization path should branch. The possibility of crossing a bifurcation during gradient minimization is something not usually considered in algorithms for computational chemistry.

## 4.2 Randomized Exploration of the Energy Landscape

If instead of locally optimizing a conformation, we are interested in a more global exploration of the energy landscape or in determining low energy paths connecting two given conformations, the T-RRT approach can be used [31]. T-RRT is a sampling-based method for efficiently exploring energetically favorable regions of an energy landscape. Basically, the method is driven by a double strategy. On the one hand, a tree-based exploration strategy inspired from Robotics eagerly explores yet unexplored regions of the space growing the tree toward randomly sampled conformations. On the other hand, a Monte Carlo like transition test rejects samples with too high energy, and guides the expansion toward energetically favorable regions. Moreover, an important feature of the method is the automatic balance between these two strategies achieved thanks to a self-tuning mechanism.

The T-RRT method cannot be applied straightforwardly to molecules involving loops since the conformations randomly generated will not respect the loop-closure constraints. However, it is possible to intertwine the T-RRT exploration with the atlas construction, where the atlas constructed so far is used to sample new conformations for the T-RRT, and the T-RRT is used to determine directions of expansion for the atlas. More precisely, to generate a random sample first a chart  $\mathcal{C}_i$  is selected at random and then rejection sampling is used to get a point in  $\mathcal{P}_i$ , that is projected on the conformational space using the corresponding  $\psi_i$  mapping. The atlas is initialized with just one chart and new charts are added as the tree branches escape from the balls associated with the charts. Note that recently, a related technique also using the tangent space to generate samples has proved useful in studying the local flexibility of proteins with several loops and hundreds of degrees of freedom [63, 64], but in this approach, only one tangent space is used and the projection from the tangent space to the variety is not considered.

## 5 Experiments

The tools described in Sections 3 and 4 have been implemented in C and integrated into the CuikSuite [34], using the GNU Scientific Library for the lineal algebra operations [24] and SOLID [60] for the collision detection. The management of the molecular information as well as the evaluation of potential energies is done via version 2.3.1 of OpenBabel [43, 57]. In all cases, the energies are evaluated using the MMFF94 force field [27], although the presented approach is independent of the actual force field used. The experiments reported below were executed on a single CPU of an Intel Core i7 at 2.93 Ghz running Mac OS.

Figure 4 shows the four molecules used in the experiments. The first one is a multi-loop academic benchmark previously used in [35]. Its structure defines a tetrahedron composed of 4 carbon, 4 hydrogen, and 14 sulfur atoms. For this particular experiment and following the previous work, the bond distance for carbon-sulfur, carbon-hydrogen, and sulfur-sulfur are set

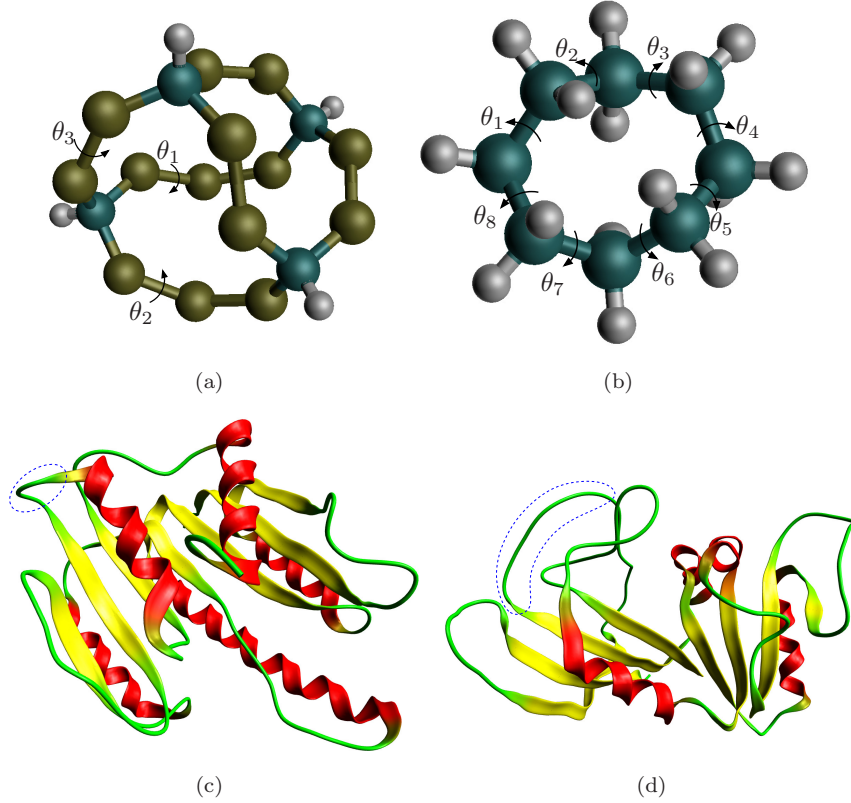


Figure 4: The four molecules used in the experiments. **(a)** A synthetic tetrahedron like molecule. Carbon atoms are represented in cyan, hydrogen in white, and sulfur in gold. **(b)** The cyclooctane using the same color code as in the previous example. **(c)** The FTSJ of escherichia coli (PDB code 1EJ0). The loop marked by a dashed ellipse formed by the residues 74 to 77 is analyzed. **(d)** The dihydrofolate reductase of the lactobacillus casei (PDB code 3DFR). The loop in the dashed region formed by the residues 120 to 124 is analyzed.

to 1.804 Å, 2.019 Å, and 1.120 Å, respectively. Moreover, the carbon and sulfur bond angles are set to  $109.5^\circ$  and  $135^\circ$ . These values are not the canonical ones for the energy field, but they are used to avoid excessive van der Waals overlap [58]. The second molecule used in the experiments is the cyclooctane, a cyclic molecule including eight carbon and sixteen hydrogen atoms. This molecule has received significant attention [16, 33, 50, 20, 4, 11, 47, 46, 39] because it has a complex energy landscape, despite its structural simplicity. The third test case is the loop formed by the residues 74 to 77 of the FTSJ of escherichia coli (PDB code 1EJ0). This loop has been also studied previously [9, 36] and it is used here to show the potential applicability of the proposed method to the protein completion problem. Finally, we consider the loop formed by the residues 120 to 124 of the dihydrofolate reductase of the lactobacillus casei (PDB code 3DFR). This loop that has been analyzed before [23, 17], is used here to demonstrate the applicability of the proposed techniques to problems with conformational spaces of higher dimensionality. In all cases, the initial point to start the analysis is obtained with a Jacobian pseudo-inverse numerical inverse kinematics process [59].

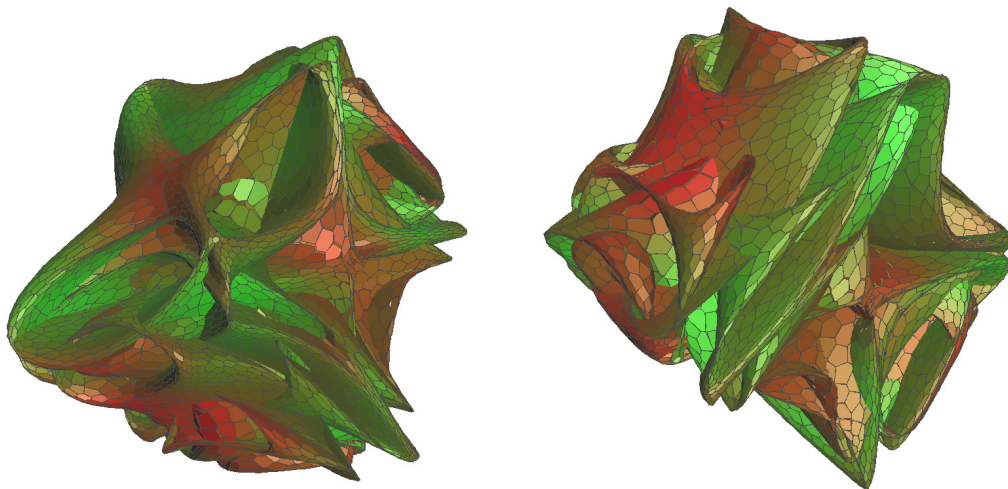


Figure 5: Two complementary views of the full atlas of the tetrahedron molecule projected on  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  (see Fig. 4-a). Each polygon is a chart. Red and green correspond to conformations with high and low energy, respectively.

Figure 5 shows two complementary views of the atlas of the tetrahedron molecule projected on the three dihedral angles  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  labeled in Fig. 4-a. In Fig. 5, each polygon is a chart and the colors correspond to the energy of the conformation at the center of the chart, using red for high energies and green for low ones. The full atlas includes about 15500 charts computed in 34 seconds with  $r = 0.4$ ,  $\epsilon = 0.4$ , and  $\alpha = 0.9$  radians. As a reference, the sampling procedure introduced in [35], takes 60 minutes to generate 500 samples, although using a significantly different software and hardware infrastructure, which makes the direct comparison difficult. The atlas reveals a very intricate structure that would be hard to cover via sampling. Moreover, a collection of samples without any further post-process would not provide information about the possible transitions and might lead to incorrect analysis. However, once the structure of the conformational space is captured in the atlas, it clearly gets easier to compute the local minima of the associated energy landscape and the expected transitions between them, fulfilling the loop-closure conditions at any time. This is the advantage of systematic methods and, in particular of the one proposed herein, with respect to basic sampling-based methods. Additionally, analytic tools could be used to infer the topology of the conformational space from the atlas structure [39], but this analysis is out of the scope of this paper. In this example no bifurcation is detected, but this does not mean that the conformational space is manifold. Singularities other than bifurcations might exist. However, the singularities typically lead to numerical problems around them and we did not observe any numerical issue at the used resolution.

The full atlas of the second example, the cyclooctane molecule, is shown in Fig. 6, projected on  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$ . In this case, the atlas includes about 14000 charts computed in 6 seconds with  $r = 0.2$ ,  $\epsilon = 0.2$ , and  $\alpha = 0.65$  radians. This short execution time is in contrast with the more than 9000 CPU hours required by our previous technique to obtain a full description of the cyclooctane conformational space [46]. The fine resolution of the obtained atlas leads to a smooth estimation of the energy landscape. This estimation is overlaid over the atlas in Fig. 6, using red and green for high and low potential energies, respectively.

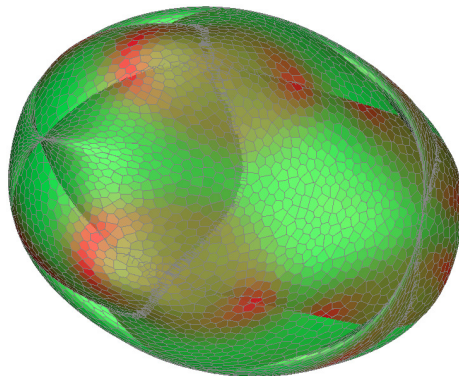


Figure 6: The full atlas of the cyclooctane made of two manifolds. Each polygon is a chart. Red and green correspond to conformations with high and low energy, respectively.

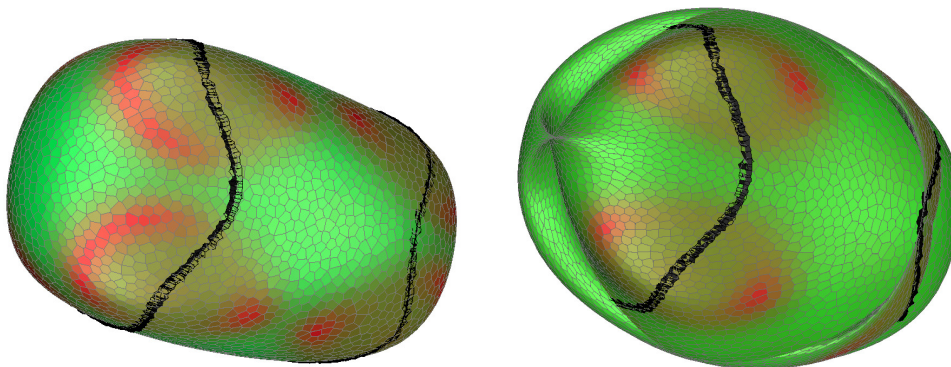


Figure 7: The two manifolds that compose the conformational space of the cyclooctane. Each polygon is a chart. Red and green correspond to conformations with high and low energy, respectively. The bifurcations are represented in black.

When generating the atlas of the cyclooctane, two bifurcations are detected, which confirms that its conformational space is not manifold. This is consistent with the thorough analysis recently provided in [39], departing from a set of samples generated from the loop-closure equations. Figure 7 shows the two manifolds forming the cyclooctane conformational space together with the bifurcation connecting them, shown in black. According to [39], the manifold in Fig. 7-left has the topology of a sphere and the one in the right has a Klein-bottle topology.

The generation of the atlas is an expensive process for high-dimensional conformational spaces. However the tools introduced in Section 4 can be used to partially explore these spaces. As a proof of concept, Fig. 8 shows in blue the path followed by a local energy minimization over the cyclooctane conformational space. Note that in this experiment, a bifurcation is crossed, and, thus, the minimization path branches leading to two different local minima. In the figure, the full atlas is only overlaid as a reference, since it is not computed during the minimization.



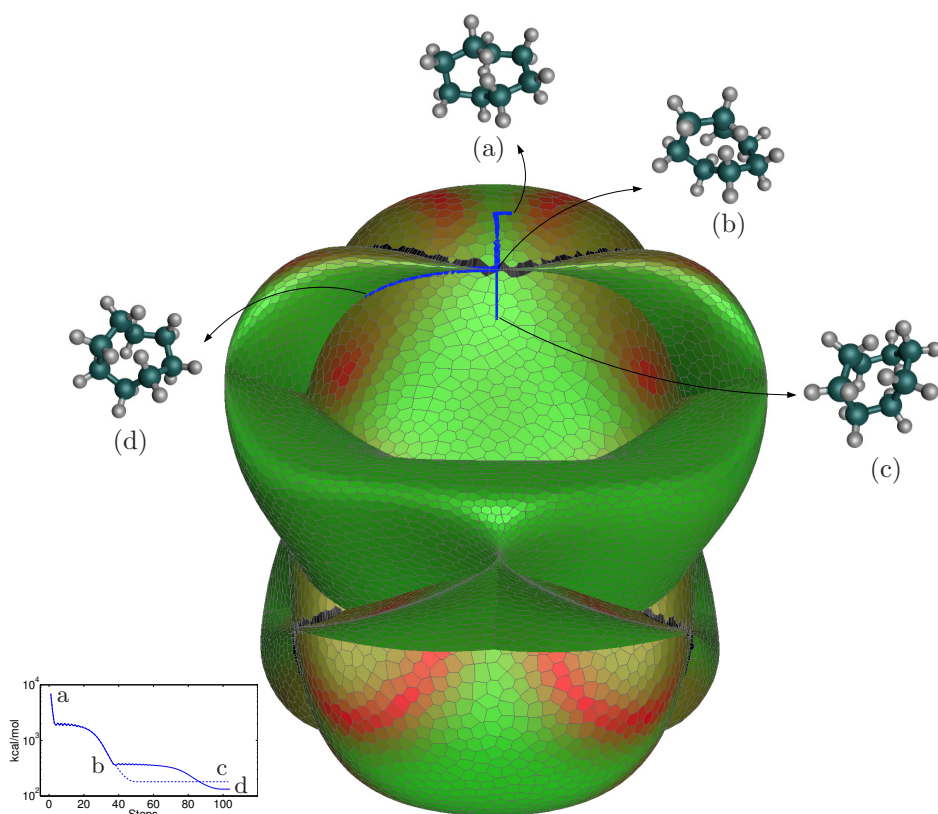


Figure 8: In blue, the projection on  $\theta_1$ ,  $\theta_5$ , and  $\theta_7$  of the paths followed by a local minimization constrained to the set of valid conformations of the cyclooctane. Note that when traversing the bifurcation, the minimization path branches and each subpath leads to a different local minimum. The insets show: a) the initial conformation with high energy due to steric clashes, b) the conformation at the bifurcation, c) a boat conformation that is the local minimum reached on the same manifold as the initial conformation, and d) a TCB conformation that is the local minimum at the other manifold. This second local minimum has a lower potential energy than the first one. The lower-left inset shows the energy profile along the two paths. The path between conformations b) and d) seems to follow an intersection between the two manifolds, but this is just an artifact of the projection.

If we are not just interested in a local minimization, but in determining the minimum potential energy transition between local minima, we can use the T-RRT approach. Figure 9 shows the transition path between two local minima on the cyclooctane. This transition agrees with the diagram given in Fig. 5 in [39].

In the minimization and in the T-RRT experiments on the cyclooctane only 120 and 425 charts are generated, respectively. This is a significant low number taking into account that the full atlas includes about 14000 charts.

In the third benchmark reported in this paper, we analyze the loop formed by the residues 74 to 77 in the 1EJ0 protein, where only the degrees of freedom of the backbone are modeled. Figures 10 left and center show two complementary views of the atlas including 2700 charts

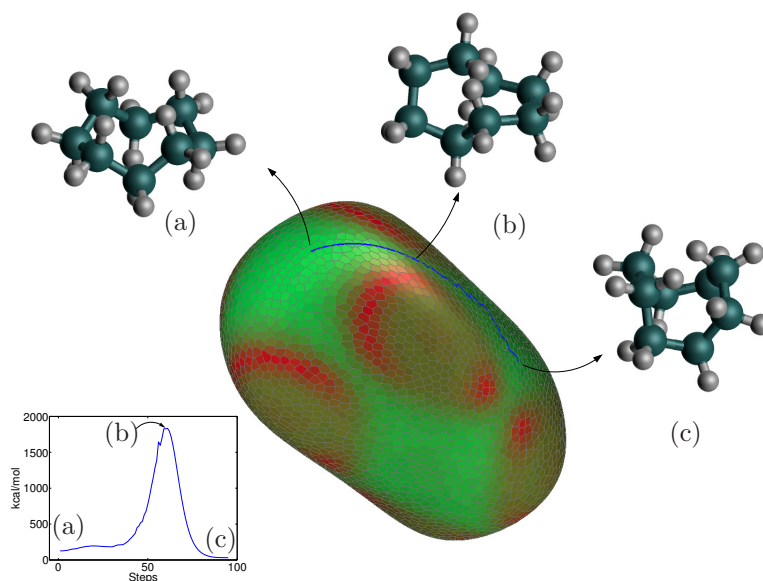


Figure 9: Transition path (in blue) between two local minima in the cyclooctane obtained using T-RRT. This algorithm only generates a partial atlas, but here, the full atlas is shown as a reference. The insets show: a) the initial conformation (a crown), b) the transition state, and c) the final conformation (a boat). The lower-left inset shows the energy profile along the transition path.

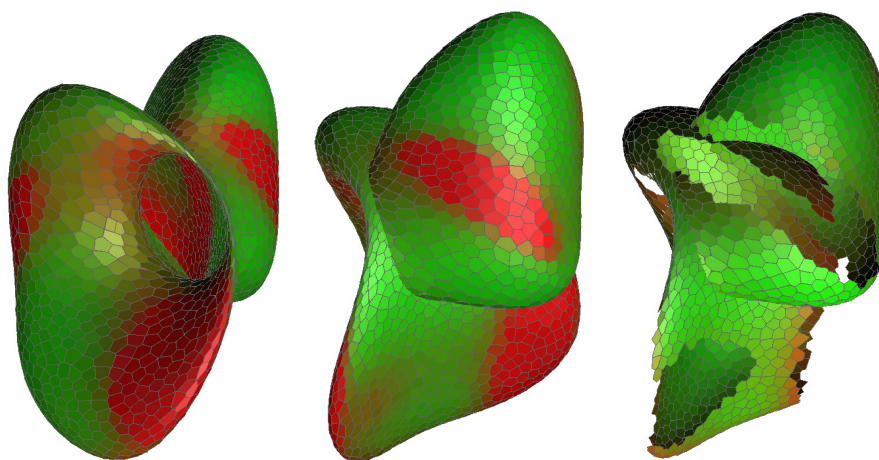


Figure 10: Representations of the conformational space of the 1EJ0-74-77 protein loop. **Left** and **center** Two complementary views of the full atlas. **Right** The atlas without considering steric clashes. Each polygon is a chart. Red and green correspond to conformations with high and low energy, respectively.



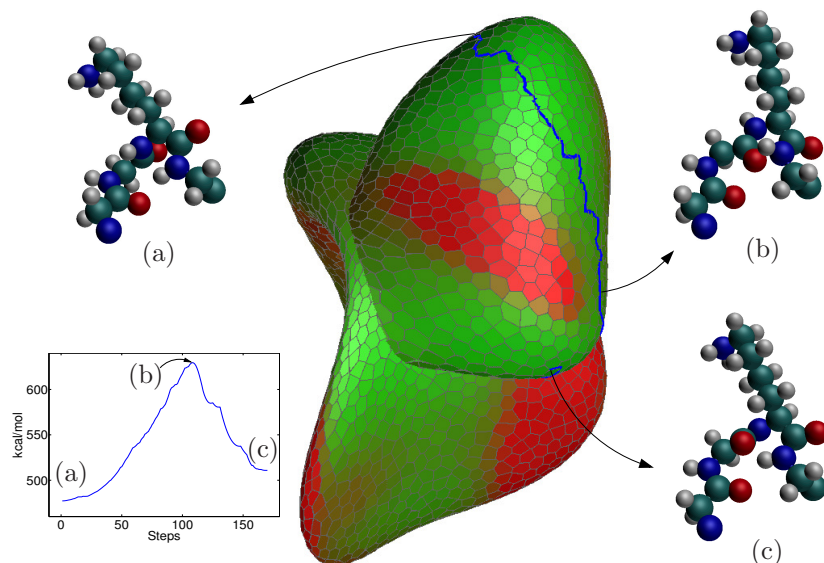


Figure 11: Transition path between two local minima in the analyzed loop of the 1EJ0 protein obtained using T-RRT. The path is shown in blue and the full atlas is not actually generated, but displayed as a reference. The insets show: a) the initial conformation, b) the transition state, and c) the final conformation. In the conformations, only the atoms in the loop are shown and carbon atoms are represented in cyan, hydrogen in white, nitrogen in blue, and oxygen in red. The lower-left inset shows the energy profile along the transition path.

obtained in 10 seconds with  $r = 0.2$ ,  $\epsilon = 0.1$ , and  $\alpha = 0.45$  radians projected on the  $\phi$  dihedral angle of residue 75 and on the  $\phi$  and  $\psi$  angles of residue 76. The topology of the conformational space seems to be a Klein bottle, but further analysis is required to confirm this point [39]. The minimum root mean square error (RMSE) of the charts centers with respect to the crystal structure in the PDB considering all the atoms in the loop, including side chains, is about  $0.2 \text{ \AA}$ , which is comparable to the error obtained with other methods [23, 17]. The conformation in the atlas giving the minimum RMSE is the one with minimum energy all over the conformational space at the given resolution. Setting  $r = 0.1$ , the full atlas is obtained in about 30 seconds and the minimum RMSE is lowered to  $0.1 \text{ \AA}$ .

The conformational space as shown in Figs. 10 left and center includes many areas with high energy due to steric clashes. The atlas construction method introduced in this paper can explore the conformational space avoiding the regions with steric clashes. To this end, collisions between atoms are detected representing the atoms as spheres with a radius proportional the corresponding van der Waals radius (the proportional factor is 0.4 in our experiments). Charts whose center is in collision are considered bounded and not taken for further expansion at line 3 of Algorithm 1. Figure 10-right shows the conformational space free of steric clashes for the considered loop of the 1EJ0 protein. This atlas only includes 1900 charts, a significant reduction with respect to the full atlas for this loop. Any path on this subset of the conformational space has an acceptable energy, however, to determine minimum energy transitions between conformations an approach like the T-RRT should be used. Figure 11 shows the transition path between two conformations of the analyzed loop.

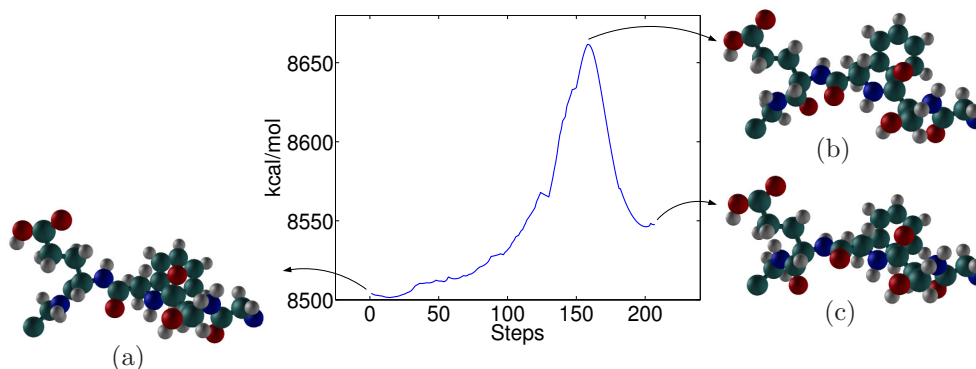


Figure 12: Energy profile along the transition path between two local minima in the analyzed loop of the 3DFR protein obtained using T-RRT. The insets show: a) the initial conformation, b) the transition state, and c) the final conformation. In the conformations, only the atoms in the loop are shown using the same color code as in Fig. 11.

Finally, we analyzed the loop formed by the residues 120 to 125 of the 3DFR protein. Since the conformational space is four-dimensional, it cannot be visualized in 3D like the previous benchmarks. In this case, setting  $r = 0.4$ ,  $\epsilon = 0.1$ , and  $\alpha = 0.45$  the coverage of the conformational space was not achieved after 90 minutes and, at this point, the atlas included more than half a million charts. With  $r = 0.8$ , the full atlas is computed in about 50 minutes and it includes 230000 charts. However, when we activate the collision detection to avoid steric clashes, the atlas with  $r = 0.4$  can be computed in only 500 seconds, including about 43000 charts. This is a clear example of how the extra constraints can dramatically reduce the search space. A similar reduction could be obtained adding the constraints derived from the Ramachandran plots. In the presented approach, this can be achieved defining the adequate ranges for the variables representing the dihedral angles or considering inequalities involving these variables.

In the 3DFR experiment, setting  $r = 0.8$  and avoiding the steric clashes we obtain a RMSE with respect to the crystal structure that is below  $0.45 \text{ \AA}$  in less than 50 seconds. With  $r = 0.4$ , the RMSE reduces to  $0.2 \text{ \AA}$ , but, as mentioned, the computational time increases to 500 seconds. A multi-resolution technique where first the area with minimum energy is identified at a coarse-grained resolution and then it is refined at a fine-grained resolution, could significantly speed up the identification of the most stable conformation, but we leave this as a future work.

Figure 12 showcases the transition path between two conformations of the analyzed loop of the 3DFR protein obtained using the adaptation of the T-RRT algorithm introduced in Section 4.2. This shows that this algorithm can be used to obtain relevant information about the energy landscape, even when the conformational spaces are of high dimensionality.

## 6 Conclusions

In this paper, we have introduced the use of higher-dimensional continuation tools to explore the energy landscape of molecules (or fragments of molecules) including loops. We have shown how these tools can be applied to generate an atlas representing a conformational space, including the case where this space has bifurcations and, thus, it is not manifold everywhere. The atlas provides a complete information about the energy landscape, something that is not readily available when using other types of exploration. Thus, the atlas can be used to determine the local minima and

the transitions between them, information that is crucial for understanding the biological function of the molecule under consideration. For low dimensional conformational spaces, the quality of the information provided by the atlas and the low computational effort required to compute it are remarkable.

This paper also proposed to exploit the higher-dimensional continuation tools to adapt existing algorithms so that they can operate on the varieties arising in the presence of molecular loops. In this line, we presented and evaluated the gradient-based local minimization restricted to the variety, and an adaptation of the T-RRT to find low energy transition paths between local minima. These tools only provide partial information about the conformational space, but they can be used even on high-dimensional varieties, where the construction of the full atlas would be prohibitively expensive.

In the future, we aim to exploit the symmetries that appear in many problems [53] and to speed up the steric clash detection [52]. These improvements might help to reduce the computational burden of the approach in more challenging problems.

## Acknowledgments

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under project DPI2010-18449. Léonard Jaillet was supported by the CSIC under a JAE-Doc fellowship partially founded by the ESF.

## References

- [1] D. Baker and A. Sali. Protein structure prediction and structural genomics. *Science*, 294:93–96, 2001.
- [2] D. D. Beusen, E. R. B. Shands, S. F. Karasek, G. R. Marshall, and R. A. Dammkoehler. Systematic search in conformational analysis. *Journal of Molecular Structure: THEOCHEM*, 370(12):157–171, 1996.
- [3] W.-J. Beyn, A. Champneys, E. Doedel, W. Govarets, U. A. Kuznetsov, A. Yu, and B. Sandstede. *Handbook of Dynamical Systems (Vol 2)*, chapter Numerical Continuation, and Computation of Normal Forms, pages 149–219. Elsevier, 2002.
- [4] R. J. Bharadwaj. Conformational properties of cyclooctane: a molecular dynamics simulation study. *Molecular Physics*, 98(4):211–218, 2000.
- [5] O. Bohigas. Branch switching from singular points. Institut de Robòtica i Informàtica. CSIC-UPC. Technical Report, 2011.
- [6] W. M. Brown, S. Martin, S. N. Pollock, E. A. Coutsiás, and J.-P. Watson. Algorithmic dimensionality reduction for molecular structure analysis. *Journal of Chemical Physics*, 129(6):064118, 2008.
- [7] R. E. Bruccoleri and M. Karplus. Conformational sampling using high-temperature molecular dynamics. *Biopolymers*, 29(14):1847–1862, 1990.
- [8] A. Caflisch, P. Niederer, and M. Anliker. Monte Carlo minimization with thermalization for global optimization of polypeptide conformations in Cartesian coordinate space. *Proteins: Structure, Function, and Bioinformatics*, 14(1):102–109, 1992.

- [9] A. A. Canutescu and R. L. Dunbrack Jr. Cyclic Coordinate Descent: A robotics algorithm for protein loop closure. *Protein Science*, 12(5):963–72, 2003.
- [10] L. Carlacchi and S. W. Englander. Loop problem in proteins: Developments on Monte Carlo simulated annealing approach. *Journal of Computational Chemistry*, 17(8):1002–1012, 1996.
- [11] Z. Chen and F. A. Escobedo. A configurational-bias approach for the simulation of inner sections of linear and cyclic molecules. *Journal of Chemical Physics*, 113:11382, 2000.
- [12] C. Clementi. Coarse-grained models of protein folding: toy models or predictive tools? *Current Opinion in Structural Biology*, 18(1):10–15, 2008.
- [13] J. Cortés, T. Siméon, V. Ruiz de Angulo, D. Guieysse, M. Remaud-Siméon, and V. Tran. A path planning approach for computing large-amplitude motions of flexible molecules. *Bioinformatics*, 21(suppl 1):i116–i125, 2005.
- [14] J. Cortés, T. Siméon, M. Remaud-Siméon, and V. Tran. Geometric algorithms for the conformational analysis of long protein loops. *Journal of Computational Chemistry*, 25(7):956–967, 2004.
- [15] E. A. Coutsiyas, C. Seok, M. P. Jacobson, and K. A. Dill. A kinematic view of loop closure. *Journal of Computational Chemistry*, 25(4):510–528, 2004.
- [16] G. Crippen. Exploring the conformational space of cycloalkanes by linearized embedding. *Journal of Computational Chemistry*, 13(3):351–361, 1992.
- [17] C. M. Deane and T. L. Blundell. A novel exhaustive search algorithm for predicting the conformation of polypeptide segments in proteins. *Proteins: Structure, Function, and Bioinformatics*, 40(1):135–144, 2000.
- [18] A. R. Dinner. Local deformations of polymers with nonplanar rigid main-chain internal coordinates. *Journal of Computational Chemistry*, 21(13):1132–1144, 2000.
- [19] M. P. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976.
- [20] I. Z. Emiris and B. Mourrain. Computer algebra methods for studying and computing molecular conformations. *Algorithmica*, 25(2-3):372–402, June 1999.
- [21] K. Fidelis, P. S. Stern, D. Bacon, and J. Moulton. Comparison of systematic search and database methods for constructing segments of protein structure. *Protein Engineering*, 7(8):953–960, 1994.
- [22] R. M. Fine, H. Wang, P. S. Shenkin, D. L. Yarmush, and C. Levinthal. Predicting antibody hypervariable loop conformations II: Minimization and molecular dynamics studies of MCP603 from many randomly generated loop conformations. *Proteins: Structure, Function, and Bioinformatics*, 1(4):342–362, 1986.
- [23] A. Fiser, R. K. G. Do, and A. Sali. Modeling of loops in protein structures. *Protein Science*, 9(9):1753–1773, 2000.
- [24] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi. *GNU Scientific Library Reference Manual*. Network Theory Ltd., 2009.
- [25] K. D. Gibson and H. A. Scheraga. Energy minimization of rigid-geometry polypeptides with exactly closed disulfide loops. *Journal of Computational Chemistry*, 18(3):403–415, 1997.

- [26] N. Go and H. A. Scheraga. Ring closure and local conformational deformations of chain molecules. *Macromolecules*, 3(2):178–187, 1970.
- [27] T. A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 17(5-6):490–519, 1996.
- [28] M. E. Henderson. Multiple parameter continuation: Computing implicitly defined k-manifolds. *International Journal of Bifurcation and Chaos*, 12(3):451–476, 2002.
- [29] M. E. Henderson. Multiparameter parallel search branch switching. *International Journal of Bifurcation and Chaos in Applied Science and Engineering*, 15(3):967–974, 2005.
- [30] J. B. Hendrickson. Molecular Geometry. I. Machine Computation of the Common Rings. *Journal of the American Chemical Society*, 83(22):4537–4547, 1961.
- [31] L. Jaillet, F. J. Corcho, J.-J. Pérez, and J. Cortés. Randomized tree construction algorithm to explore energy landscapes. *Journal of Computational Chemistry*, 32(16):3464–3474, 2011.
- [32] L. Jaillet and J. M. Porta. Path planning with loop closure constraints using an atlas-based RRT. *International Symposium on Robotics Research*, 2011.
- [33] I. Kolossváry and W. Guida. Comprehensive conformational analysis of the four- to twelve-membered ring cycloalkanes: Identification of the complete set of interconversion pathways on the MM2 potential energy hypersurface. *Journal of the American Chemical Society*, (115):2107–2119, 1993.
- [34] KRD Group. The CuikSuite. <http://www.iri.upc.edu/research/webprojects/cuikweb>, Last accessed August 2012.
- [35] A. Lee, O. Brock, and I. Streinu. A methodology for efficiently sampling the conformation space of molecular structures. *Physical Biology*, 2(4):S108–15, 2005.
- [36] J. Lee, D. Lee, H. Park, E. A. Coutsiias, and C. Seok. Protein loop modeling by using fragment assembly and analytical loop closure. *Proteins: Structure, Function, and Bioinformatics*, 78(16):3428–3436, 2010.
- [37] S.-H. Lee, K. Palmo, and S. Krimm. A new formalism for molecular dynamics in internal coordinates. *Chemical Physics*, 265(1):63–85, 2001.
- [38] D. Manocha and J. F. Canny. Efficient inverse kinematics for general 6R manipulators. *IEEE Transactions on Robotics and Automation*, 10:648–657, 1994.
- [39] S. Martin, A. Thompson, E. A. Coutsiias, and J.-P. Watson. Topology of cyclo-octane energy landscape. *Journal of Chemical Physics*, 132:234115, 2010.
- [40] N. Metropolis and N. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(267):335–341, 1949.
- [41] R. J. Milgram, G. Liu, and J.-C. Latombe. On the structure of the inverse kinematics map of a fragment of protein backbone. *Journal of Computational Chemistry*, 29(1):50–68, 2008.
- [42] J. Moult and M. N. G. James. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins: Structure, Function, and Bioinformatics*, 1(2):146–163, 1986.

- [43] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(33), 2011.
- [44] K. A. Palmer and H. A. Scheraga. Standard-geometry chains fitted to X-ray derived structures: Validation of the rigid-geometry approximation. I. chain closure through a limited search of “loop” conformations. *Journal of Computational Chemistry*, 12(4):505–526, 1991.
- [45] J. M. Porta, L. Jaillet, and O. Bohigas. Randomized path planning on manifolds based on higher-dimensional continuation. *The International Journal of Robotics Research*, 31(2):201–215, 2012.
- [46] J. M. Porta, L. Ros, F. Thomas, F. Corcho, J. Cantó, and J.-J. Pérez. Complete maps of molecular-loop conformational spaces. *Journal of Computational Chemistry*, 28(13):2170–2189, 2007.
- [47] J. Pérez, K. Nolsøe, M. Kessler, L. García, E. Pérez, and J. L. Serrano. Bayesian methods for the conformational classification of eight-membered rings. *Acta Crystallographica Section B*, 61(5):585–594, 2005.
- [48] G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, 1963.
- [49] W. C. Rheinboldt. MANPACK: A set of algorithms of computations on implicitly defined manifolds. *Computers and Mathematics with Applications*, 32(12):15–28, 1996.
- [50] W. R. Rocha, J. R. Pliego, S. M. Resende, H. F. Dos Santos, M. A. De Oliveira, and W. B. De Almeida. Ab initio conformational analysis of cyclooctane molecule. *Journal of Computational Chemistry*, 19(5):524–534, 1998.
- [51] C. A. Rohl, C. E. M. Strauss, D. Chivian, and D. Baker. Modeling structurally variable regions in homologous proteins with Rosetta. *Proteins: Structure, Function, and Bioinformatics*, 55(3):656–677, 2004.
- [52] V. Ruiz de Angulo, J. Cortés, and J. M. Porta. Rigid-CLL: Avoiding constant-distance computations in cell linked-lists algorithms. *Journal of Computational Chemistry*, 33(3):294–300, 2012.
- [53] V. Ruiz de Angulo and C. Torras. Exploiting single-cycle symmetries in continuous constraint problems. *Journal of Artificial Intelligence Research*, 24:499–520, 2009.
- [54] M. Saunders. Stochastic search for the conformations of bicyclic hydrocarbons. *Journal of Computational Chemistry*, 10(2):203–208, 1989.
- [55] A. Shehu, C. Clementi, and L. E. Kavraki. Modeling protein conformational ensembles: From missing loops to equilibrium fluctuations. *Proteins: Structure, Function, Bioinformatics*, 65(1):164–179, 2006.
- [56] A. Shehu and L. E. Kavraki. Modeling structures and motions of loops in protein molecules. *Entropy*, 14:252–290, 2012.
- [57] The Open Babel Package. Version 2.3.1. <http://www.openbabel.org>, Last accessed August 2012.
- [58] M. F. Thorpe and M. Lei. Macromolecular flexibility. *Philosophical Magazine*, 84(13-16):1323–1331, 2004.

- [59] H. van den Bedem, I. Lotan, J.-C. Latombe, and A. M. Deacon. Real-space protein-model completion: an inverse-kinematics approach. *Acta Crystallographica Section D*, 61(1):2–13, 2005.
- [60] G. van den Bergen. The SOLID web page. <http://www.dtecta.com>, Last accessed August 2012.
- [61] W. F. van Gunsteren and H. J. C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition in English*, 29(9):992–1023, 1990.
- [62] W. J. Wedemeyer and H. A. Scheraga. Exact analytical loop closure in proteins using polynomial equations. *Journal of Computational Chemistry*, 20(8):819–844, 1999.
- [63] P. Yao, A. Dhanik, N. Marz, R. Propper, C. Kou, G. Liu, H. van den Bedem, J.-C. Latombe, I. Halperin-Landsberg, and R. B. Altman. Efficient algorithms to explore conformation spaces of flexible protein loops. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5(4):534–545, October 2008.
- [64] P. Yao, L. Zhang, and J.-C. Latombe. Sampling-based exploration of folded state of a protein under kinematic and geometric constraints. *Proteins: Structure, Function, and Bioinformatics*, 80(1):25–43, 2012.
- [65] A. Yershova and S. M. LaValle. Improving motion planning algorithms by efficient nearest neighbor searching. *IEEE Transactions on Robotics*, 23(1):151–157, 2007.
- [66] Q. Zheng, R. Rosenfeld, S. Vajda, and C. Delisi. Determining protein loop conformation using scaling-relaxation techniques. *Protein Science*, 2(8):1242–1248, 1993.