



**HAL**  
open science

# Algorithmics of motion: From robotics, through structural biology, toward atomic-scale CAD

Juan Cortés

► **To cite this version:**

Juan Cortés. Algorithmics of motion: From robotics, through structural biology, toward atomic-scale CAD. Robotics [cs.RO]. Université Toulouse III Paul Sabatier, 2014. tel-01110545

**HAL Id: tel-01110545**

**<https://theses.hal.science/tel-01110545>**

Submitted on 28 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# HABILITATION A DIRIGER DES RECHERCHES (HDR)

delivrée par

*l'Université de Toulouse*

présentée au

*Laboratoire d'Analyse et d'Architecture des Systèmes*

par : **Juan Cortés**

Chargé de Recherche CNRS

---

## ALGORITHMICS OF MOTION: FROM ROBOTICS, THROUGH STRUCTURAL BIOLOGY, TOWARD ATOMIC-SCALE CAD

---

Soutenue le **18 Avril 2014**

devant le Jury composé de :

<b>Michael Nilges</b>	Institut Pasteur, Paris	<b><i>Rapporteurs</i></b>
<b>Frédéric Cazals</b>	INRIA Sophia Antipolis-Mediterrane	
<b>Stéphane Doncieux</b>	ISIR, UPMC-CNRS, Paris	
<b>Oliver Brock</b>	Technische Universität Berlin	<b><i>Examineurs</i></b>
<b>Jean-Pierre Jessel</b>	IRIT, UPS, Toulouse	
<b>Jean-Paul Laumond</b>	LAAS-CNRS, Toulouse	
<b>Pierre Monsan</b>	INSA, Toulouse	<b><i>Membre invité</i></b>
<b>Thierry Siméon</b>	LAAS-CNRS, Toulouse	<b><i>Parrain</i></b>



# Contents

	<b>Curriculum Vitæ</b>	<b>1</b>
<b>1</b>	<b>Research Activities</b>	<b>5</b>
1.1	Introduction . . . . .	5
1.2	Methodological advances . . . . .	6
1.2.1	Algorithmic developments in a general framework . . . . .	6
1.2.2	Methods for molecular modeling and simulation . . . . .	13
1.3	Applications in robotics . . . . .	17
1.3.1	Manipulation and Grasping . . . . .	17
1.3.2	Motion planning for aerial robots . . . . .	19
1.4	Applications in structural biology, biotechnology and materials science . . . . .	21
1.4.1	Modeling conformational transitions of proteins and peptides . . . . .	21
1.4.2	Analyzing and predicting molecular interactions . . . . .	24
1.4.3	Polymer modeling . . . . .	26
<b>2</b>	<b>Other Activities</b>	<b>27</b>
2.1	Teaching and Advising Activities . . . . .	27
2.1.1	Teaching . . . . .	27
2.1.2	Advising . . . . .	28
2.2	Organizational and Editorial Activities . . . . .	29
2.2.1	Chairing activities, organization de conferences et workshops . . . . .	29
2.2.2	Editorial activities . . . . .	30
2.2.3	Other organizational and evaluation duties . . . . .	30
2.3	Projects and collaborations . . . . .	30
2.3.1	Contractual projects . . . . .	30
2.3.2	Other collaborations . . . . .	32
2.4	Mobility . . . . .	34
2.4.1	Disciplinary mobility . . . . .	34
2.4.2	Geographic mobility . . . . .	34
2.5	Software development . . . . .	35
2.5.1	Move3D . . . . .	35
2.5.2	MoMA . . . . .	35
2.6	Technology transfer, industrial relationships . . . . .	36
<b>3</b>	<b>Research Project</b>	<b>37</b>
3.1	Overview . . . . .	37
3.2	Methodological developments . . . . .	38
3.3	Robotics applications: motion planning and control of aerial robots . . . . .	39
3.4	Towards integrative approaches in structural biology . . . . .	40
3.5	Computational Protein Design . . . . .	41

3.6	Towards atomic-scale CAD . . . . .	42
4	<b>List of Publications</b>	<b>45</b>
	<b>Annex 1: Selected Publications</b>	<b>51</b>

# Curriculum Vitæ

**Full name:** Juan Cortés-Mastral  
**Date of birth:** 15/08/1974  
**Place of birth:** Zaragoza, Spain  
**Marital status:** Married, 1 child



**Professional address:** LAAS-CNRS  
7 Av. Colonel Roche, F-31400 Toulouse, France  
**Phone:** +33 5 61 33 63 45 **Fax:** +33 5 61 33 64 55  
**E-mail:** [juan.cortes@laas.fr](mailto:juan.cortes@laas.fr)  
**Web:** <http://homepages.laas.fr/jcortes/>

## Professional Experience

- **CNRS Researcher** at LAAS-CNRS (Toulouse) From Oct. 2004  
*Area:* Algorithmics of motion, with applications in robotics, structural biology, biotechnology and materials science
- **Invited Professor** at IRI (CSIC-UPC, Barcelona) Nov. 2008 - Nov. 2009  
*Subject:* New approaches to peptide and polymer modeling
- **Post-doctoral Researcher** at UFIP (Univ. Nantes-CNRS) Mar. 2004 - Sep. 2004  
*Subject:* Computational prediction and analysis of protein loop motions
- **PhD thesis preparation** at LAAS-CNRS (Toulouse) Oct. 2000 - Dec. 2003  
*Subject:* Motion planning algorithms for close-chain mechanisms
- **Undergrad. Research Assistant** at LAAS-CNRS (Toulouse) Aug. 1999 - Sep. 2000  
*Subject:* Participation to the development of Move3D:  
a generic path-planning software platform

## Education

- **PhD** - Computer science / Robotics December 2003  
Institut National Polytechnique de Toulouse (France)
- **Diplôme d'Etudes Approfondies (Master)** - Control / Robotics July 2000  
Institut National Polytechnique de Toulouse (France)
- **Engineering Degree** - Specialization in Control and Robotics June 2000  
Centro Politécnico Superior - Universidad de Zaragoza (Spain)

## Awards

- **Prix Jean Lagasse** : Award to the best PhD theses of the CNRS STIC department, 2004
- **Prix Léopold Escande** : Award to the best PhD theses of the INP Toulouse, 2004

## Summary of Research and Teaching Activities

### Research Topics

- **Basic research, Robotics:** Algorithmics of motion. Motion computation and analysis for complex systems. Applications in robotics, mainly to motion planning and manipulation task planning.
- **Interdisciplinary research:** Development of computational methods to study protein flexibility and interactions, for protein design, and polymer modeling. Applications in structural biology, biotechnology and materials science.

### Publications

- 20 articles in scientific journals on robotics (*IEEE Trans. Robot., Int. J. Robot. Res.*), bioinformatics and theoretical physical chemistry (*Bioinformatics, J. Comput. Chem., Proteins, Polymer, Nucl. Acids Res.*, among others).
- 4 book chapters.
- 23 publications in proceedings of international conferences and workshops.
- 5 invited talks, and 9 seminars in French and foreign universities.
- Google Scholar *h-index* = 20 ; 1133 citations.

### Advising activities

- Advisor or co-advisor of 8 PhD students and 6 post-docs.
- Advisor of 12 undergraduate students.
- Scientific advisor of 1 CNRS Research Engineer.

### Teaching

- Co-organizer and teacher of the international, interdisciplinary school “Algorithms in Structural Bioinformatics”, 2012, 2013. ( <http://algosb.sciencesconf.org/> )
- Member of the teaching team of the master program “Ingénierie de la Matière : Modélisation des Processus Physiques”, Université de Toulouse, since 2008.
- Teacher in the “Master en Simulación de Polímeros y Biopolímeros”, Universidad Politécnica de Catalunya (Barcelona, Spain), 2011.
- Teacher in the “Master en Systèmes Automatiques, Informatiques et Décisionnels”, Université de Toulouse, 2008-2010.
- Teacher in the “Master de Robótica” of the Universidad Politécnica de Valencia (Spain), 2008.
- Teacher in the “Summer School on Image & Robotics”, CIC Mexico, 2007.

---

## Organizational and editorial activities

- Co-chair of the IEEE RAS Technical Committee on Algorithms for Planning and Control of Robot Motion, since 2009.
- Co-organizer of the AAAI Workshop on “Artificial Intelligence and Robotics Methods in Computational Biology”, 2013.
- Co-organizer of the ACM-BCB Tutorial “From Robot Motion Planning to Modeling Structures and Motions of Biological Molecules”, 2013.
- Co-organizer of the RSS Workshop “Motion Planning: From Theory to Practice”, 2010.
- Program Committee Member of the Computational Structural Bioinformatics Workshop (CSBW), 2011-2013.
- Program Committee Member of the international conference ECAI 2008, 2010.
- Associated Editor of the IEEE international conferences ICRA and IROS, 2009-2013.
- Reviewer of : IEEE Journals and Conferences (T-RO, T-ASE, ICRA, IROS, ...), The International Journal of Robotics Research, Workshop on the Algorithmic Foundations of Robotics, Journal of Computations Chemistry, Journal of Chemical Theory and Computation, Bioinformatics, Archives of Biochemistry and Biophysics, ...

## Software development

- Coordinator of the developments of MoMA, a software platform for molecular modeling.
- Participation in the development of Move3D, a generic motion planning software.

## Mobility

- **Disciplinary mobility:** Being robotics my basic research area, I am interested in the development of methods in structural biology and materials science, as well as in applications to biotechnologies and nanotechnologies.
- **Geographic mobility:** One-year stay (Nov. 2008 - Nov. 2009) at the l'Universitat Politècnica de Catalunya (Barcelona, Spain), in the framework of international mobility programs of the CNRS and the INPT.

## Participation in projets

- **Robotics projects:** 6 European projects (ARCAS, SAPHARI, DEXMART, PHRIENDS, MOVIE, MOLOG), 1 French national project (IRASIS).
- **Computational biology / Molecular modeling projects:** 3 French national projects (ProtiCAD, GlucoDesign, NanoBioMod), 1 regional project (AMYLO), 1 trans-regional project (AMOBIO), 3 local projects (ITAV-ALMA, OTIMASU, AMORO), 1 CNRS project (BioMove3D), 1 ADEME project.



# Chapter 1

## Research Activities

### 1.1 Introduction

Algorithmics of motion is a fundamental part of robotics. It deals with the development of methods to compute, simulate and analyze motions of physical systems such as articulated mechanisms or mobile robots. Algorithmics of motion is my basic research domain, and this is the reason why I carry out my research activities within a robotics group. Nevertheless, my scientific interests go beyond robotics. Indeed, since the end of my PhD thesis, my activities are focused on interdisciplinary research in the areas of structural bioinformatics and materials science. In particular, I am interested in the development of methods for modeling biological molecules and polymers, as well as in the applications of these methods in biology, biotechnology and nanotechnology. Motivated by this scientific interest, I have actively participated in the emergence of a new research axis at LAAS-CNRS on “Molecular Motion Algorithms”. This interdisciplinary research topic involves interactions and collaboration with experts in biochemistry, biophysics and biology. The continuous contact with these other disciplines along the past ten years has been a very rich experience that has allowed me to acquire knowledge in these areas, as well as the ability to communicate with scientists from different background, which is essential in this context.

The main problem addressed by my research concerns motion of complex systems. Structural bioinformatics and materials science are very interesting areas for the development of methods since they involve extremely challenging problems related to motion. Doubtless, biological macromolecules (such as proteins or RNA) and polymers, at the atomic scale, can be seen as extremely complex mobile systems. The development of methods for modeling/simulating motions of such systems is essential to better understand their physicochemical properties and biological functions. These methods are also of key importance for advances in application areas such as health (e.g. drug design), biotechnology (e.g. engineered enzymes for bio-catalysis) or nanotechnologies (e.g. nucleic acid-based nano-devices). Nevertheless, whenever possible, methodological developments motivated by the complexity of molecular systems are mostly carried out in a general, ab-

stract manner, thus enabling their application in other domains, particularly in robotics. Even if it correspond to a minor part in my research activities, I continue working on robot motion and manipulation planning, mostly in the framework of European projects.

The next sections present a summary of my contributions. They are structured into three categories : methodological advances, applications in robotics, and applications in structural biology, biotechnology and materials science.

## 1.2 Methodological advances

My main methodological contributions concern motion planning algorithms for complex systems. Complexity in this context mainly concerns high-dimensionality together with multiple types of motion constraints, being kinematic loop closure an important class of them. I have also investigated variants of motion planning algorithms able to provide good-quality solutions when a cost-function is defined over the space being explored. Parallel implementations of these algorithms have also been studied.

Most algorithmic developments have been done in a general perspective, the underlying principles being applicable to problems in robotics, in structural bioinformatics, and possibly in other domains such as computer animation or computer-aided manufacturing. Nevertheless, I have focused efforts on specific variants and implementations of these algorithms to deal with molecular systems [4]. In the context of structural bioinformatics, I have contributed with the development of approaches combining motion planning algorithms and other computational techniques, as well as with geometric methods to filter data and to analyze results provided by such techniques. In addition, I have worked on basic methods associated with motion planning, such as collision detection, aiming to enhance the overall performance of the algorithms. These various methodological contributions are summarized below.

### 1.2.1 Algorithmic developments in a general framework

#### Motion planning for closed-chain mechanisms

Mobile systems with kinematic loop-closure constraints are challenging for motion planning. Indeed, such constraints imply changes in the topological properties of the configuration space, which require a reformulation of the problem with respect to the “basic” motion planning problem for open-chain mechanisms [23]. Closed-chain mechanisms are, however, frequent in robotics. Parallel robots are a clear example of this. Closed kinematic chains are also created when several manipulators grasp an object simultaneously. Algorithms for computing motions of closed kinematic chains are also important for applications in structural bioinformatics. For instance, simulating motions of cyclic molecules or flexible fragments of proteins (called loops) requires dealing with loop-closure constraints. Figure 1.1 represents several closed-chain systems.

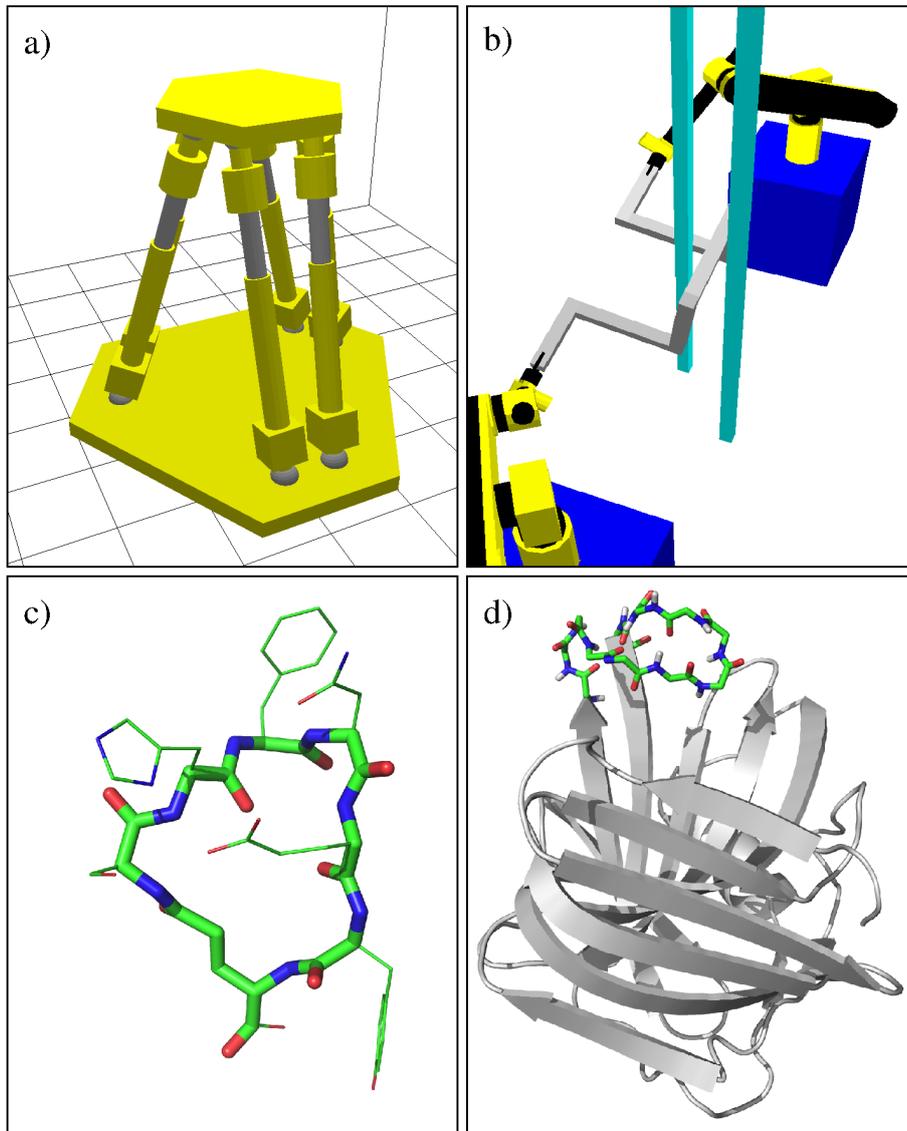


Figure 1.1: Illustrations of closed kinematic chains in robotics and structural biology: a) A parallel robot; b) Two manipulators grasping an object; c) A cyclic peptide; d) A flexible protein fragment with fixed ends.

My thesis work [77, 23] proposed a framework for the extension of sampling-based motion planning algorithms to deal with closed-chain mechanisms. In the following years, I have continued with the improvement of algorithms in this direction, and I have investigated other approaches to closed-chain motion planning based on space-subdivision techniques. The following paragraphs provide more detailed explanations on these works.

*Extending sampling-based algorithms:* The configuration space of a closed-chain mechanism is a lower-dimensional subset of the joint space (i.e. the space of all the variable parameters describing the spatial configuration of the system). While the joint space can be represented by an  $n$ -dimensional polytope, the configuration space is composed by a set of manifolds, whose parameterization is not straightforward. Therefore, sampling and connecting configurations cannot be done using simple techniques as for open-chain mechanisms. Indeed, sampling-based motion planning algorithms, such as the *Probabilistic Roadmap* (PRM) planner or the *Rapidly-exploring Random Tree* (RRT) planner, require extensions to enable their applicability to closed-chain mechanisms. My thesis work [77, 23] proposed a framework for such extensions, and provided algorithmic tools for efficient configuration sampling. During the PhD thesis of Mokhtar Gharbi, we have introduced an improvement that permits to compute paths involving the reconfiguration of the system, which may be necessary to solve some classes of problems, particularly when manipulating objects with several robotic arms. The proposed algorithm [39] extends the PRM planner by including a multi-layer structure of the configuration space, and by considering an explicit treatment of singular configurations that connect these layers.

*A resolution-complete, subdivision-based method:* Sampling-based motion planning algorithms are efficient and general methods, able to treat difficult problems in high-dimensional spaces. However, they present a weak completeness guarantee, namely *probabilistic completeness*, which simply states that the method will find an existing solution if sufficient computing time is provided. Although such a weak completeness property is not a real drawback for most practical applications, some other applications may require stronger guarantees. For instance, when using motion planning algorithms in virtual prototyping for product development, the capacity to determine if a design is incorrect because motion between two configurations is impossible is of great importance. In the context of structural biology, completeness may be important for an accurate analysis of transitions between meta-stable states of molecular systems. In collaboration with researchers at IRI (CSIC-UPC, Barcelona), we have developed a resolution-complete motion planning method for closed-chain mechanisms [40]. The approach is based on a *branch and prune* subdivision algorithm, which was originally developed to solve inverse kinematics problems of serial mechanisms and direct kinematics of parallel robots. This algorithm is able to provide a complete representation of the configuration space of a closed-chain mechanism, with an arbitrary resolution. In the framework of a collaborative project, we have extended this algorithm for considering collision detection, and we have combined it with several motion planning techniques. This approach is currently applicable (in practice) to systems involving a moderate number of degrees of freedom. Variants and extensions to deal with a wider range of problems could be investigated in the future.

### Exploration of very-high-dimensional spaces

High-dimensionality is probably the main challenge for motion planning algorithms, as for any method that performs exploration of a search/state space. Sampling-based motion planning algorithms, such as PRM and RRT, are able to efficiently solve problems involving a moderately high number of variables (i.e. a few tens of degrees of freedom). This is enough for most applications in robotics. In structural biology, however, most problems require to deal with a much higher number of variables. Indeed, models of biological macromolecules may involve thousands of degrees of freedom. We have investigated two directions to deal with very-high-dimensional spaces: One is based on a functional decomposition of the degrees of freedom, and the other applies a dimensionality reduction approach.

*Decoupling degrees of freedom:* We have developed an extension of the RRT algorithm, called Manhattan-like RRT (ML-RRT) [42, 16], to treat high-dimensional problems that admit a particular structuring of the configuration space. The main idea is to decompose degrees of freedom into groups depending on their relevance/role in the motion planning problem. *Active* degrees of freedom are associated with the motion of objects/parts that are essential for treating the problem, whereas *passive* parameters correspond to parts that need to move only if they hinder the motions of other mobile parts (active or passive). ML-RRT treats active parameters at each iteration, as in a basic RRT algorithm. The subset of parts associated with passive parameters that may block the active parts motion are then identified, and moved subsequently. Such a decoupled treatment of degrees of freedom significantly improves the performance of the algorithm, enabling the solution of classes of problems up to now intractable in practice.

The active/passive partition can be clearly illustrated in the case of *molecular disassembly* problems, in which path planning algorithms can be used to simulate protein-ligand unbinding. In this type of problems, the important parameters are those defining the relative pose (i.e. position and orientation) of the ligand with respect to the protein. These parameters are considered to be active parameters within ML-RRT. The parameters corresponding to the internal flexibility of both molecules are considered to be passive, since they only need to be modified if rigid-body disassembly is not possible. The idea is illustrated in Figure 1.2. Section 1.4 will further explain the application of ML-RRT to structural biology problems.

During the PhD thesis of Duc Thanh Le, we have developed several extensions of ML-RRT. A first extension was aimed to enlarge the applicability of the algorithm. It involves a hierarchical representation of the mobile system, with different degrees of mobility (also called passiveness levels) associated to parts. The main interest of this extension was to better deal with protein flexibility when simulating protein-ligand interactions [12], but the idea could also be applied to disassembly problems in constrained workspaces involving a large number of mobile parts. With respect to the basic ML-RRT, which was able to treat the flexibility of the ligand and the protein side-chain, the extended version enables

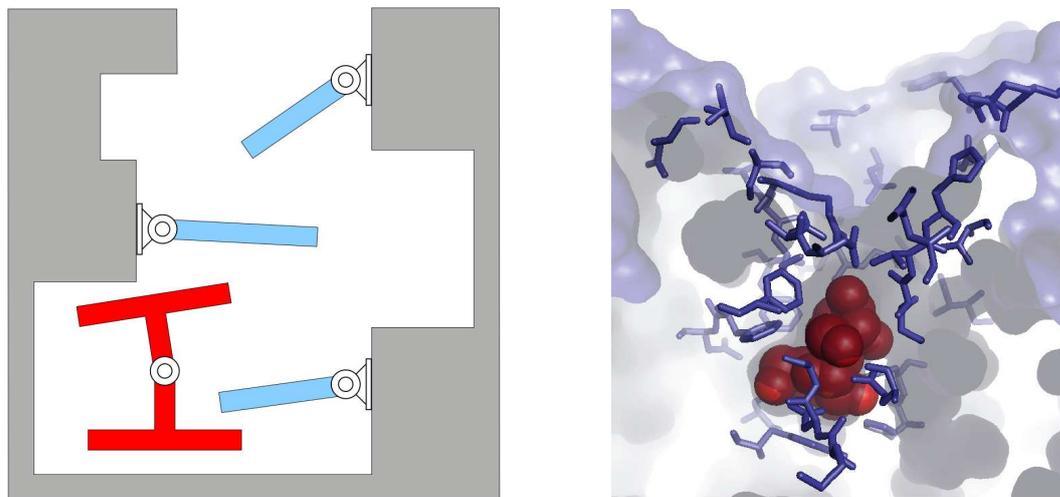


Figure 1.2: The image on the left illustrates an academic disassembly planning problem for two articulated objects. An analogy can be made with the protein-ligand disassembly problem represented in the right-hand image.

the introduction of flexible backbone fragments. This extension permits the simulation of movements of parts of the protein structure (e.g. loops, helices) induced by interactions with the ligand.

Another variant, called Iterative-ML-RRT (I-ML-RRT), was developed for simultaneously (dis)assembly sequencing and path planning [36]. Indeed, both are parts of a same problem that can be formulated in a general path planning framework. The idea is to perform iterative calls to the ML-RRT algorithm, using different active/passive partitions, until all the parts of a general assembly are extracted. Since the method does not rely on any sophisticated geometric computations but only uses collision detection, it can be applied to complex CAD models.

Finally mention that we have also applied the principle of functional decomposition in a PRM framework [35], in the particular context of manipulation planning with multi-arm robots, during the PhD Thesis of Mokhtar Gharbi (this application will be further discussed in Section 1.3). In this case, the idea is to compute independent roadmaps for different parts of the system, and then, to construct a super-graph for planning motions of the whole system by the composition of elementary roadmaps.

*Dimensionality reduction using collective degrees of freedom:* Another approach to deal with very-high-dimensions spaces applies dimensionality reduction techniques to enable the tractability of problems. The underlying principle is to project the high-dimensional space on a lower-dimensional one, which is sufficiently representative for solving the problem (or a simplified instance of it). One of the most popular dimensionality reduction techniques is Principal Component Analysis (PCA). Although PCA has been successfully applied in a

few works in the context of motion planning, it presents two important drawbacks. First, for performing the analysis, it requires a data set that can be difficult to obtain. Second, it is based on a linear approximation, which can be unsuitable for many systems.

We have proposed a different approach, which can be applied to compute motions of complex systems that move by the action of a set of forces represented by a potential energy function. Our developments have been made in the context of molecular system, but the approach could be applied to other types of complex systems. For instance, the application to compute motions of self-organized multi-agent systems could be investigated. The approach relies on the combination of Normal Mode Analysis (NMA) and motion planning algorithms [17, 28, 1]. NMA is based on an harmonic approximation of the potential energy. Several works have shown that low-frequency normal modes are a good approximation of large-amplitude motions of molecules, in which atom motions are highly correlated. Our method searches for paths in the (high-dimensional) conformational space of the molecule by performing an RRT-like exploration of linear combinations of low-frequency normal modes. Since the harmonic approximation of NMA is valid only in the vicinity of the initial structure used for the analysis, normal modes are recomputed at intermediate states in order to enable an accurate simulation of large-amplitude motions. More explanations on the interest of the approach will be provided in Section 1.4.

### Exploration of continuous cost-spaces

Motion planning has traditionally aimed at finding feasible, collision-free paths for a mobile system. However, beyond feasible solutions, in many applications it is important to compute good-quality paths with respect to a given cost criterion. When a cost function is defined on the configuration space of the system, motion planning becomes a path-finding problem in a continuous cost-space. Figure 1.3 illustrates on simple examples the difference between a basic motion planning problem and a problem involving a cost-function. In robotics, the cost function associated with robot configurations may be defined from the distance to obstacles in order to find high-clearance solution paths. It may also be related to controllability, to energy consumption, or to many other different criteria. In molecular applications, the cost function is usually defined by the potential energy or the free energy of the molecular system. Computing low energy paths is important since they correspond to the most probable conformational transitions.

We have developed a variant of the RRT algorithm, called Transition-RRT (T-RRT) [13], to compute good-quality paths in high-dimensional continuous cost-spaces. The idea is to integrate a stochastic state-transition test, similarly to the Metropolis Monte Carlo method, which makes the exploration get focused on low-cost regions of the space. The algorithm involves a self-tuning mechanism that controls the difficulty of this transition test depending on the evolution of the exploration process, and which significantly contributes to the overall performance of the method. T-RRT is a simple and general algorithm that can take into account any type of continuous, smooth cost function defined on the configuration space. It has been successfully applied to diverse robot path-planning

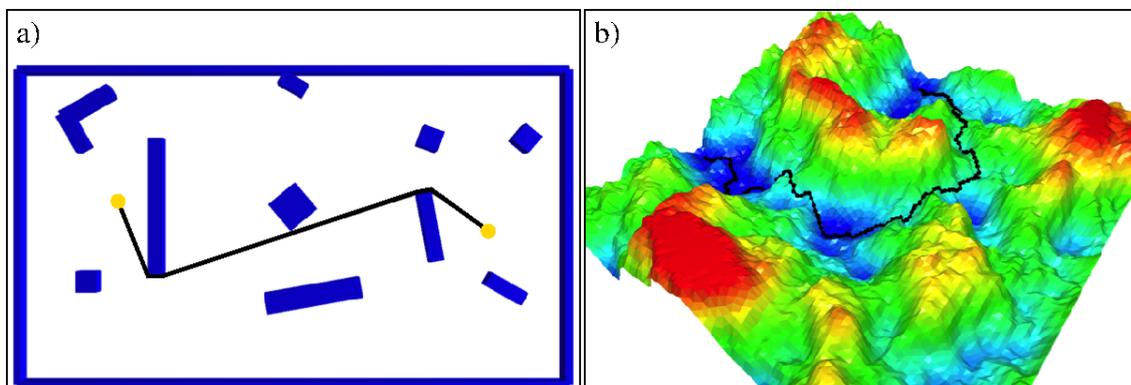


Figure 1.3: a) Illustration of a basic motion planning problem in 2D: All collision-free paths are equally valid. Optimality may depend on the path length. b) Cost-based motion planning in 2D: The quality of the solution depends on the cost values along the path.

problems [13, 33, 21] as well as structural biology problems [7]. Some of these applications will be further discussed in next sections.

Recently, we have worked on variants and improvements of the basic T-RRT algorithm. One of these variants involves its combination with the ML-RRT algorithm. This variant, called MLT-RRT [30], is aimed to treat more efficiently classes of problems in which the cost/energy function can be decomposed into elementary terms associated with different groups of degrees of freedom. Other improvements of T-RRT are aimed to enhance the efficiency of the the algorithm for treating problems in large-scale workspaces. We have investigated different strategies building on improvements proposed in the literature for the basic RRT [27]. The analysis of the performance of these variants on a set of problems shows that devising a general strategy that performs well in all the scenarios is not easy. Significant improvements were only obtained when using a bi-directional implementation of the algorithm, which simultaneously constructs two trees rooted at the initial and the goal configurations. The idea has been further developed with the Multi-T-RRT algorithm [25], which constructs  $n$  trees from a set of configurations. These configurations may correspond to waypoints defined by the user in the case of robotics applications [25], or can be a set of low-energy conformations of a molecule obtained with energy minimization techniques [26].

### Parallel algorithms

The complexity of the problems we treat, both in robotics and in the molecular context, tend to increase along the years. The mobile systems tend to have a higher number of degrees of freedom, which generally implies that motion planning algorithms require a higher number of iterations to solve problem. In addition, motion constraints and cost functions tend to be more complex, which implies that additional computing time is required to evaluate the feasibility or the quality of configurations of the system during the

planning/exploration process. A possible way to face such an increasing computational cost is to develop parallel algorithms, able to exploit multiprocessor architectures. We have developed parallel versions of RRT-like algorithms, which are a type of methods we largely apply in our works. More precisely, we have analyzed several parallelization paradigms, and we have proposed implementation schemas on distributed-memory architectures, using the *Message Passing Interface* (MPI) [31, 3].

### 1.2.2 Methods for molecular modeling and simulation

#### A new mechanistic approach to protein modeling

In the framework of our methodological developments in the molecular context, we have proposed a new mechanistic approach to protein modeling. The main principle is to divide long polypeptide chains into short fragments involving three amino acid residues, which will be called *tripeptides* from now on. The idea is illustrated in Figure 1.4. Such a subdivision enables a multi-scale treatment of the protein. At the coarse-grained scale, each tripeptide is considered as a single (oriented) particle, to which a reference frame is attached. The atomic-scale model can be efficiently reconstructed from the tripeptide-based, coarse-grained model using methods from robot kinematics. Indeed, a tripeptide can be modeled in a similar way as a robot manipulator with six revolute joints. Thus, a semi-analytical inverse kinematics method (like the one explained below) can be applied to obtain the coordinates of the all-atom model from the pose of the reference frames associated with the particles of the coarse-grained model.

On the basis of this mechanistic model, we have proposed a unified approach to devise move classes for a more efficient sampling of the conformational space of proteins using Monte-Carlo-like methods [29] (an extended version of this work, including a more detailed analysis of results, is in preparation). This new modeling approach has also allowed us to improve the performance of a method to simulate conformational transitions of proteins [1] (see Section 1.4.1), and other applications will be explored.

#### Geometric filtering for the prediction of molecular interactions

In addition to methods to sample or to explore the conformational space of molecular systems, we have proposed a geometric filter to improve the performance of protein-protein docking prediction techniques (see Section 1.4.2 for explanation about the problem). Some of these techniques apply an exhaustive sampling of the relative pose of the two proteins in a first stage of the docking prediction procedure. The methods we have developed is aimed at filtering samples that will probably not lead to a stable, low-energy conformation of the protein complex. The proposed filtering method, called *Ray Casting Filter* (RCF) [8], is based on computer graphics techniques. It is able to efficiently identify conformations presenting a weak complementarity between the surfaces of the two proteins, and which, in principle, are unlikely starting points for a docking process. The integration of RCF

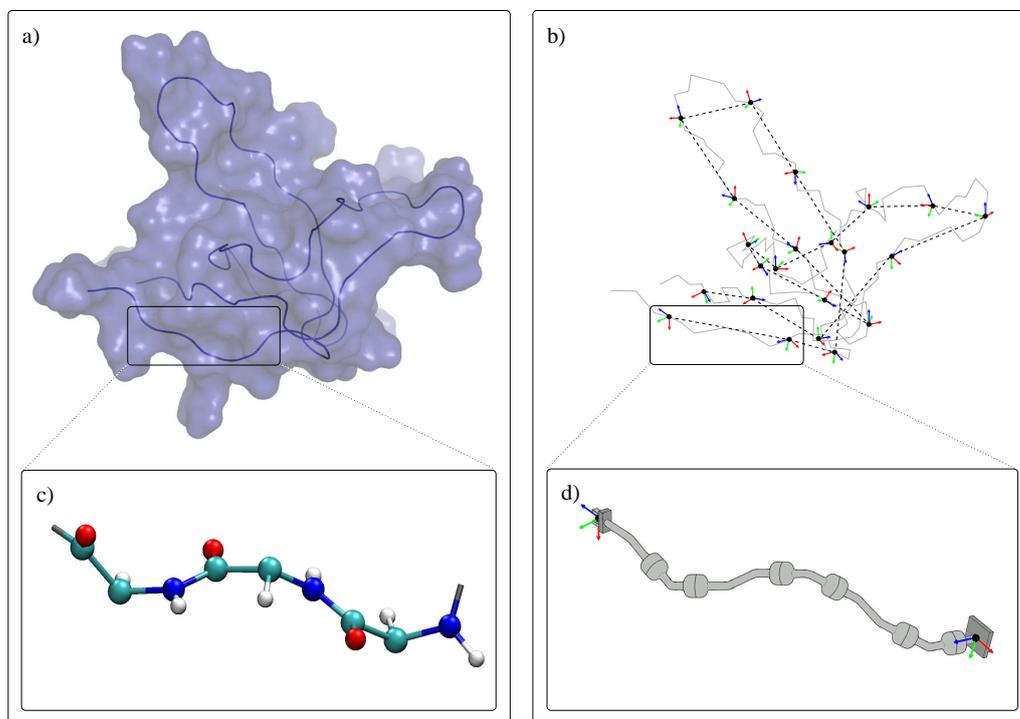


Figure 1.4: Illustration of the protein subdivision approach. Fragments of three amino-acid residues are treated as kinematic chains, similar to robot manipulators.

within ATTRACT, a popular protein-protein docking method, has provided very good results, especially for enzyme-enzyme inhibitor complexes. In addition to a significant speedup (results are obtained about 15 times faster in average compared with the basic ATTRACT method), the filtering process removes some false positives, which improves the quality of docking predictions.

It is interesting to mention here that the basic ideas of RCF have been subsequently applied in robotics to devise a grasping method [32]. The method computes enveloping grasps with a multi-fingered robotic hand, searching to maximize the contact surface between the held object and the hand's surface in a given pre-grasp configuration.

### Improvements of basic methods

A very significant part of the computing time required by all the aforementioned algorithms is spent in basic operations such as distance computation, nearest neighbor search, collision checking, or forward/inverse kinematics calculations. Thus, the efficiency of these basic methods is essential for the good performance of the overall algorithm. As a part of my research, I have participated in the development of several such important “low-level” methods.

*Distance computation and collision checking:* In collaboration with colleagues at the Institut de Robòtica i Informàtica Industrial (IRI, CSIC-UPC, Barcelona), we have developed several algorithms for distance computation and collision checking on atomic-scale molecular models. A first algorithm, called BioCD [22], was specifically designed to be used within sampling-based motion planning algorithms applied to proteins described as kinematic chains. BioCD uses hierarchical data structures to approximate the shape of the molecules at successive levels of detail, making the number of atom pairs tested for collision to be significantly reduced.

The drawback of BioCD is that it strongly relies on implicit structural information inherent in proteins, and thus requires reimplementations for other molecule types. To circumvent this problem, we have developed another method for inter-atomic distance computation and collision checking that can be applied to any molecule type. The method is a variant of the *Cell Linked-Lists* (CLL) algorithm, which is widely used in molecular simulations. CLL-like algorithms use a three-dimensional grid to reduce the computational complexity of neighbor search. We have proposed a variant involving a reorganization of the stages of the algorithm, and the insertion of the atom in the grid by rigid groups, aiming to avoid distance computations between atom pairs that have a fixed relative position. Results show that, in many cases, this improvement can reduce computing time by more than one order of magnitude with respect to the standard algorithm.

*Inverse kinematics:* In collaboration with Marc Renaud (Professor at INSA, Toulouse), we have improved and implemented a semi-analytical inverse kinematics method for  $6R$  mechanisms (i.e. serial mechanisms involving six revolute joints) with arbitrary geometry. The method is based on algebraic elimination theory, and develops an ad-hoc resultant formulation inspired by the seminal work of Lie and Liang. The elimination procedure leads to a system of quadratic polynomial equations, which can be linearized and treated as a generalized eigenvalue problem, for which efficient and robust numerical solutions are available. This inverse kinematics method is a key component in our conformational sampling and exploration algorithms for proteins, peptides and polymers [11, 29, 1]. A publication of the method is in preparation.

### Tools for the analysis of results

Methods to explore the conformational space of biological macromolecules, including molecular dynamics simulations, Monte Carlo methods and robotics-inspired algorithms, generate huge amounts of data that can be difficult to analyze. To facilitate this analysis, one possibility is to project the results of the exploration of a high-dimensional space into a lower dimensional data structure. We have developed such a projection technique using a voxel map [10]. The voxel map representation has been chosen for different reasons: 1) It is a simple and regular structure, which facilitates operations such as nearest-neighbor search. 2) It is 3D, which permits a visual rendering of the information gathered during the conformational search. The choice of the three dimensions of the voxel map depend on

the application. They may be directly a subset of the conformational parameters, or other variables that can be computed from the conformation. The voxel size also depends on the application, and on the chosen coordinates. Indeed, the resolution is chosen depending on the motion amplitude, and on the cost of the ulterior treatment (i.e. geometric and/or energetic analysis). In our works, we have applied this technique to analyze results provided by RRT-based methods. Nevertheless, the voxel map representation is a general approach that could be applied to analyze results obtained with other conformational exploration methods such as molecular dynamics simulations.

### **Coupling methods for the understanding of molecular systems**

The algorithms presented in Section 1.2.1, when applied to molecular models, can directly provide very useful information about molecular motions and interactions [4]. Nevertheless, such robotics-inspired algorithms can be further more interesting when integrated in a more global methodology, together with other computational and experimental methods, for a better understanding of molecular systems.

From our first works on the study of protein flexibility [18], we have investigated an interdisciplinary approach, coupling robotics-inspired methods and molecular modeling techniques. We have mainly developed a two-stage approach: 1) In a first stage, path-planning algorithms are applied to explore the conformational space of a mechanistic model of the molecular system. 2) In a second stage, results of the previous explorations are refined and analyzed using more accurate models and state-of-the-art molecular modeling techniques, such as energy minimization. This approach enables the simulation of large-amplitude conformational transitions with very few computation resources. As will be discussed in Section 1.4, such molecular motions are essential for protein function, especially in the case of enzymes [10, 9, 5].

First results provided by such a combination of methods are very promising. As will be explained in Chapter 3, I expect to go further in this direction by the combination of computational and experimental methods within integrative structural biology approaches.

## 1.3 Applications in robotics

The methodological developments presented in Section 1.2 can be applied to solve several types of problems related to robotics such as motion planning, manipulation planning and assembly planning. These applications have been briefly mentioned in that section. This section focuses on two applications that I have further investigated, mainly in the context of European projects.

### 1.3.1 Manipulation and Grasping

Since the origins of robotics, manipulation is probably the topic that has attracted more research attention. Indeed, object manipulation is the most important action a robot can perform. Object manipulation with a robot involves planning at several levels: First, given the models of their robot gripper and of the object to be manipulated, possible grasping configurations have to be identified. Then, the manipulation task itself has to be planned, possibly requiring decomposition into subtasks due to object re-grasping requirements. Finally, motions to achieve all these subtasks (e.g. go to grasp, transport, reconfigure for re-grasping, etc) have to be computed before execution. The overall problem is even more complex when several manipulators are involved in a manipulation task. In this case, additional sub-problems, such as coordinated manipulation planning, have to be solved.

Manipulation has been the main focus of my work applied to robotics since my arrival at LAAS-CNRS. During my thesis, I worked on coordinated manipulation planning [45, 23], which requires closed-chain motion planning, and on a continuous instance of the manipulation task planning problem, simultaneously involving task decomposition and motion planning [44, 24, 20]. In the following years, I have continued working on algorithms for these applications, aiming to solve more complex problems, or to improve the capacity of the methods to provide good-quality solutions.

A method based on the composition of a “meta”-graph from a set of probabilistic roadmaps constructed for parts of the system [35] enables the solution of constrained motion planning problems for multi-arm robot systems in very short computing times (between one and two orders of magnitude faster) compared with a standard approach. Other improvements were aimed to enhance the capacities of algorithms for coordinated manipulation planning for such type of robot systems. The multi-layer approach developed during the PhD thesis of Mokhtar Gharbi [39] enables the reconfiguration of the manipulators during the execution of the manipulation task, which increases the volume of the workspace that can be reached. Also related to manipulation with multi-arm systems, a method that extends previous work on pick and place planning was developed for this type of robot systems [34]. This method is able to automatically identify and plan for re-grasping operations involving two hands. when required for solving some manipulation queries. In addition, we have developed an original approach to identify suitable multi-fingered arm configurations to grasp objects defined as continuous surfaces [32], which

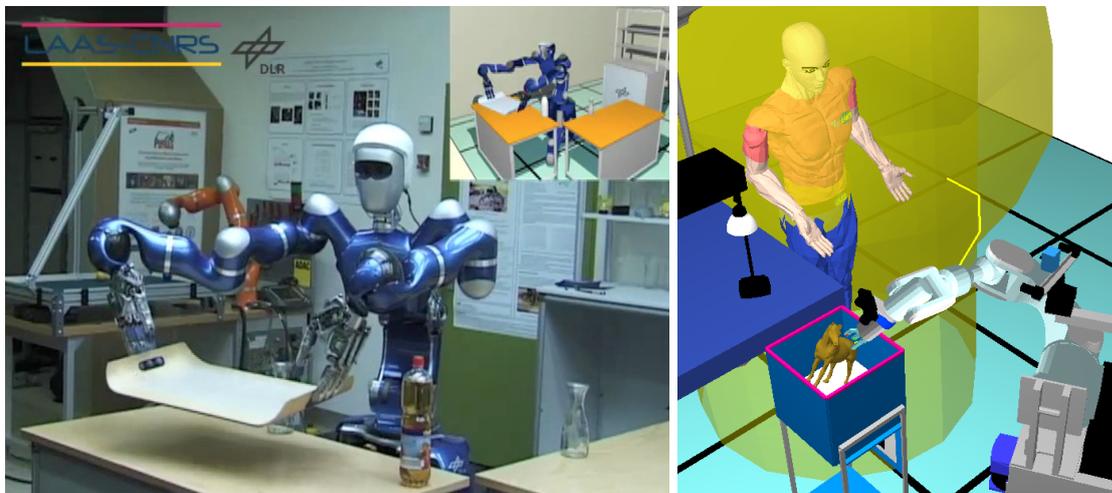


Figure 1.5: *Left:* Two-arm manipulation with the DLR's Justin robot.  
*Right:* Mobile manipulator interacting with a human.

can be applied in conjunction with the aforementioned algorithms. These methods were implemented and successfully tested on real systems in the framework of the European project DEXMART, as illustrated in Figure 1.5.

A current trend in the manufacturing industry is to bring robots out of their cages. This involves the development of a new generation of manipulators able to safely interact with humans. In addition to the mechatronic design and to control methods, the motion planning problem also needs to be revisited. Indeed, motion planning in this context is not only aimed at computing paths that minimize execution time, as is usually the case for industrial robots. In addition, and most importantly, motions must be human-aware. This means that the algorithms need to take into account the presence of humans in the robot workspace in order to compute safe motions, whose intention should be easily interpreted by the human co-worker. Such criteria can be considered to devise a cost function associated with the robot configurations, and then the T-RRT algorithm presented in Section 1.2 can be used to compute good-quality motions with respect to it [33]. For instance, in the example presented in the right-hand image in Figure 1.5, the robot will not execute a direct, straight-line motion to bring the object to the human; instead, it will approach the object following a path that the human can easily see and interpret. We have investigated this type of approach in the context of the European project PHRIENDS, and work in this direction is being continued in the framework of a new project: SAPHARI.

### 1.3.2 Motion planning for aerial robots

Aerial robots are challenging systems for roboticists, and are interesting for diverse applications (in addition to military applications) such as transport and rescue operations, or inspection and supervision in industrial, agricultural and forest environments. Motion planning for aerial robots involves additional difficulties compared to other classes of robots such as simple wheeled robots or manipulators. The main issues are related to controllability. Whereas a purely geometric approach is suitable for planning motions of other types of systems that are more easily controllable, aerial robots require to consider dynamics in order to ensure flyability, and to satisfy constraints such as maximum velocity and acceleration along a planned trajectory. In addition, planned motions should be robust with respect to uncertainties in the position and orientation of the robot, which are the effect of control and sensing errors. These errors can be particularly significant in outdoor environments (especially under variable wind conditions).

We are currently developing motion planning methods for aerial robots in the framework of the European project ARCAS. In this context, we investigate the use of simplified dynamic models of quadrotors and helicopters for motion planning. The idea is to compute realistic and easily controllable motions, while avoiding the use of computationally expensive kino-dynamic planning algorithms. For this, we are developing a local method to connect nearby configuration sampled by the planner, in a similar way to sampling-based motion planning methods for non-holonomic car-like robots. The method (a publication is in preparation), based on polynomial interpolation, guarantees position, velocity and acceleration continuity along the trajectory, and is aimed to minimize flying time, while respecting physical limits of the system. First results obtained (in simulation) with a preliminary implementation of the method are very encouraging.

We are also investigating the application of the T-RRT algorithm for the computation of good-quality paths with respect to safety and controllability metrics in this context. First results, considering a cost function that takes into account the distance between the robot and the obstacles, show the ability of the method to compute high-clearance solution paths in cluttered environments [27, 25]. The Multi-T-RRT algorithm [25] is particularly efficient when solving problems in large-scale workspaces. This is of interest for industrial inspection applications. Given a set of points to be inspected, the algorithm is able to find high-clearance paths to move between them, and automatically determines the passage order that optimizes a criterion defined by the operator, which can be the length of the overall solution path or the integral of a cost function along the path, for instance. Even in very large and geometrically complex workspaces, the algorithm is able to provide good-quality solutions within a few seconds of CPU time.

In addition to the aforementioned works on motion planning, we have developed an original approach to 6-dimensional (6-D) manipulation of objects using flying robots and cables [21]. Aerial towed-cable systems have been mostly applied to transportation, only monitoring the position of the carried load. However, little work has been done on trying to govern the load in both position and orientation, and existing approaches require a

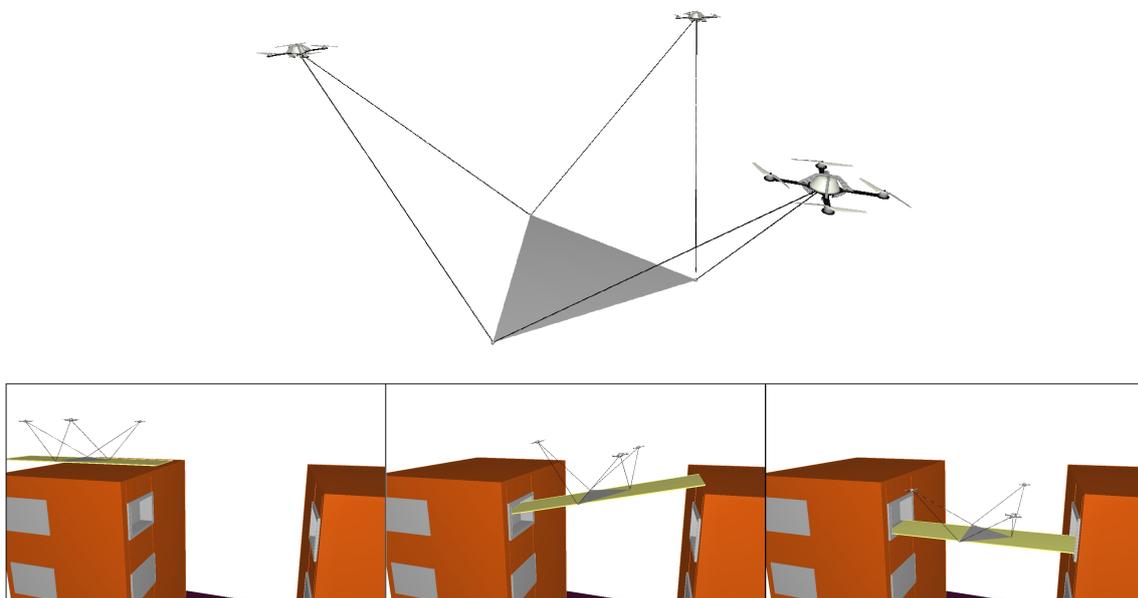


Figure 1.6: *Top*: A version of the FlyCrane. *Bottom*: Example of possible application of the system to install a platform in a rescue operation.

given sequence of load poses along the trajectory, which may be too restrictive, especially in constrained workspaces. In contrast, the approach we have developed for 6-D quasi-static manipulation with an aerial towed-cable system only requires a start and a goal configurations and provides a feasible path to achieve the desired manipulation task. In addition to being feasible, all intermediate configurations along the path fulfill adequate physical properties related to the forces applied to the system and to the cable tensions. Such a quality associated with configurations is measured by a formal criterion derived from the static analysis of the system, based on a similar formulation as that used for cable-driven manipulators. Given this quality measure, a cost-based motion planner such as T-RRT can be applied to compute good-quality paths. In addition to the methodology, we have devised a system to perform 6- D manipulation tasks, that we have called FlyCrane. This system consists of a moving platform attached to three flying robots by means of six fixed-length cables linked by pairs to each robot. An version of this system is illustrated in Figure 1.6. The approach has been validated in simulation on several 6-dimensional quasi-static manipulation problems.

## 1.4 Applications in structural biology, biotechnology and materials science

### 1.4.1 Modeling conformational transitions of proteins and peptides

*Protein loop and domain motions:* Proteins are long polypeptide chains that use to fold (at least partially) into a relatively stable, functional shape. Nevertheless, proteins are not completely rigid molecules. Furthermore, protein flexibility and conformational transitions are generally related to their capacity to recognize and interact with other molecules. The study of protein motions is therefore essential for the understanding of protein functions. Unfortunately, it is very difficult to obtain this type of dynamic information at the atomic scale using available experimental techniques. Modeling protein conformational transitions with conventional computational methods is also challenging because, in many cases, these transitions are rare, slow events. Standard molecular dynamics (MD) simulations with current computational resources cannot be applied in practice to model large-amplitude (slow time-scale) conformational transitions. Such simulations require variants of MD methods that enhance sampling of rare events or that bias the exploration in a given direction, or, alternatively, to have access to outstanding computational power. Modeling conformational transitions in proteins has motivated the development of specific methods, computationally more efficient than MD simulations. Many of these methods are based on the deformation of a trivial initial path between the two given conformations toward the minimum energy path connecting them. These methods use to suffer from local minima problems, which makes them unreliable when applied to complex systems.

As described in Section 1.2, we have developed methods to compute motions of complex systems that can be applied to simulate molecular motion [4]. We have particularly investigated their application to model conformational transitions in proteins, both at a local level (e.g. loop motions) [19, 18], and at a global level [17, 1] (Figure 1.7). These robotics-inspired methods are aimed at providing qualitative information about the conformational transition using few computational resources. For this, they exploit the efficiency of sampling-based exploration algorithms applied to simplified molecular models. Despite the simplifications, the results provided by this type of methods can be very useful to help undemanding molecular mechanisms at the atomic scale. They can also be used as input for other classes of methods to perform a more accurate analysis of conformational transitions.

The first application of an RRT-based algorithm for computing protein loop motions was described in our seminal work in this domain [19]. Results obtained with this initial work served to demonstrate the effectiveness of motion-planning-based methods for studying the mobility of long protein loops. An improved version of the method, which integrates ideas of ML-RRT, was applied subsequent to investigate loop and domain motions in several enzymes in collaborations with researchers at the *Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés* (LISBP, UMR CNRS-INRA-INSA, Toulouse) and

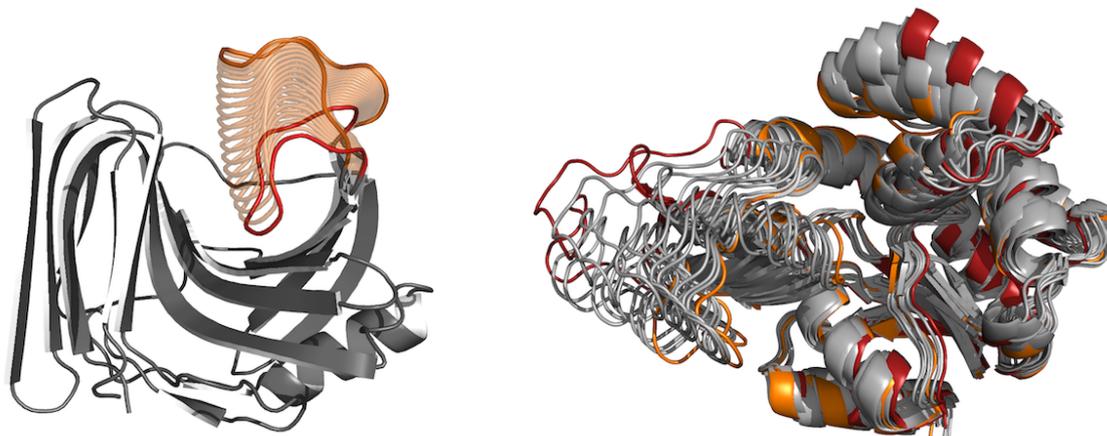


Figure 1.7: Cartoon representations of local and global conformational transitions in proteins.

the *Unité Fonctionnalité et Ingénierie des Protéines* (UFIP, CNRS-Université de Nantes). Some of our joint works [56, 9] were aimed to validate the computational approach, and showed that conformational transition paths computed with robotics-inspired methods are globally similar to those obtained with molecular dynamics simulations, which are orders of magnitude more computationally expensive. These results highlight the ability of our robotics approach to provide satisfying estimations of molecular motions at a very low computational cost. Most importantly, applied together with site-directed mutagenesis experiments, our methods have helped to understand the functional role of a protein loop motion [5].

We have also applied RRT-based algorithms to analyze protein domain motions. In particular, we investigated conformational changes of a DNA binding protein in collaboration with researchers at the *Institut de Pharmacologie et de Biologie Structurale* (IPBS, UMR CNRS-UPS, Toulouse). Results provided by a combination of robotics-inspired algorithms and molecular modeling techniques helped to reinforce the hypothesis about the existence of a continuous running from free to pre-bound conformations of the protein, which are recognized by regulatory DNA regions [10]. We are currently applying the most recent variant of our algorithm [1] to investigate other functional motions of proteins in collaboration with researchers in structural biology at IPBS.

*Peptide conformational transitions:* From a purely chemical point of view, peptides can be seen as small proteins. Indeed, proteins and peptides are polymers formed from amino acids. From the structural point of view, however, a significant difference in addition to size is that, unlike proteins, peptides do not fold into a stable shape. They use to fluctuate between multiple energy minima that correspond to very different conformations. Although peptides may be of less biological significance compared to proteins, they also have important roles in biology, and are of interest for applications in pharmacology and biotechnologies.

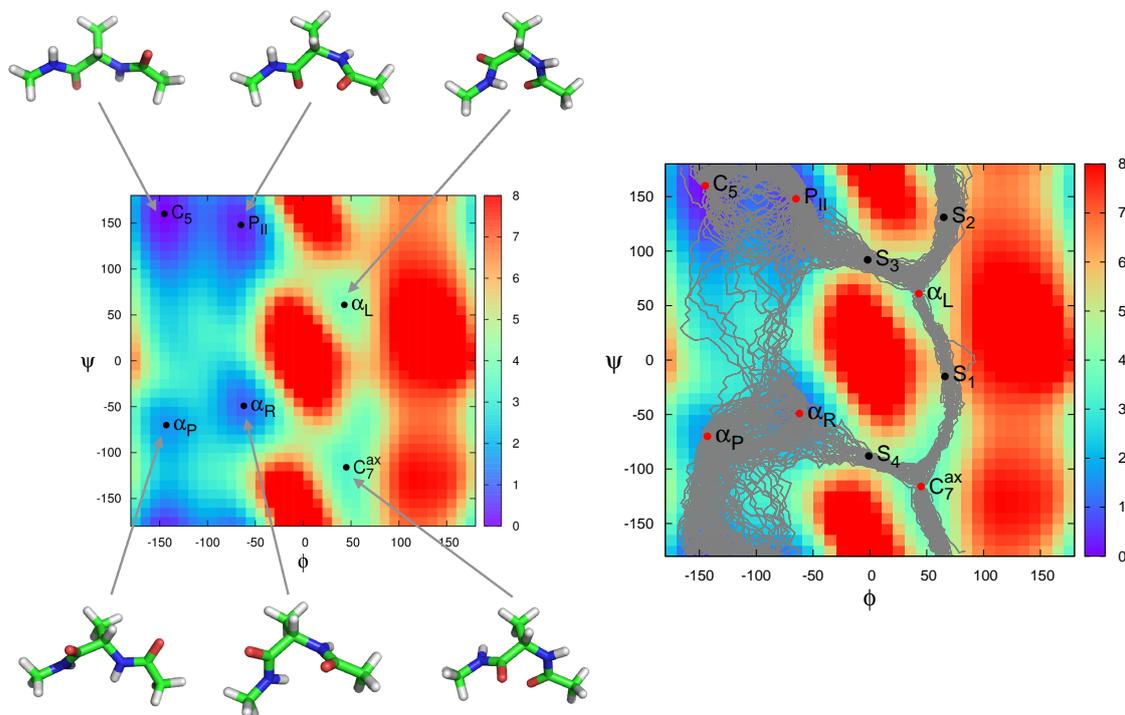


Figure 1.8: Results provided by T-RRT on the energy landscape exploration of a small peptide. *Left*: Identification of the 6 local minima. *Right*: Ensemble of conformational transition paths between the 6 minima, through 4 transition states.

Because of the intrinsic flexibility, the structural characterization of peptides with experimental methods remains extremely difficult. Several computation techniques, mainly derived from molecular dynamics simulation and Monte Carlo methods, have been developed over the years to determine low energy conformations of peptides and the transitions between them. Nevertheless, none of these techniques is fully satisfactory.

In the framework of a collaboration with researchers at the *Laboratori d'Enginyeria Molecular* (UPC, Barcelona) we have shown that the T-RRT algorithm is an interesting method to explore the conformational energy landscape of small peptides [7]. Figure 1.8 presents some results that show the ability of the method to find the local energy minima and a variety of low-energy paths connecting them, which also enables the identification of transition states. Encouraged by these first results, we have continued working on improvements of the method [26], and we envisage its combination with other computational methods aiming to provide a meaningful representation of the conformational energy landscape, which would help to understand structural and dynamic properties of these molecules.

## 1.4.2 Analyzing and predicting molecular interactions

*Protein-ligand interactions:* A significant part of my work in molecular modeling and simulation has been focused on the analysis of protein-ligand interactions. In particular, I have studied interactions taking place along the access/exit pathway between the surface of the protein and the active site, which can be deep inside the protein, especially in the case of enzymes. The activity and the specificity of a protein can be conditioned by these interaction, and not only by the interaction with the ligand bound in the active site. Despite their importance, protein-ligand interactions along the access/exit pathway have been rarely studied, mainly because of the difficulty to obtain accurate information about them with current experiential methods. Computational methods may help to better understand such interactions. However, simulating ligand (un)binding, particularly when the active site is deeply buried into the protein, is also a challenging problem for computational approaches (Figure 1.9 illustrates the problem).

We have shown that the application of motion planning algorithms using a mechanistic representation of molecules is an interesting approach that enables the simulation of protein-ligand (un)binding at a very low computational cost [18, 12]. Such a robotics inspired approach can directly provide qualitative information about protein-ligand interactions taking place far away from the active site, which is very useful for applications in biotechnology and pharmacology as explained below. When more accurate information is required, the solutions provided by these simplified simulations are a suitable first approximation that can be further refined using state-of-the-art energy models and molecular modeling techniques.

The majority of my research on protein ligand-interactions has been conducted in collaboration with researches at the *Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés* (LISBP, UMR CNRS-INRA-INSA, Toulouse), in the framework of several interdisciplinary research projects, of which ALMA (ITAV) and the GlucoDesign (ANR) are the most representative ones (see Section 2.3 for the list of collaborative projects). Together, we have investigated the application of the aforementioned method to rational enzyme engineering [15, 14]. Experimental results have shown the ability of our methods to identify interesting positions for site-directed mutagenesis, leading to new enzymes with improved activity and enhanced selectivity. Such an interesting research direction will be continued in the framework of the ProtiCAD project (ANR), whose objective is the development of novel methods and computational tools for protein design (see Chapter 3 for additional explanations).

We are also collaborating with researchers at the drug design department of *SANOFI* (pharmaceutical company). The objective is to develop improved computational methods to estimate the affinity between protein targets and small molecules, which could be promising candidates for drug development. We are investigating a two-stage approach consisting of the application of motion planning algorithms to provide an approximate reaction path that is then used as input for meta-dynamics simulations. First results obtained with this combined approach are very promising. The computing time and the need

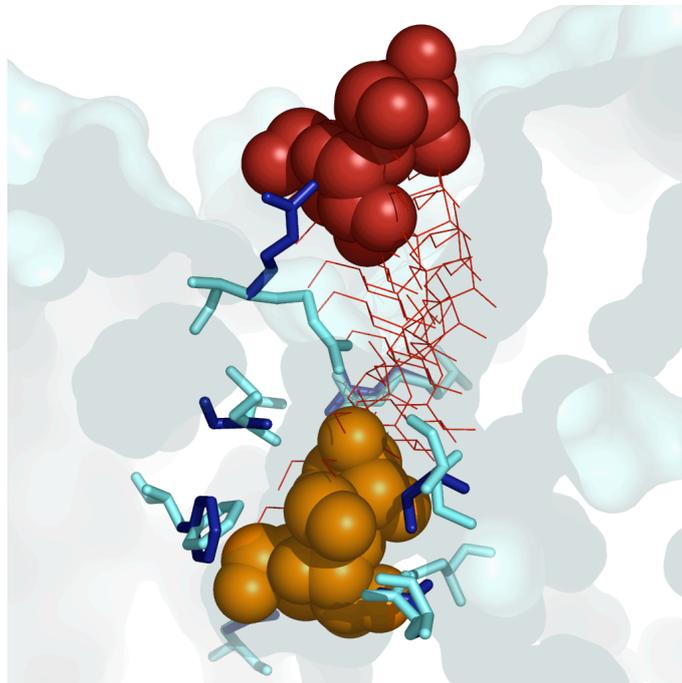


Figure 1.9: Illustration of a protein-ligand (un)binding problem. The figure shows a transversal cut of a protein with a ligand (represented with spherical atoms) occupying different locations: in the active site (orange) and on the surface (red). Some intermediate conformations of the ligand along the exit path are represented with red lines, and some side-chains that change their conformation during the ligand exit are represented with blue sticks.

of user intervention are significantly reduced compared to the previous method applied by the researchers at *SANOFI*. In addition, the method seems to converge more rapidly to an accurate estimation of the free energy variation. Further tests are being conducted to confirm the interest of the approach, which would be the object of a joint publication.

Aiming to enable easy access to our methods for this type of applications to a larger scientific community, we have developed a web application called MoMA-LigPath [2], which is freely available at : <http://moma.laas.fr/> . Starting from the model of a protein-ligand complex, MoMA-LigPath computes the ligand unbinding path from the active site to the surface of the protein. The application provides a set of solution paths, together with information about protein-ligand interactions identified along these paths. In only a few months, more than 300 jobs have been submitted to the web server, and we have received positive feedback, which encourages us to develop other web applications in the future.

*Protein-protein interactions:* In addition to protein-ligand interactions, although in a less extent, we have also addressed the protein-protein interaction prediction problem. Protein-protein interactions are extremely important in biology, since they are at the basis of most molecular processes in the cell. Furthermore, some of these important processes, such as RNA transcription or DNA replication, are carried out by molecular machines that are mainly built from a set of self-assembled proteins. Protein-protein interactions are also important in other domains such as bio-nanotechnology, for the conception of nano-sensors. Many questions remains about the mechanisms of protein recognition and complex formation. This is mainly due to the relatively small number of structures of protein complexes that have been determined experimentally. Computational methods are therefore essential to gain insight into protein-protein interactions. Significant efforts have been done in the last decades to develop methods able to predict the structure of protein complexes given the structure of the individual proteins. We have contributed in this domain with the introduction of the a filtering technique, called RCF (see Section 1.2.2), to improve the performance of a class of protein-ligand interaction prediction methods. Nevertheless, available methods are still inaccurate and unreliable, mainly when dealing with very flexible proteins. I expect to participate in future advances in this application domain, through the development of methods to better deal with molecular flexibility in protein-protein interaction prediction.

### 1.4.3 Polymer modeling

Atomic scale simulation of polymer materials is a topic of interest since it permits to reduce costly experiments to determine their physicochemical properties. In this context, modeling heterogeneously ordered multi-chain systems such as amorphous polymers remains a challenging problem. During my stay in Barcelona in 2008-2009, I initiated a collaboration with Carlos Alemán's group at the *Innovation in Materials and Molecular Engineering Laboratory* (IMEM, UPC), aiming to develop new approaches in this domain.

Our first joint work concerned the development of an improved version of a method to model amorphous polymers recently proposed by Carlos Alemán's group. The method consists of iteratively generating polymer structures using a simplified energy model, and subsequently relaxing the system, considering a more accurate model, in order to reduce its potential energy. We improved this method by integrating a novel relaxation technique, which applies analytical *rebridging* moves inspired by robotics [11]. A comparative analysis using models of amorphous polyethylene with different sizes and densities showed the efficiency of the improved version, which provides a significant performance gain with respect to the original method. We expect to continue this fruitful interdisciplinary collaboration in the framework of future projects.

# Chapter 2

## Other Activities

### 2.1 Teaching and Advising Activities

#### 2.1.1 Teaching

Many of today's scientific and technological advances are the result of interdisciplinary research. Researchers from different domains have to share their knowledge and specific background for working together on challenging problems. This requires some basic knowledge in the other domains, as well as the capacity of communicating with people from other disciplines. Unfortunately, few students hold such an interdisciplinary profile, particularly in the context of my research activities. For this reason, I am devoting more and more time to teaching, trying to promote interdisciplinary education in several departments and at different levels. My teaching activities are summarized next:

- I am co-organizer of the international interdisciplinary school *Algorithms in Structural Bioinformatics*, supported by the CNRS and Inria. This one-week school aims to bring together people from different disciplines interested in advanced methods in structural bioinformatics. The specific subjects being addressed vary each year. In 2012, the first edition of the school was focused on computational methods to model/predict interactions between biomolecules (<http://www-sop.inria.fr/manifestations/algoSB/>). In 2013, the focus was on protein flexibility (<http://algosb.sciencesconf.org/>).
- I am member of the pedagogical team of the Master School *Ingénierie de la Matière: Modélisation des Processus Physiques* (IM2P2) at the University of Toulouse since 2008. Since 2009, I also give a 10-hour lecture entitled “Molecular robotics” on robotics-inspired methods for molecular modeling.
- In 2011, I was invited to give a 15-hour lecture on “Robotics algorithms for molecular modeling and simulation” at the Master School *Simulación de Polímeros y Biopolímeros* of the Universidad Politècnica de Catalunya (Barcelona, Spain).
- In 2008 and 2009, I gave a short lecture (3 hours) on the application of motion planning

algorithms to solve some problems in structural biology at the Master School *Systèmes Automatiques, Informatiques et Décisionnels* (M2R SAID), University of Toulouse.

- In 2008, I gave a 10-hour lecture series on “Motion Planning Algorithms” in the Master School *Robotica* of the Universidad Politécnica de Valencia (Spain). A part of this lecture was dedicated to structural bioinformatics.
- In 2007, I gave a 4-hour lecture at the French-Mexican *Summer School on Image & Robotics* on sampling-based motion planning algorithms and their applications in structural bioinformatics.
- I have also taken some actions to introduce interdisciplinary education at the undergraduate level. In particular, I propose project subjects on molecular modeling for 4-year and 5-year students in the department *Génie Mathématique et Modélisation* of INSA-Toulouse. The possibility of creating a “*Module d’ouverture*” in this interdisciplinary topic is under study.

### 2.1.2 Advising

- Advisor or co-advisor of 8 PhD students:
  - Duc Thanh Le. PhD thesis defended in 2010.
    - \* Subject: (Dis)assembly path planning for complex objects in robotics and biology.
    - \* Main advisor, co-directed with Thierry Siméon (LAAS-CNRS).
  - Mokhtar Gharbi. PhD thesis defended in 2010.
    - \* Subject: Motion and manipulation planning for multi-arm robot systems.
    - \* Co-advisor, co-directed with Thierry Siméon (LAAS-CNRS).
  - Ibrahim Al-Bluwi. PhD thesis defended in 2012.
    - \* Subject: Robotics-inspired methods to model conformational changes in proteins.
    - \* Main advisor, co-directed with Thierry Siméon (LAAS-CNRS).
  - Romain Iehl. Last-year PhD student.
    - \* Subject: Path-planning-based algorithms applied to the rational design of enzymes.
    - \* Co-advisor, co-directed with Thierry Siméon (LAAS-CNRS).
  - Didier Devaurs. Third-year PhD student.
    - \* Subject: Algorithms to explore high-dimensional (continuous) cost-spaces.
    - \* Main advisor, co-directed with Thierry Siméon (LAAS-CNRS).
  - Victor Gil. Third-year PhD student.
    - \* Subject: Coupling physical-chemistry and robotics approaches to model protein flexibility.
    - \* Co-advisor, directed by Victor Guallar (BSC, Barcelona).
  - Alexandre Boeuf. Second-year PhD student.
    - \* Subject: Motion planning for aerial manipulators.
    - \* Co-advisor, co-directed with Thierry Siméon (LAAS-CNRS).
  - Laurent Denarie. First-year PhD student.
    - \* Subject: Robotics-inspired methods for computational protein design.
    - \* Main advisor, co-directed with Thierry Siméon (LAAS-CNRS).

- Advisor or co-advisor of 6 post-docs:
  - Svetlana Kirillova, from the National Academy of Sciences of Belarus. 2005-2006.  
Subject: Development of a novel method combining Normal Mode Analysis and path planning algorithms to compute conformational transitions of proteins.
  - Léonard Jaillet, former PhD student at LAAS-CNRS. 2007-2009.  
Subject: Motion planning algorithms with applications to structural biology.
  - Sophie Barbe, from the ENS Cachan. 2006-2009.  
Subject: Interfacing robotics-inspired algorithms and molecular modeling techniques.
  - Yi Li, from Simon Fraser University. 2008-2009.  
Subject: A geometric filtering technique to enhance protein-protein docking methods.
  - Montserrat Manubens, visiting postdoc from IRI (CSIC-UPC, Barcelone). 2012.  
Subject: Motion planning methods for manipulation with aerial robots and cables.
- Advisor of 12 undergraduate students at different levels, from 2<sup>nd</sup> year university degree to Master. I have proposed and supervised small projects (from 2 to 6 months) on different topics related to robot motion planning and molecular modeling.
- Scientific advisor of a CNRS Research Engineer: Marc Vaisset.

## 2.2 Organizational and Editorial Activities

### 2.2.1 Chairing activities, organization de conferences et workshops

- Since 2009, I am co-chair of the l'IEEE RAS Technical Committee on Algorithms for Planning and Control of Robot Motion (<http://www.robotmotion.org/>). The goals of this international committee are: to be a reference point for researchers on motion planning and control in robotics (and other domains), to coordinate and to support the organization of workshops and tutorials, and to promote scientific exchanges in this community.
- Co-organizer of the Workshop on “Artificial Intelligence and Robotics Methods in Computational Biology”, during the international conference of the Association for the Advancement of Artificial Intelligence (AAAI), 2013.
- Co-organizer of the Tutorial “From Robot Motion Planning to Modeling Structures and Motions of Biological Molecules”, at the ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB), 2013.
- Co-organizer of the Workshop “Motion Planning: From Theory to Practice”, during the international conference Robotics Science and Systems (RSS), 2010.
- Program Committee member of the Computational Structural Bioinformatics Workshop (CSBW), 2011-2013.
- Program Committee member of the European Conference on Artificial Intelligence (ECAI), 2008, 2010.

### 2.2.2 Editorial activities

- *Associate Editor* of the IEEE Robotics and Automation Society conferences: ICRA 2009-2012 ; IROS 2010-2013.
- *Reviewer* of journals and conferences of robotics: IEEE Transactions on Robotics, International Journal of Robotics Research, IEEE Transactions on Automation Science and Engineering, Workshop on the Algorithmic Foundations of Robotics, ...
- *Reviewer* of journals of bioinformatics, computational biology and physical chemistry: Journal of Chemical Theory and Computation, Bioinformatics, Archives of Chemistry and Biophysics, Journal of Computational Chemistry, ...

### 2.2.3 Other organizational and evaluation duties

- Member of the *Jury de recrutement des CR2 d'Inria Sophia Antipolis*, 2013.
- Since 2011, member of the *Commission de recrutement du Département Génie Electrique et Informatique* (DGEI) of INSA-Toulouse.
- Between 2007 and 2009, member of the *Commission de Projets LAAS*, in charge of the evaluation of interdisciplinary projects between groups of the laboratory.

## 2.3 Projects and collaborations

### 2.3.1 Contractual projects

- ProtiCAD - National Project ANR *Modèles Numériques* (2013-2016)
  - *Subject*: Methodological advances and novel tools for computational protein design.
  - *Partners*: LAAS-CNRS, BIOS-Ecole Polytechnique, LISBP-INSA, Kineo-Siemens.
  - *Responsibilities*: Coordinator.
- ARCAS - European Project FP7-ICT (2011-2015)
  - *Subject*: A first multi-robot aerial system for the assembly and construction of structures.
  - *Partners*: CATEC (ES), DLR (DE), UNINA (IT), LAAS-CNRS (FR), USE (ES), UPC (ES), STI (DE), AIR (CH).
  - *Responsibilities*: Co-leader of one work-package.
- SAPHARI - European Project FP7-ICT (2011-2015)
  - *Subject*: New paradigms for the design of robots interacting with humans.
  - *Partners*: UNIHB (DE), UNIPI (IT), UNINA (IT), TUM (DE), LAAS-CNRS (FR), IOSB (DE), IIT (IT), EADS (FR), DLR (DE), UNIROMA1 (IT), KUKA (DE).
- OTIMASU - LAAS-CNRS Project (2009-2012)
  - *Subject*: Methods to model interactions between biological molecules and inorganic surfaces.
  - *Partners*: RIS and N2IS groups of LAAS-CNRS.
  - *Responsibilities*: Coordinator.

- GlucoDesign - National Project ANR *Biologie-Santé* (2009-2012)
  - *Subject*: Low-cost chemo-enzymatic routes to produce glycovaccines against *Shigella flexneri*.
  - *Partners*: Institut Pasteur, LISBP-INSA, LAAS-CNRS.
- DEXMART - European Project FP7-ICT (2008-2012)
  - *Subject*: Service robots with dexterous and autonomous dual-hand manipulation capabilities.
  - *Partners*: UNINA (IT), LAAS-CNRS (FR), Univ. Bologna (IT), DLR (DE), Univ. Karlsruhe (DE), Univ. Saarlandes (DE), OMG plc (UK).
- AMYLO -Regional Project Midi-Pyrénées (2007-2010)
  - *Subject*: Coupling simulation and haptic devices for modeling functional dynamics of proteins.
  - *Partners*: LAAS-CNRS, LGP-ENIT, LISBP-INSA, Kineo-CAM.
- AMOBIO - Transregional Project *Communauté de Travail des Pyrénées* (2007 - 2009)
  - *Subject*: Motion of structures with kinematic loops in robotics and biochemistry.
  - *Partners*: LAAS-CNRS, UPC-CSIC (ES), Crystax Pharmaceuticals (ES), Kineo-CAM.
  - *Responsibilities*: Coordinator in Midi-Pyrénées.
- PHRIENDS - European Project STReP IST (2006-2009)
  - *Subject*: Developing robots that can co-exist and safely co-operate with humans.
  - *Partners*: Univ. Pisa (IT), DLR (DE), LAAS-CNRS (FR), KUKA (DE), UNIROMA1 (IT), UNINA (IT).
- ALMA - Regional Project ITAV (2006 - 2009)
  - *Subject*: New approaches, combining robotics methods and molecular modeling techniques for *in silico* analysis and prediction of interactions between biological molecules.
  - *Partners*: LAAS-CNRS, LISBP-INSA, IPBS-CNRS, Inst. Claudius Regaud-INSERM.
- NanoBioMod - National Project ANR *Jeunes Chercheurs* (2006-2008)
  - *Subject*: Multi-model approaches for the simulation of molecular systems in biotechnology.
  - *Partners*: LAAS-CNRS, IPBS-CNRS, Rutgers Univ. (USA).
- Xylanase - National Project ADEME 02-01051 (2004-2006)
  - *Subject*: Optimizing the production process of bio-fuel from cereal co-products using the catalytic properties of a xylanase.
  - *Partners*: UFIP (Nantes), INRA (Reims).
- AMORO - LAAS-CNRS Project (2003-2005)
  - *Subject*: Developing new methodologies for modeling and predicting protein-ligand and protein-protein interactions.
  - *Partners*: MIS and RIA groups of LAAS-CNRS.
- IRASIS - National Project Robea (2003-2005)
  - *Subject*: Developing a robot system for tumor destruction using an image-guided percutaneous procedure.
  - *Partners*: LSIIT (Strasbourg), IRCAD (Strasbourg), LAAS-CNRS.

- MOVIE - European Project IST-2001-39250 (2003-2005)
  - *Subject*: Development of techniques able to compute realistic motions of several virtual entities in complex environments.
  - *Partners*: Universiteit Utrecht (NL), LAAS-CNRS (FR), Tel Aviv University (IL), Kineo-CAM (FR).
- BioMove3D - PIR CNRS (2002-2004)
  - *Subject*: Development of methods to simulate protein-ligand access/exit paths based on recent robot motion planning algorithms.
  - *Partners*: LAAS-CNRS, INSA (Toulouse), INRA (Nantes).
- MOLOG - European Project Esprit 28226 (2000-2002)
  - *Subject*: Extend the range of CAD systems applications to help the operator in decision making via the integration of new techniques of geometric reasoning, including motion and task planning.
  - *Partners*: LAAS-CNRS (FR), Universiteit Utrecht (NL), CADCENTRE (UK), EDF-DER (FR), Kineo-CAM (FR).

### 2.3.2 Other collaborations

- Collaborations with several members of the **Institut de Robòtica i Informàtica Industrial** (IRI, CSIC-UPC, Barcelona), before, during, and after my stay in this laboratory in 2008-2009. We have collaborated on the development of improved motion planning algorithms for closed-chain mechanism, and on distance computation and collision detection methods for molecular models. In addition, the mobility of people between the two laboratories is very active. Léonard Jaillet moved to the IRI as a postdoc after his PhD at LAAS. Montserrat Manubens, postdoc at IRI, did a 8-month stay at LAAS in 2012. I was also external participant in the Spanish National project CUIK++, coordinated by Lluís Ros, between 2011 and 2013.
- Collaboration with Vinh Tran, molecular modeling expert at the **Unité Fonctionnalité et Ingénierie des Protéines** (UFIP, CNRS-Université de Nantes). After the end of the ADEME project between 2004 and 2006, which funded joint research, we have continued an informal collaboration on the combination of robotics-inspired algorithms and molecular modeling techniques for understanding functional dynamics of proteins. With the recent arrival of Yves-Henri Sanejouand (expert on methods to model the flexibility of proteins) at the UFIP, we expect to extend and reinforce our collaboration.
- Collaboration with the group of Carlos Alemán at the **Innovation in Materials and Molecular Engineering Laboratory** (IMEM, UPC, Barcelona) on the development of polymer modeling methods. I have co-advised the work of a master student (Sergio Carrión) and a PhD student (Esther Córdoba) in this group. We are currently planning to extend the collaboration with the participation of Ian W. Hamley (Dept. of Chemistry, University of Reading) for treating new problems involving modeling of polymer-peptide conjugates. The elaboration of a European project is under discussion.

- Collaboration with Marc Bianciotto and Jean-Philippe Rameau of the *Drug Design* team of **SANOFI** R&D. We develop improved computational methods to estimate the affinity between druggable compounds and protein targets. We have co-advised a master student in 2013, Fabien Contaut, who has worked on this topic. First promising results have been obtained, and more experiments are being conducted in order to collect data for a joint publication. We expect to continue this fruitful collaboration in the near future.
- Informal collaboration with Victor Guallar, leader of the Electronic and Atomic Protein Modelling Group at the **Barcelona Supercomputing Center** (BSC, UPC, Barcelona). We exchange ideas on the combination of methods for modeling protein flexibility, and we promote the mobility of students between our groups. I co-advise the work of Victor Gil, PhD students at the BSC. In 2011, he did a one-month stay at LAAS. Didier Devaurs, PhD student at LAAS, visited the BSC during two months in 2011.
- Project of collaboration with Lionel Perrin, researcher at the **Laboratoire de Physique et Chimie de Nano-Objets** (LPCNO, UMR INSA-CNRS-UPS, Toulouse), on peptide self-assembly modeling. We would like to co-advise a PhD student for working on a multi-scale multi-physics approach. Unfortunately, funding applications to the University of Toulouse have been unsuccessful. We are looking for alternatives.
- Project of collaboration with Pau Bernadó, researcher at the **Centre de Biologie Structurale** (CBS, UMR CNRS-INSERM-UM1, Montpellier), on modeling of intrinsically disordered proteins. Our first attempt to get funding through a proposal of *Projets Exploratoires Pluridisciplinaires* (PEPS) in 2012 was unsuccessful. Nevertheless, we think that the approach we have in mind, combining experimental and computational methods, would provide very interesting results. We are planning to propose a larger project on this topic in the short future.
- Project of collaboration with the groups of Lionel Mourey et Alain Milon at the **Institut de Pharmacologie et de Biologie Structurale** (IPBS, UMR CNRS-UPS, Toulouse). These groups are experts in experimental structural biology methods, such as X-ray crystallography, small-angle X-ray scattering (SAXS), and nuclear magnetic resonance (NMR). In April 2013, we organized a brainstorming meeting on the combination of experimental and computational method in structural biology. Several joint research directions raised from this meeting. We have initiated informal collaborations on some topics, and we are planning to apply for funding for collaborative projects to develop integrative structural biology approaches coupling experiment and simulation.

## 2.4 Mobility

### 2.4.1 Disciplinary mobility

In 2002, when I was still a PhD student in the robotics group of the LAAS-CNRS, I started to be interested in problems related to protein modeling. This interest has been reinforced along the years. My postdoctoral stay in a laboratory of biochemistry at the Université de Nantes in 2004 was an important stage that strengthened my motivation to carry out interdisciplinary research. Since 2004, when I was recruited as permanent researcher at LAAS-CNRS, a notable part of my research activities is focused on the development of methods to model proteins and their interactions with other molecules. These methods have been applied to several problems in structural biology, biotechnology and materials science in collaboration with experts in these domains. At present, I am envisaging the development of methods for modeling hybrid molecular systems, involving biomolecules and synthetic/inorganic molecules. Such systems have potential applications in bio-nanotechnologies, which is a domain of increasing interest at LAAS-CNRS. This point is further developed in my research project (Chapte 3).

### 2.4.2 Geographic mobility

Between November 2008 and November 2009, I moved to Barcelona (Spain) for a one-year stay as Invited Professor at the *Universidad Politècnica de Catalunya* (UPC), funded by international mobility programs of the CNRS and the INPT. This stay allowed me to reinforce the relationship with colleagues at the *Institut de Robòtica i Informàtica Industrial* (IRI, CSIC-UPC). I also had the opportunity to meet researchers in neighboring laboratories with the aim of establishing future collaborations. In particular, I met Carlos Alemán (Innovation in Materials and Molecular Engineering Laboratory, UPC) and Victor Gual-lar (Barcelona Supercomputing Center), with whom I have very interesting interactions at present.

## 2.5 Software development

### 2.5.1 Move3D

Since my arrival at LAAS-CNRS, I participate in the development of Move3D, a general-purpose motion planning software. This software has been mainly used for the prototyping of algorithms. Modules developed from Move3D libraries are currently embedded in some of the robots at LAAS-CNRS and have been integrated in demonstrators in the framework of several european Projects.

Note that Move3D was at the origin of the creation of the start-up company Kineo-CAM, which commercializes software solutions for CAD/CAM based on motion planning algorithms originating from robotics research.

### 2.5.2 MoMA

Since 2006, I lead the development of a new molecular modeling software platform called MoMA (for Molecular Motion Algorithms). In collaboration with Marc Vaisset (Research Engineer at LAAS-CNRS), we have conceived and developed a modern, modular and evolutive set of libraries for the implementation of software applications dedicated to the simulation of molecular motions and interactions. Two basic modules, PSF and AMC, are already available in the public domain (under CeCill license) : <https://softs.laas.fr/Psf-Amc> . The module called PSF (for Protein Structure Format) is aimed to parse and organize the informations contained in standard files to represent bimolecular structures (i.e. the Protein Data Bank file format). The module AMC (for Articulated Molecular Chain) concerns the mechanistic representation of molecular objects. The core module of MoMA, called AMC-Motion, which contains a collection of algorithms to explore the conformational space of molecular systems is under development. A first applications developed from MoMA, called MoMA-LigPath, is accessible via a web server: <http://moma.laas.fr/>. Other applications will be available soon.

## 2.6 Technology transfer, industrial relationships

- Since its creation in 2000, I have had a close contact with the company **Kineo-CAM**. This LAAS-CNRS spin-off, which was acquired by Siemens in 2012, develops software solutions integrating motion planning technologies into CAD/CAM packages. Although the activities of the company are mainly focused on industrial manufacturing, they aim to spread over other areas. In particular, they are studying the possibility of developing molecular modeling software based on robotics-inspired algorithms. With this aim in mind, they have been partners in some projects coordinated by LAAS-CNRS. Through my regular cooperation with Kineo-CAM, I contribute to technology transfer from academic research to industry.
- In 2009, I was contacted by the company **DNASTAR**. Specialized in the development of software tools for genomics, this USA company wanted to extend its activities toward structural bioinformatics. Being interested in the type of robotics-inspired methods we develop, DNASTAR envisaged either the possibility of establishing a research collaboration with LAAS-CNRS, or proposing me to serve as a consultant. Unfortunately, some administrative constraints and a change in the scientific direction of the company in 2010 made that the collaboration project was not accomplished.

# Chapter 3

## Research Project

### 3.1 Overview

The core of my research will continue focused on the development of algorithms to explore high-dimensional spaces. Despite significant advances in this area during the last decades, a number of problems related to path finding and optimization in (non-convex) high-dimensional spaces remain open. Developing better methods to solve these problems is of interest for robot motion planning and control, as well as in other domains related to the motion of physical or virtual systems, such as structural bioinformatics. For the development of improved algorithms, I will investigate connections with other areas of computer science, especially with other branches of artificial intelligence, and with other theoretical disciplines. Indeed, sampling and exploration algorithms are the basis of numerous methods in diverse disciplines, being applied mathematics and statistical physics the two more clear examples.

In addition to methodological developments, their application in different domains will continue playing an important role in my research. I wish to continue investigating applications in robotics. Even though this is not a central part of my research project, this is important for me in order keep rooted in my basic discipline. In particular, motion planning and control of aerial robots is a challenging area in which I intend to work together with other researchers at LAAS-CNRS.

As during the past ten years, the main motivation for my research will be the study of molecular systems. In the context of structural biology, I am interested in coupling computational methods and experimental methods for a more accurate structural and dynamic characterization of biological molecules, especially of protein. I am also particularly interested in the development of new approaches to computational protein design, taking into account protein flexibility. Going further in this direction, and for a longer-term research, my project is to develop novel algorithmic approaches for the design of systems at the atomic scale, with applications in bio-nanotechnologies.

Next sections provide more details on my research project from a fundamental and an applicative point of view.

## 3.2 Methodological developments

As we have shown in recent years, combining the underlying principles of sampling-based motion planning algorithms and ideas from other methods, such as stochastic optimization methods, is a suitable approach to address challenging problems in high dimension. This is an interesting research direction that I propose to continue.

In the short future, I will continue working on further improvements and extensions of the T-RRT algorithm, as well as on a more rigorous analysis of its properties. T-RRT is a very general algorithm that can be applied to find good-quality paths in robotics applications, or low-energy (highly-probable) conformational transitions paths of biomolecules. Nevertheless, more specialized variants of the algorithm, with specific properties, can be developed depending on the application. For instance, in robotics, an *any-time* variant of the algorithm that will eventually converge to an optimal solution if additional computing time is available for motion planning, would be of particular interest. In molecular applications, an important aim for new algorithmic variants is to solve more efficiently problems in very-high dimension. For this, the development of more sophisticated sampling strategies and suitable search heuristics are required. In addition, I want to further investigate the properties of the solutions provided by the algorithm with respect to thermodynamic properties of molecular systems.

Another interesting direction that I intend to investigate is the introduction of machine learning approaches in motion planning algorithms. Thanks to their conceptual simplicity, sampling-based algorithms are very general techniques, able to solve many classes of problems in moderately-high dimension. However, their simplicity can also be a drawback in some cases; for instance, when the feasible (or low-cost/energy) subsets of the search space have a complicated topology, or when solving problems in very-high dimension. In such cases, the number of samples required to solve the problem can become prohibitively high. A suitable approach to better solve these difficult classes of problems is to learn about the structure of the regions of the space that are incrementally explored, and to use this information to guide subsequent exploration. The choice of the parameters or properties to be learnt and of the learning method itself are not straightforward. Several possibilities have to be studied, taking care of preserving ergodicity, completeness, and convergence properties of the exploration algorithms when introducing heuristics.

The aforementioned research directions concern extensions or improvements of algorithms that explore continuous high-dimensional spaces by sampling points (i.e. states, configurations) in order to construct a graph structure that tends to capture the topology of the subset of the space where solutions can be found. When the problem involves finding a continuous path, algorithms that directly operate in the path-space or the trajectory-space (rather than in the state-space) can be an interesting alternative. In robotics, such methods have been mostly developed for trajectory optimization, and (with few exceptions) they only present local convergence guarantees. More global methods have been developed in computational biology and statistical physics. Nevertheless, their applicability remains

mostly restricted to a particular molecular simulation framework. The development of general, global methods to explore the path/trajectory-space of a mobile system is also a promising research direction that I would like to investigate in the future.

### 3.3 Robotics applications: motion planning and control of aerial robots

During decades, different subproblems related to robot motion have been usually treated separately. Motion planning has traditionally considered only geometric aspects of the problem, disregarding time or controllability issues. Algorithms have been mostly developed to compute collision free paths, considering perfect models of the mobile system and the environment, and neglecting or simplifying dynamic considerations, which are treated subsequently by trajectory generation and control methods. It is worth to mention that the purely geometric problem is difficult enough, mainly when considering mobile systems with many degrees of freedom in cluttered workspaces, so that it has motivated research in the robotics community since the 70's. Furthermore, such a decoupled approach (geometric path planning + trajectory generation + control) is suitable in many applications, especially when the mobile system is easily controllable, which is usually the case for a serial manipulator or a wheeled robot moving at moderate speed. However, systems that are more difficult to control, such as legged robots or flying robots, require a more integrated approach.

I have started to work on more integrated methods for motion planning and control of aerial robots in the framework of the European project ARCAS. The subject is challenging, and I expect to continue working on it after the end of this project. Indeed, aerial robotics is becoming a central research topic in the Robotics and Interactions (RIS) team at LAAS-CNRS, and more generally in the academic and industrial context of Toulouse. In addition, the arrival of a new CNRS researcher from January 2014, Antonio Franchi, expert on control methods for aerial robots, will significantly contribute to reinforce these aspects inside the team. Together with Antonio Franchi, Simon Lacroix and other colleagues, we envision the development of more integrated robot system architectures, involving stronger connections between sensing, planning and control. Problems related to aerial manipulation are a new venue for the development of this type of methodology. Following the lines initiated within the ARCAS project, we will continue working on methods to generate motions of aerial robots equipped with arms or cables for object manipulation.

In addition to aerial robotics, I will continue participating in other activities of the RIS team related to motion planning and task planning, mainly through co-advising of PhD students and postdocs.

### 3.4 Towards integrative approaches in structural biology

A complete description of biological activity at the molecular and atomic levels requires taking into account the role of conformational dynamics. However, no single computational or experimental approach can provide a comprehensive yet detailed characterization of the ensemble of conformations that flexible biological macromolecules, such as proteins and RNA, can adopt. Understanding molecular phenomena through such characterization would be central to novel insights and treatments of numerous diseases related to protein dysfunction. It would also be interesting in other domains such as biotechnology and nanotechnology, where proteins and RNA have numerous applications.

Advances in this area require the development of integrative approaches, combining theory, computation, and experiment to provide an accurate structural and dynamic description of biological macromolecules. Several leading groups in the U.S.A. and in Europe are starting very interesting work in this direction. My project is to participate in the development of an original approach by the introduction of robotics-inspired methods as new computational tools in an integrative structural biology framework. I have started discussions on several ideas with structural biology experts, and an informal collaboration has been initiated with some of them while searching for opportunities to fund our joint research.

The most direct application of robotics-inspired models and algorithms would be the flexible fitting of high-resolution structures into low-resolution data provided by small-angle X-ray scattering (SAXS) or microscopy to produce accurate models of large systems. This is an important open problem in structural biology, which is not well solved with current methods. I expect to work on this in the short future in collaboration with Lionel Mourey (IPBS, Toulouse).

Another even more interesting direction for future research in this area is the combination of motion-planning-based algorithms and statistical physics methods with experimental techniques, such as nuclear magnetic resonance (NMR) and SAXS, to accurately model conformational ensembles corresponding to (meta)stable states, and transitions between them. The idea is to use experimental data to bias and/or to restrain simulations in an intelligent way. Some directions to put this idea into practice have been envisaged in collaboration with Alain Milon (IPBS, Toulouse) and Michael Nilges (Institut Pasteur, Paris).

Following a similar approach on the combination of experimental and computational methods, an extremely challenging problem that I want to address, in collaboration with Pau Bernadó (CBS, Montpellier), is the structural and dynamic characterization of intrinsically disordered proteins (IDPs). Despite lacking permanent secondary and/or tertiary structure, IDPs are fully functional and indeed perform highly specialized biological tasks that complement those done by globular (folded) proteins. Furthermore, numerous IDPs are directly associated with human diseases, including cancer, cardiovascular diseases, and amyloidogenic pathologies such as Alzheimer's, Parkinson's, or diabetes II. IDPs pose an

enormous challenge to structural biology as traditional approaches are not useful in such an extreme case of conformational variability. Structural information can be obtained with NMR and SAXS methods, but the interpretation of the observables measured is not straightforward. New computational methods are necessary for a physically meaningful interpretation in terms of structure and dynamics of experimental data obtained for IDPs and to understand how these proteins work. The complexity of the problem will require improved exploration algorithms, as well as sophisticated parallel implementations for high-performance computing. We have applied for funding for a co-advised PhD student to start working on such a challenging and largely unexplored subject.

### 3.5 Computational Protein Design

Due to their large range of possible functions, the study of proteins interests other fields in addition to biology. For instance, proteins are pharmaceutical targets and drugs, their catalytic properties are widely used in biotechnology, and they are used as components of nano-devices in the rising field of bio-nanotechnology. Although the properties of natural proteins can be directly exploited, new, designed proteins, with novel functions or improved activities, are of major interest in all these application areas.

An important part of my research in the next years will be focused on the development of computational methods and suitable tools to assist the synthesis of proteins with specific properties and new functions. The problem is extremely challenging since the number of possible combinations of amino acids to be tested is astronomically large. Despite impressive advances in recent years, the activity of proteins designed with available methods is usually far below that of natural proteins. The main reason for this is that dynamics aspects are mostly neglected by current design approaches. Indeed, in addition to the intrinsic combinatorial complexity of the protein design problem, computational methods have to face the natural flexibility of proteins, which may play essential roles in their functions. Solving such a challenging problem requires the development of novel approaches, involving the definition of appropriate models and the implementation of efficient search/optimization algorithms, beyond the state-of-the-art.

The ANR project ProtiCAD, of which I am the coordinator, represents a first step in this direction. Started in January 2013 in collaboration with researchers at LISBP-INSA (Toulouse) and BIOS-Polytechnique (Palaiseau), and with the technical support of Kineo-SIEMENS (Toulouse), the project aims to yield advances in computational protein design by a synergistic combination of cutting-edge methods in computational biology with efficient algorithms originating from robotics. The goal of this interdisciplinary approach is to overcome the current limitations of available techniques, in particular regarding protein backbone flexibility and ligand accessibility. The adequacy between models, energy functions, and exploration algorithms is a key issue for successful protein design that we are addressing in this project.

Among all the possible applications of the methods developed in this project, special attention will be given to enzyme design for applications in biotechnology such as the production of high-valued molecules, the development of eco-friendly bioprocesses and the valorization of renewable carbon resources. Such applications are of high interest to the pre-industrial demonstrator Toulouse White Biotech (TWB) Center and to the Competitiveness Cluster AgriMip, which are supporters of our project.

Computational protein design is a fascinating topic that I expect to continue investigating beyond the end of the ProtiCAD project. In addition to the theoretical challenge, I am particularly motivated by the development of methods to assist the conception of new enzymes used in biotechnologies, which would contribute to the rise of an innovative bio-economy, respectful of the environment. The local context in Toulouse, mainly through collaborations with the LISBP-INSA and with the support of the TWB Center, is very favorable for this.

### 3.6 Towards atomic-scale CAD

Research in bio-nanotechnology advances very rapidly. Ground-breaking discoveries in this area happen almost everyday, offering great opportunities for new technological developments. Indeed, bio-nanotechnology is leading to totally new classes of devices and systems, and is providing new insights into the understanding of biological systems. To date, advances in bio-nanotechnology have been mostly led by experimental research, generally based on expensive trial-and-error procedures. The main reason is that current theoretical methods and computational tools for treating the complex problems that arise in this area are very limited. Like in today's industrial manufacturing, where CAD/CAM techniques drastically reduce cost and time for the development of new products and facilitate product life-cycle management, the use of appropriate computational tools will completely change the current procedures in bio-nanotechnology. The revolution can be even more significant than the one represented by the introduction of digital mockups in the automotive industry during the 80s.

The development of CAD methods and tools for atomic-scale design of bio-nano materials and systems is an exciting direction for long-term research. This is a big challenge that will require a strong collaboration between theoreticians, experimentalists and technologists. First, the development of fundamental methods beyond the state-of-the-art for modeling, simulating and characterizing complex molecular systems will require a narrow collaboration with physicists and theoretical chemists. Besides, computational methods will often be used in tandem with experimental methods, and experiments will be necessary to validate them. Finally, technologists will define specific needs and functionalities to facilitate decision making for molecular engineering. Building on a fruitful experience over the past ten years, during which I have been at the core of interdisciplinary research projects in computational structural biology, and extremely motivated by the theoretical and technological problems to be addressed, I feel able to conduct such a challenging exploratory project.

In the short-medium term, my research project will be focused on the development of fundamental methods in this context, and on their application to particular systems. The main topics that I plan to address in the coming years are:

- **The analysis of interactions between biomolecules and surfaces:**

Developments in bio-nanotechnology usually involve interactions between biomolecules and synthetic/inorganic surfaces. For instance, biomolecules are used to functionalize material surfaces and nano-particles, to improve the biocompatibility of medical implants, or as components of electronic nano-devices. Better understanding of the basic underlying mechanisms of these interactions, through the combination of experimental and computational approaches, would greatly contribute to significant advances in all these application domains. I wish to investigate the applicability of robotics-inspired algorithms in this context. A first work in this direction will be conducted in collaboration with Christian Schön (Max Planck Institute for Solid State Research, Stuttgart). A two-month visit of Christian Schön has already been scheduled for 2014. In addition, this is a topic of interest for some colleagues at the N2IS team of LAAS-CNRS, which would enable internal collaborations between the two teams.

- **The design of peptide-based nano-structured materials:**

The natural ability of biomolecules that self-organize to form more complex structures has inspired in the last few years the construction of peptide-based nano-structured materials such as nanotubes, nanowires and monolayers. Although the very intense research activity on peptide-based materials is leading to a rapid advancement of the knowledge in this area, the mechanisms of peptide self-assembly are, in general, not well understood. New methods are therefore necessary in order to complement available experimental and computational approaches for the design and characterization of peptide-based materials. I have started discussions with Lionel Perrin (LPCNO-INSA, Toulouse) on a multi-scale, multi-physics approach, coupling quantum physics calculations, mechanistic models, and motion-planning-based algorithms, to simulate peptide self-assembly. We are looking for opportunities to fund a PhD thesis on this subject.

- **The study of amphiphilic peptide-polymer conjugates:**

Amphiphilic molecules, which present both hydrophilic and hydrophobic properties, are very interesting elements for constructing nano-structured objects. In particular, amphiphiles containing biomimetic peptides can be designed to interact with proteins or DNA in a very specific manner, which makes these novel materials be very interesting for medical applications. The analysis of structural, thermodynamic and physical properties of these materials at the atomic scale by means of computer simulations is a topic of growing interest. However, the large size of these molecular complexes is a big challenge for existing simulation techniques. Currently, atomic-scale simulations are applied to small parts of these complexes, while larger systems are studied with coarse-grained models. In collaboration with Carlos Alemán (IMEM-UPC, Barcelona, Spain), we envision the development of new molecular modeling and simulation approaches for the study of these new materials.

Finally mention that this topic, atomic-scale CAD, would be a great framework to carry out novel theoretical research, to promote interdisciplinary education, and to develop new technologies with high valorization potential. Besides, the increasing interest in bio-nanotechnology at LAAS-CNRS, where several teams work on the technological development and the characterization of bio-nano devices, makes this laboratory an ideal place to achieve this (long-term) research project. Furthermore, my research project fits perfectly in one of the two transverse research axes of the laboratory, ALIVE, whose objectives are two-fold: the development of novel bio-inspired nanotechnologies, and the development of technologies for the analysis and understanding of biological systems.

## Chapter 4

### List of Publications

#### Publications in scientific journals

- [1] I. Al-Bluwi, M. Vaisset, T. Siméon, J. Cortés. Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods. *BMC Structural Biology*, 13(Suppl 1):S2, 2013.
- [2] D. Devaurs, L. Bouard, M. Vaisset, C. Zanon, I. Al-Bluwi, R. Iehl, T. Siméon, J. Cortés. MoMA-LigPath: a web server to simulate protein-ligand unbinding. *Nucleic Acids Research*, 41(W1):W297-W302, 2013.
- [3] D. Devaurs, T. Siméon, J. Cortés. Parallelizing RRT on large-scale distributed-memory architectures. *IEEE Transactions on Robotics*, 29(2):571-579, 2013.
- [4] I. Al-Bluwi, T. Siméon, J. Cortés. Motion planning algorithms for molecular simulations: A survey. *Computer Science Review*, 6(4):125-143, 2012.
- [5] G. Paës, J. Cortés, T. Siméon, M.J. O'Donohue, V. Tran. Thumb-loops up for catalysis: A structure/function investigation of a functional loop movement in a GH11 xylanase. *Computational and Structural Biotechnology Journal*, 1(2):e201207001, 2012.
- [6] V. Ruiz de Angulo, J. Cortés, J.M. Porta. Rigid-CLL: Avoiding constant-distance computations in cell linked-lists algorithms. *Journal of Computational Chemistry*, 33(3):294-300, 2012.
- [7] L. Jaillet, F.J. Corcho, J.J. Pérez, J. Cortés. Randomized tree construction algorithm to explore energy landscapes. *Journal of Computational Chemistry*, 32(16):3464-3474, 2011.
- [8] Y. Li, J. Cortés, T. Siméon. Enhancing systematic protein-protein docking methods using ray casting: Application to ATTRACT. *Proteins: Structure, Function, and Bioinformatics*, 79(11):3037-3049, 2011.
- [9] S. Barbe, J. Cortés, T. Siméon, P. Monsan, M. Remaud-Siméon I. André A mixed molecular modeling-robotics approach to investigate lipase large molecular motions. *Proteins: Structure, Function, and Bioinformatics*, 79(8):2517-2529, 2011.
- [10] J. Cortés, S. Barbe, M. Erard, T. Siméon. Encoding molecular motions in voxel maps. *IEEE Transactions on Computational Biology and Bioinformatics*, 8(2):557-563, 2011.

- [11] J. Cortés, S. Carrión, D. Curcó, M. Renaud, C. Alemán. Relaxation of amorphous multichain polymer systems using an inverse kinematics method originating from robotics. *Polymer*, 51(17):4008-4014, 2010.
- [12] J. Cortés, D.T. Le, R. Iehl, T. Siméon. Simulating ligand-induced conformational changes in proteins using a mechanical disassembly method. *Physical Chemistry Chemical Physics*, 12(29):8268-8276, 2010.
- [13] L. Jailliet, J. Cortés, T. Siméon. Sampling-based path planning on configuration-space costmaps. *IEEE Transactions on Robotics*, 6(4):635-646, 2010.
- [14] V. Lafaquière, S. Barbe, S. Puech-Guenot, D. Guieysse, J. Cortés, P. Monsan, T. Siméon, I. André, M. Remaud-Siméon. Control of lipase enantioselectivity by engineering the substrate binding site and access channel. *ChemBioChem*, 10:2760-2771, 2009.
- [15] D. Guieysse, J. Cortés, S. Puech-Guenot, S. Barbe, V. Lafaquière, P. Monsan, T. Siméon, I. André, M. Remaud-Siméon. Structure-controlled lipase enantioselectivity investigated by a path planning approach. *ChemBioChem*, 9:1308-1317, 2008.
- [16] J. Cortés, L. Jailliet, T. Siméon. Disassembly path planning for complex articulated objects. *IEEE Transactions of Robotics*, 24(2):475-481, 2008.
- [17] S. Kirillova, J. Cortés, A. Stefaniu, T. Siméon. An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins. *Proteins: Structure, Function, and Bioinformatics*, 70(1):131-143, 2008.
- [18] J. Cortés, T. Siméon, V. Ruiz de Angulo, D. Guieysse, M. Remaud-Siméon, V. Tran. A path planning approach for computing large-amplitude motions of flexible molecules. *Bioinformatics*, 21(Suppl.1):i116-i125, 2005.
- [19] J. Cortés, T. Siméon, M. Remaud-Siméon, V. Tran. Geometric algorithms for the conformational analysis of long protein loops. *Journal of Computational Chemistry*, 25(7):956-967, 2004.
- [20] T. Siméon, J.-P. Laumond, J. Cortés, A. Sahbani. Manipulation planning with probabilistic roadmaps. *International Journal of Robotics Research*, 23(7-8):729-746, 2004.

## Book chapters

- [21] M. Manubens, D. Devaurs, L. Ros, J. Cortés. Motion planning for 6-D manipulation with aerial towed-cable systems. In: P. Newman, D. Fox, D. Hsu (Eds.) *Robotics: Science and Systems IX*, in press.
- [22] V. Ruiz de Angulo, J. Cortés, T. Siméon. BioCD : An efficient algorithm for self-collision detection and distance computations between highly articulated molecular models. In: S. Thrun, G. Sukhatme, S. Schaal, O. Brock (Eds.) *Robotics: Science and Systems I*. MIT Press, Cambridge. 2005.
- [23] J. Cortés, T. Siméon. Sampling-based motion planning under kinematic loop-closure constraints. In: M. Erdmann, D. Hsu, M. Overmars, A.F. van der Stappen (Eds.). *Algorithmic Foundations of Roadmaps VI*. Springer-Verlag, Berlin. 2005.
- [24] T. Siméon, J. Cortés, A. Sahbani, J.-P. Laumond. A general manipulation task planner. In: J.-D. Boissonnat, J.W. Burdick, K. Goldberg, S. Hutchinson (Eds.) *Algorithmic Foundations of Roadmaps V*. Springer-Verlag, Berlin. 2004.

## Proceeding of international conferences (peer-reviewed)

- [25] D. Devaurs, T. Siméon, J. Cortés. Multi transition-based rapidly-exploring random trees. *IEEE Int. Conf. on Robotics and Automation*, submitted.
- [26] D. Devaurs, M. Vaisset, T. Siméon, J. Cortés. A multi-tree approach to compute transition paths on energy landscape. *Workshop on Artificial Intelligence and Robotics Methods in Computational Biology, AAAI*, 2013.
- [27] D. Devaurs, T. Siméon, J. Cortés. Enhancing the transition-based RRT to deal with complex cost spaces. *IEEE Int. Conf. on Robotics and Automation*, 2013.
- [28] I. Al-Bluwi, M. Vaisset, T. Siméon, J. Cortés. Coarse-grained elastic networks, normal mode analysis and robotics-inspired methods for modeling protein conformational transitions. *Computational Structural Biology Workshop, ACM-BCB*, 2012.
- [29] J. Cortés, I. Al-Bluwi. A robotics approach to enhance conformational sampling of proteins. *ASME Mechanisms and Robotics Conference*, 2012.
- [30] R. Iehl, J. Cortés, T. Siméon. Costmap planning in high-dimensional configuration spaces. *IEEE/ASME Int. Conf. on Advanced Intelligent Mechatronics*, 2012.
- [31] D. Devaurs, T. Siméon, J. Cortés. Parallelizing RRT on distributed-memory architectures. *IEEE Int. Conf. on Robotics and Automation*, 2011.
- [32] Y. Li, J.P. Saut, J. Cortés, T. Siméon, D. Sidobre. Finding enveloping grasps by matching continuous surfaces. *IEEE Int. Conf. on Robotics and Automation*, 2011.
- [33] J. Mainprice, E. A. Sisbot, L. Jaillet, J. Cortés, R. Alami, T. Siméon. Planning human-aware motions using a sampling-based costmap planner. *IEEE Int. Conf. on Robotics and Automation*, 2011.
- [34] J.-P. Saut, M. Gharbi, J. Cortés, D. Sidobre, T. Siméon. Planning pick and place tasks with two-hand regrasping. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2010.
- [35] M. Gharbi, J. Cortés, T. Siméon. Roadmap composition for multi-arm systems path planning. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2009.
- [36] D.T. Le, J. Cortés, T. Siméon. A path planning approach to (dis)assembly sequencing. *IEEE Conf. on Automation Science and Engineering*, 2009.
- [37] J. Cortés, S. Barbe, M. Erard, T. Siméon. Encoding molecular motions in voxel maps. *IEEE Int. Conf. on Robotics and Automation*, 2009.
- [38] L. Jaillet, J. Cortés, T. Siméon. Transition-based RRT for path planning in continuous cost spaces. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2008.
- [39] M. Gharbi, J. Cortés, T. Siméon. A sampling-based path planner for dual-arm manipulation. *IEEE/ASME Int. Conf. on Advanced Intelligent Mechatronics*, 2008.
- [40] J. M. Porta, J. Cortés, L. Ros, F. Thomas. A space decomposition method for path planning of loop linkages. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2007.
- [41] J. Cortés, T. Siméon. Disassembly path planning for objects with articulated parts. *IFAC Workshop on Intelligent Assembly and Disassembly*, 2007.
- [42] J. Cortés, L. Jaillet, T. Siméon. Molecular disassembly with RRT-like algorithms. *IEEE Int. Conf. on Robotics and Automation*, 2007.

- [43] J. Cortés, T. Siméon. Probabilistic motion planning for parallel mechanisms. *IEEE Int. Conf. on Robotics and Automation*, 2003.
- [44] A. Sahbani, J. Cortés, T. Siméon. A probabilistic algorithm for manipulation planning under continuous grasps and placements. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2002.
- [45] J. Cortés, T. Siméon, J.-P. Laumond. A random loop generator for planning the motions of closed kinematic chains using PRM methods. *IEEE Int. Conf. on Robotics and Automation*, 2002.
- [46] T. Siméon, J. Cortés, A. Sahbani, J.-P. Laumond. A manipulation planner for pick and place operations under continuous grasps and placements. *IEEE Int. Conf. on Robotics and Automation*, 2002.
- [47] T. Siméon, J.-P. Laumond, C. van Geem, J. Cortés. Computer aided motion: Move3D within MOLOG. *IEEE Int. Conf. on Robotics and Automation*, 2001.

### Other publications in conferences and workshops

- [48] D. Devaurs, L. Bouard, M. Vaisset, C. Zanon, I. Al-Bluwi, R. Iehl, T. Siméon, J. Cortés. MoMA-LigPath: a web server to simulate protein-ligand unbinding. *Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM)*, 2013.
- [49] S. Barbe, R. Iehl, J. Cortés, T. Siméon, M. Remaud-Siméon, I. André. How to consider molecular motions and flexibility in enzyme catalysis and design? *Congrès du Group. Graphisme et Modélisation Moléculaire (GGMM)*, 2013.
- [50] I. Al-Bluwi, M. Vaisset, T. Siméon, J. Cortés. Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods. *Congrès du Group. Graphisme et Modélisation Moléculaire (GGMM)*, 2013.
- [51] L. Jaillet, J. Cortés. Transition-based RRT for exploring the energy landscape of biomolecules. *Workshop on the Algorithmic Foundations of Robotics*, 2011.
- [52] L. Jaillet, J. Cortés. A randomized tree construction algorithm to explore energy landscapes. *Robotics Science and Systems Workshop on Motion Planning: From Theory to Practice*, 2010.
- [53] V. Lafaquière, S. Barbe, S. Guenot, D. Guieysse, J. Cortés, T. Siméon, P. Monsan, I. André, M. Remaud. A structured-controlled lipase enantioselectivity investigated by a path planning approach. *ESF-EMBO Symposium Protein Design and Evolution for Biocatalysis*, 2009.
- [54] S. Barbe, V. Lafaquière, D. Guieysse, J. Cortés, P. Monsan, T. Siméon, M. Remaud, I. André. Insights into molecular motions of Burkholderia cepacia lipase inferred from mixed molecular modelling and robotic-based path planning approaches. *ESF-EMBO Symposium Protein Design and Evolution for Biocatalysis*, 2009.
- [55] V. Lafaquière, S. Barbe, S. Guenot, D. Guieysse, J. Cortés, T. Siméon, P. Monsan, I. André, M. Remaud. Toward the building of a predictive model of Burkholderia cepacia lipase enantioselectivity. *8<sup>th</sup> International Symposium on Biocatalysis and Biotransformations*, 2008.
- [56] K. Bastard, J. Cortés, G. Paës, M. Levitt, V. Tran. Computing ligand release pathway efficiency: A comparison between robotics approach and molecular dynamics simulations. *Trends in Enzymology*, 2008.

- 
- [57] S. Barbe, I. André, J. Cortés, M. Remaud, T. Siméon. A robotics-based path planning approach for computing large-amplitude motions of flexible proteins. *Congrès du Group. Graphisme et Modélisation Moléculaire (GGMM)*, 2007.
- [58] S. Barbe, V. Lafaquière, S. Guenot, G. Guieysse, J. Cortés, T. Siméon, P. Monsan, I. André, M. Remaud. Toward the building of a predictive model of Burkholderia cepacia lipase enantioselectivity. *Congrès du Group. Graphisme et Modélisation Moléculaire (GGMM)*, 2007.
- [59] V. Tran, G. Paës, J. Cortés, T. Siméon. Couplage robotique-modélisation moléculaire pour l'analyse dynamique de mouvements locaux. *Congrès du Group. Graphisme et Modélisation Moléculaire (GGMM)*, 2007.
- [60] J. Cortés, T. Siméon. A robotic-based path planning approach for computing large-amplitude motions of flexible molecules. *3<sup>ème</sup> Journées Annuelle "GénoToul"*, 2006.
- [61] J. Cortés. Path planning for closed kinematic chains. *Journées Doctorales d'Automatique*, 2001.
- [62] J. Cortés. Planification de mouvements pour chaînes cinématiques fermées. *14<sup>èmes</sup> Journées des Jeunes Chercheurs en Robotique*, 2001.

### Invited talks

- [63] J. Cortés. Simulating large time-scale molecular motions within CPU minutes : A robotics approach. *International Symposium Frontiers in Protein Research*, Lisbon, 2011.
- [64] J. Cortés. Extending RRTs to solve problems in structural biology. *Robotics Science and Systems Workshop on Motion Planning: From Theory to Practice*, Zaragoza, 2010.
- [65] J. Cortés. Algorithmics of motion: from robotics, through structural biology, toward molecular CAD/CAM. *Workshop on NanoBiotechnology & IT for Life Sciences and Medicine*, Barcelona, 2010.
- [66] J. Cortés. A robotics approach to rational enzyme engineering. *GDR-CNRS "New Approaches to Directed Evolution of Proteins"*, Bats-mer, 2010.
- [67] J. Cortés. Path planning algorithms: from robotics to structural biology. *International Meeting on Flexibility and Biological Recognition*, Sophia-Antipolis, 2009.

### Seminars

- [68] J. Cortés. Simulating molecular motions with robotics-inspired algorithms. Seminar at the *Institut de Recherche en Informatique et Systèmes Aléatoires*, Rennes, 2013.
- [69] J. Cortés. Robotics-inspired algorithms for modeling protein motions and interactions. Seminar at the *Centre de Biologie Structurale*, Montpellier, 2012.
- [70] J. Cortés. Simulating molecular motions with robotics-inspired algorithms. Seminar at the *Laboratoire de Biologie Moléculaire des Eucaryotes*, Toulouse, 2012.
- [71] J. Cortés. Simulating molecular motions with robotics-inspired algorithms. Seminar at the *Institut Européen de Chimie et Biologie*, Bordeaux, 2011.
- [72] J. Cortés. Computing molecular motions with robotics algorithms. Seminar at the *Barcelona Supercomputing Center*, Barcelona, 2010.

- [73] J. Cortés. Path planning algorithms: from robotics to structural biology. Seminar at the *Institut de Biologie Physico-Chimique*, Paris, 2009.
- [74] J. Cortés. Path planning in structural biology. Seminar at the *Institut de Robòtica i Informàtica Industrial*, Barcelona, 2008.
- [75] J. Cortés. Path planning in structural biology. Seminar at the *University of California*, Berkeley, 2007.
- [76] J. Cortés. Path planning algorithms for computing protein motions. Seminar at *Tel-Aviv University*, Tel-Aviv, 2005.

## PhD Thesis

- [77] J. Cortés. Algorithmes pour la planification de mouvements de mécanismes articulés avec chaînes cinématiques fermées. *Thèse de l'Institut National Polytechnique de Toulouse*, 2003.

## Research reports

- [78] D. Sidobre, X. Broquere, J. Mainprice, W. He, J. Cortés, T. Siméon, F. Ingrand, M. Herrb, J. Manhes, R. Alami. Learning from previous manipulation actions and planned results. European Project DEXMART ICT-FP7216239. *Rapport LAAS n.12381*, 2012.
- [79] D. Sidobre, J.P. Saut, X. Broquere, J. Mainprice, J. Cortés, T. Siméon, M. Gharbi, F. Ingrand, M. Herrb, J. Manhes, R. Alami, R. Chatila. Model of human interaction. European Project DEXMART ICT-FP7216239. *Rapport LAAS n.10473*, 2010.
- [80] D. Sidobre, M. Herrb, J. Manhes, X. Broquere, R. Alami, R. Chatila, J. Cortés, F. Ingrand, J.P. Saut, T. Siméon. Models of information exchanged by two humans exchanging an object for several strategies of handling. European Project DEXMART ICT-FP7216239. *Rapport LAAS n.10360*, 2010.
- [81] T. Siméon, J. Cortés, E.A. Sisbot, M. Gharbi, J. Mainprice, R. Alami. Motion planning algorithms for safe and legible manipulator trajectories. European Project PHRIENDS FP6-045359. *Rapport LAAS n.08864*, 2008.
- [82] T. Siméon, J. Cortés, E.A. Sisbot, M. Gharbi, R. Alami. Models for safe and legible manipulator trajectories. European Project PHRIENDS FP6-045359. *Rapport LAAS n.07840*, 2007.
- [83] T. Siméon, J. Cortés. Handling complex kinematic environments. MOVIE Report. *Rapport LAAS n.05117*, 2005.
- [84] T. Siméon, J. Cortés, A. Sahbani, J.-L. Bouchet, M.H. Overmars, N. McPhater. MOLOG final report and third year deliverables. *Rapport LAAS n.02086*, 2002.
- [85] T. Siméon, J.-P. Laumond, C. van Geem, J. Cortés, A. Sahbani, F. Lamiroux, M.H. Overmars, N. McPhater, J.-L. Bouchet. Second year deliverables of the MOLOG project. *Rapport LAAS n.01207*, 2001.
- [86] C. Nissoux, T. Siméon, F. Lamiroux, J. Cortés, C. van Geem. Move3D Programming Manual. *Rapport LAAS n.01133*, 2001.

# Annex 1: Selected Publications

- **Motion planning for closed-chain mechanisms :**  
J. Cortés, T. Siméon. “Sampling-based motion planning under kinematic loop-closure constraints”. In: M. Erdmann, D. Hsu, M. Overmars, A.F. van der Stappen (Eds.). *Algorithmic Foundations of Roadmaps VI*. Springer-Verlag, Berlin. 2005.
- **Path planning in high-dimensional costmaps :**  
L. Jaillet, J. Cortés, T. Siméon. “Sampling-based path planning on configuration-space costmaps”. *IEEE Transactions on Robotics*, 6(4):635-646, 2010.
- **Manipulation planning :**  
T. Siméon, J.-P. Laumond, J. Cortés, A. Sahbani. “Manipulation planning with probabilistic roadmaps”. *International Journal of Robotics Research*, 23(7-8):729-746, 2004.
- **Protein loop motions :**  
J. Cortés, T. Siméon, M. Remaud-Siméon, V. Tran. “Geometric algorithms for the conformational analysis of long protein loops”. *Journal of Computational Chemistry*, 25(7):956-967, 2004.
- **Protein conformational transitions :**  
I. Al-Blawi, M. Vaisset, T. Siméon, J. Cortés. “Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods”. *BMC Structural Biology*, 13(Suppl 1):S2, 2013.
- **Protein-ligand interactions :**  
J. Cortés, D.T. Le, R. Iehl, T. Siméon. “Simulating ligand-induced conformational changes in proteins using a mechanical disassembly method”. *Physical Chemistry Chemical Physics*, 12(29):8268-8276, 2010.
- **Enzyme engineering :**  
V. Lafaquière, S. Barbe, S. Puech-Guenot, D. Guieysse, J. Cortés, P. Monsan, T. Siméon, I. André, M. Remaud-Siméon. “Control of lipase enantioselectivity by engineering the substrate binding site and access channel”. *ChemBioChem*, 10:2760-2771, 2009.



# Sampling-Based Motion Planning under Kinematic Loop-Closure Constraints

Juan Cortés and Thierry Siméon

LAAS-CNRS, Toulouse, France

**Abstract.** Kinematic loop-closure constraints significantly increase the difficulty of motion planning for articulated mechanisms. Configurations of closed-chain mechanisms do not form a single manifold, easy to parameterize, as the configurations of open kinematic chains. In general, they are grouped into several subsets with complex and a priori unknown topology. Sampling-based motion planning algorithms cannot be directly applied to such closed-chain systems. This paper describes our recent work [7] on the extension of sampling-based planners to treat this kind of mechanisms.

## 1 Introduction

Robot motion planning has led to active research over the two last decades [13]. More recently, several sampling-based approaches (e.g. [12,16]) have been proposed and successfully applied to challenging problems that remained out of scope for previously existing techniques. They allow today to handle practical motion planning problems arising in such diverse fields as robotics, graphic animation, virtual prototyping or computational biology [14].

In this paper, we consider motion planning for closed-chain mechanisms. We present an extended formulation of the motion planning problem in presence of kinematic loop-closure constraints and we introduce a framework for the development of sampling-based algorithms (Sects. 2 and 4). The additional difficulty of this instance of the problem is that feasible configurations form lower-dimensional subsets in the search-space with no available representation. The performance of the very few approaches proposed for closed-chain mechanisms [15,11] (Sect. 3) significantly degrades for reasonably complex systems, mainly due to the difficulty of generating random samples in such subsets. We propose a general and simple geometric algorithm, called Random Loop Generator (RLG), for sampling random configurations satisfying loop-closure constraints (Sect. 5). RLG enables virtually any sampling-based planning algorithm to be extended to closed-chain mechanisms. We have implemented and experimented with its integration within PRM-based and RRT-based planners, obtaining very good results (Sect. 6).

Complex articulated mechanisms with closed kinematic chains appear in all the domains where motion planning techniques can be applied. Figure 1



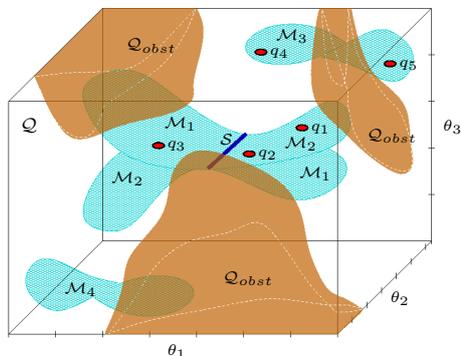
**Fig. 1.** A “closed-chain” version of the piano mover’s problem. The piano is moved by three cooperating mobile manipulators, creating multiple closed kinematic chains.

illustrates an example of coordinated manipulation of an object handled by several robots. The generality and the practical efficiency of the extended planners incorporating RLG allow to tackle such kind of problems as well as problems involving parallel robots, or problems arising in computational biology for the structural analysis of protein loops. All these applications are commented in Sect. 7. We conclude devising new directions for future research (Sect. 8).

## 2 Problem Formulation

The motion planning problem consists in finding a path between two given locations of a mobile system that satisfies intrinsic constraints as well as constraints that arise from the environment. Basically, motion constraints are due to the kinematic structure of the mechanism and to collision avoidance. Under these constraints, the problem can be posed and solved in the *configuration-space*  $\mathcal{C}$  [19]. Then, it is reduced to explore the connectivity of the subset  $\mathcal{C}_{free}$  of the collision-free configurations.

The problem has been clearly formulated for articulated mechanisms without kinematic loops (see [13] for a detailed formulation). In this case,  $\mathcal{C}$  corresponds to the space of the joint variables  $\mathcal{Q}$ , called the *joint-space*. Topologically,  $\mathcal{Q}$  is a smooth manifold, with a simple parameterization [2]. For an articulated mechanism with  $m$  independent joint variables defined in real intervals,  $\mathcal{Q}$  can be seen as a  $m$ -dimensional hypercube. If the articulated mechanism contains closed kinematic chains, then some joint variables are related by loop-closure equations [20]. A general expression of loop-closure constraints is:  $f(\mathcal{Q}) = \mathbf{I}$ , where  $f(\mathcal{Q})$  is a system of non-linear equations and  $\mathbf{I}$  is the identity displacement. The configuration-space  $\mathcal{C}$  of a closed-chain mechanism is the subset of  $\mathcal{Q}$  satisfying such equations. The stratification of  $\mathcal{C}$  leads to several  $\rho$ -dimensional manifolds  $\mathcal{M}_i$  which can be connected through sets of lower dimension  $\mathcal{S}_k$  [3,27]. Note that  $\rho$  corresponds to the global mobility of the mechanism. The  $\mathcal{M}_i$  are called *self-motion manifolds* and the  $\mathcal{S}_k$  are sets of singular configurations. The number of self-motion manifolds is bounded, and it tends to decrease as  $\rho$  increases [3].



**Fig. 2.** Illustration for the formulation of the motion planning problem under kinematic loop-closure constraints. Configurations are grouped into several subsets embedded in the joint-space.

Figure 2 illustrates a fictive example with three joint variables  $\{\theta_1, \theta_2, \theta_3\}$  and  $\rho = 2$ . Let us consider a function of the form  $f(\theta_1, \theta_2, \theta_3) = 0$ , representing loop-closure constraints. This function maps to several surfaces embedded in the joint-space  $\mathcal{Q}$ . Such surfaces are the different self-motion manifolds  $\mathcal{M}_i$ . In this example,  $\mathcal{M}_1$  and  $\mathcal{M}_2$  intersect at a singular set  $\mathcal{S}$ . We have represented the obstacle region  $\mathcal{Q}_{obst}$  in the joint-space.  $\mathcal{Q}_{free}$  is the complementary subset:  $\mathcal{Q}_{free} = \mathcal{Q} \setminus \mathcal{Q}_{obst}$ .  $\mathcal{C}_{free}$  is the intersection of  $\mathcal{Q}_{free}$  with the different  $\mathcal{M}_i$ . Several situations can rise in motion planning queries within this example. The best case is a query for a path between  $q_1$  and  $q_2$ . These configurations lie on the same self-motion manifold and in the same connected component of  $\mathcal{C}_{free}$ , thus there is a free path between them. A path is also feasible between  $q_1$  and  $q_3$ , even if it contains singular configurations. However, for  $q_4$  and  $q_5$ , the presence of obstacles makes these configurations cannot be connected by a free path. Finally, a case that does not appear for open kinematic chains can rise under closure constraints:  $q_1$  and  $q_4$  lie on the same connected component of  $\mathcal{C}$  !

### 3 Available Techniques

Only a few exact motion planning approaches have been proposed that treat kinematic loop-closure and collision avoidance simultaneously, including these two types of constraints within algebraic expressions (e.g. [4,1]). The computational complexity of these approaches makes them unpractical. Normally, different techniques have to be combined for planning motions under kinematic loop-closure constraints. First, loop-closure equations must be solved to obtain the configuration-space  $\mathcal{C}$ . Then, motion planning algorithms can be applied to compute paths in the collision-free subset  $\mathcal{C}_{free}$ .

Techniques that provide a complete solution of loop-closure equations are very limited in practice. Currently, they can be applied to non-redundant mechanisms, single loops with only a few (two or three) degrees of redundancy or particular classes of parallel mechanisms [22,21,23]. For more complex closed-chain mechanisms, only discrete points in  $\mathcal{C}$  can be obtained. The

use of a grid for globally representing  $\mathcal{C}$  is not applicable to high-dimensional spaces. The remaining possibility is then to use sampling techniques combined with numerical or algebraic techniques to obtain single configurations satisfying loop-closure equations. This fact restrains the choice of motion planning algorithms to those based on sampling.

Sampling-based planners have demonstrated their efficacy for solving difficult problems in high-dimensional spaces. The *Probabilistic RoadMap* (PRM) [12] and the *Rapidly-exploring Random Trees* (RRT) [16] are two approaches that have had a particular success. However, only two attempts had been made to extend such sampling-based planners to closed-chain mechanisms [15,11].

The first PRM-based approach able to handle mechanisms with closed chains was presented in [15]. The problem is formulated in the joint-space  $\mathcal{Q}$ . Closure constraints are expressed by error functions involving distances in the Euclidean space. Numerical optimization techniques are used to sample and to connect configurations in the subset of  $\mathcal{Q}$  satisfying these constraints within a given tolerance. The approach is general but suffers the drawbacks of optimization-based methods to solve inverse kinematic problems: they are exposed to the local minima problem and the convergence can be very slow. A technique to randomly sample the tangent space of the constraints is proposed that increases the efficiency of the process to connect sampled configurations. More details of the method and the extension of RRT-based algorithms were subsequently published in [29,30]. For the RRT approach, the random configurations used to bias the exploration are generated ignoring closure constraints. The argument is that computing closed configurations is too expensive and does not provide appreciable benefit. We will show in Sect. 6 that this last assertion is not totally right.

The approach described in [11] treats closed kinematic chains within a PRM-based planner. Each loop in the mechanism is broken into two subchains. For computing nodes, uniform random sampling is used to generate the configuration of one of the subchains (called *active subchain*) and then an inverse kinematics problem is solved to obtain the configuration of the remaining part of the loop (called *passive subchain*) in order to force closure. For computing edges, the local planner is limited to act on the active configuration parameters and the corresponding passive variables are computed for each intermediate configuration along the local path. For the efficiency of the roadmap computation, the passive subchain of each loop must be a non-redundant mechanism with closed-form inverse kinematics solution. As the authors admit, this means an important drawback when the approach is applied to a highly-redundant loop. The probability of randomly generating configurations of a long active subchain for which a configuration of the passive chain satisfying closure constraints exists is very low. The performance of the algorithm drops off significantly due to this fact.

Our approach shares some ideas used for the extension of PRM-based planners in [11]. One of our contributions is to resolve the main drawback of the referred technique by the integration of the RLG sampling technique (explained in Sect. 5).

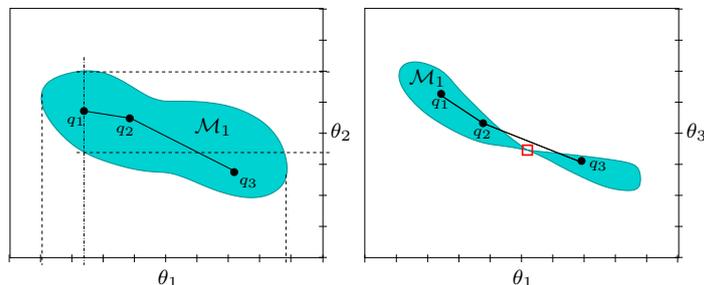
## 4 Sampling-Based Planning and Closed-Chain Mechanisms

This section presents a general framework to extend sampling-based approaches for planning the motions of general closed-chain mechanisms. The use of sampling-based planners is strongly justified since, for the kind of problems we address (see Sect. 7), there is no available technique providing a complete, exact representation of  $\mathcal{C}_{free}$ . However, there are important difficulties for sampling and for checking the connectivity of configurations of closed-chain mechanisms. Next, we discuss how to deal with these difficulties.

**On the Parameterization of  $\mathcal{C}$ :** Configurations of a closed-chain mechanism are grouped into a finite number of lower-dimensional manifolds  $\mathcal{M}_i$  embedded in the joint-space  $\mathcal{Q}$  (the search-space in our problem). These manifolds can be parameterized, at least locally, by a set of  $\rho$  independent parameters, selected from the joint variables. Points in the different manifolds can be generated by sweeping the  $\rho$  parameters through their range and evaluating the loop-closure equations that provide the value of the other (dependent) variables. An *atlas* of each one of these manifolds  $\mathcal{M}_i$  can be constituted by a finite number (in general not exceeding the dimension of  $\mathcal{Q}$ ,  $m$ ) of *local charts* considering different combinations of  $\rho$  parameters. Without loss of generality, sets of  $\rho$  consecutive joint variables in a kinematic chain can be chosen as local coordinates [27].

Following a terminology also used in [11], we call *active variables*  $q^a$  the set of the  $\rho$  joint variables chosen as parameters of a local chart and *passive variables*  $q^p$  the remaining set of the  $m - \rho$  dependent joint variables, so that  $\{q^a, q^p\} = q \in \mathcal{C} \subset \mathcal{Q}$ .

**Main Principle:** The core of our approach is to explore the connectivity of  $\mathcal{C}_{free}$  by sampling configurations and by testing feasible connections through local parameterizations of  $\mathcal{C}$ . Motion planning algorithms are applied on the local parameters  $q^a$ . Using a roadmap method such as PRM, the nodes are generated by sampling  $q^a$  and local paths are obtained by applying a local planner (also called *steering method*) to these parameters. In a similar way,  $q^a$  are the configuration parameters directly handled by incremental search methods like RRT. Obviously, for each computed value of the parameters, loop-closure equations must be solved for obtaining the whole configuration of the mechanism  $q \in \mathcal{C}$ . Therefore, the efficiency of the planner partially relies on the efficient solution of these equations.



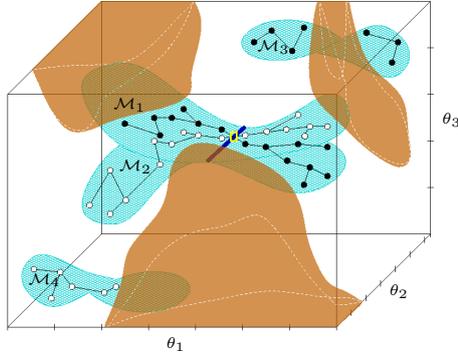
**Fig. 3.** Projection of  $\mathcal{M}_1$  on the planes  $\theta_1\theta_2$  and  $\theta_1\theta_3$ .

**Configuration Sampling:** Given a set of active variables  $q^a$ , loop-closure equations have real solutions only for a range of values of each joint variable, that we call the *closure range*. Besides, the closure range of a parameter depends on the value of the other parameters. The last assertion is illustrated in Fig. 3, that shows the projection of the manifold  $\mathcal{M}_1$  of Fig. 2 on the planes  $\theta_1\theta_2$  and  $\theta_1\theta_3$ . Let us consider  $q^a = \{\theta_1, \theta_2\}$  (the left image). If we sample first  $\theta_1$  and then  $\theta_2$  for generating a configuration  $q_1 \in \mathcal{M}_1$ , then  $\theta_1$  can be sampled in its whole closure range (i.e. the feasible range for any value of the other joint variables). However,  $\theta_2$  is valid only in a subset of its whole closure range, determined by the value of  $\theta_1$ .

There is no general and efficient method to define closure ranges of joint variables. Thus, in practice, the only possibility for sampling configurations is to use a trial method: sampling parameters  $q^a$  in the intervals defined for joint variables and solving the loop-closure equations. Nevertheless, when a closure range is very restricted with respect to the interval of a joint variable, too many samples maybe tested before finding a feasible configuration. Hence, too much computing time is spent in solving closure equations leading to imaginary values. This is an important drawback for the efficiency of motion planners, and mainly for those using a roadmap approach, such as the planner in [11]. The RLG algorithm, further discussed in Sect. 5, resolves this problem using simple geometrical operations.

**Computing Local Path:** Sampling-based roadmap methods check the connectivity of nearby configurations by local paths. Incremental search methods try to expand a configuration toward a local goal, producing also a sort of local path. Such paths can be computed by explicitly varying the local parameters  $q^a$  and solving loop-closure equations with a given resolution for obtaining  $q^p$ .

Without a topological characterization of  $\mathcal{C}$ , the number of different sets of parameters  $q^a$  required to determine if two configurations are connected or not is a priori unknown. This fact can easily be understood on Fig. 3. Let us consider that a linear interpolation of the parameters  $q^a$  produces a kinematically feasible path (i.e. there are no differential constraints). Then,



**Fig. 4.** Probabilistic roadmap capturing the connectivity of  $\mathcal{C}_{free}$  for the fictive motion planning problem involving kinematic closure constraints of Sect. 2.

any two configurations lying on  $\mathcal{M}_1$  can be connected (directly or indirectly) using  $q^a = \{\theta_1, \theta_2\}$  as local parameters. On the contrary, choosing  $q^a = \{\theta_1, \theta_3\}$  leads to a non-complete solution of motion planning queries within  $\mathcal{M}_1$ . A solution path between configurations  $q_1$  and  $q_2$  can be immediately obtained since they are directly connected by a local path. However, a path between  $q_1$  and  $q_3$  can not be found using sampling-based techniques. The point indicated by the small square is a singularity of this parameterization. The probability of generating this point by sampling values of  $q^a = \{\theta_1, \theta_3\}$  is null, as well as the probability of sampling two points on a line (local path) passing through this singularity. Thus, finding a feasible path between  $q_1$  and  $q_3$  requires another set of local parameters than  $\{\theta_1, \theta_3\}$ .

Local paths have to satisfy other motion constraints besides loop-closure, such as collision avoidance. In this paper, we do not talk about these other constraints, that can be checked along local paths by techniques (e.g. [18]) similar to those used for open-chain mechanisms.

**Dealing with Kinematic Singularities:** Up to now we have limited our discussion to the case of a single manifold. However,  $\mathcal{C}$  may be composed of several manifolds. These manifolds are either disjoint, or they intersect at lower-dimensional subsets corresponding to kinematically singular configurations. Therefore, exploring the connectivity of  $\mathcal{C}_{free}$  requires to deal with these singularities.

The roadmap represented in Fig. 4 has been built using  $q^a = \{\theta_1, \theta_2\}$  as only set of parameters. Each point  $\{\theta_1, \theta_2\}$  maps to two configurations (i.e. there are two solutions for  $q^p = \theta_3$ ) on different manifolds for all the domain under  $\mathcal{M}_1$  and  $\mathcal{M}_2$  except for the singular set  $\mathcal{S}$  where they intersect, so that configurations in these manifolds are connected by this singular set. This kinematic singularity corresponds to the singularity of the parameterization  $q^a = \{\theta_1, \theta_3\}$  commented above. It can be traversed by a local path on the current parameterization. When a local path is computed between two configurations at different sides of  $\mathcal{S}$ , the bifurcation of the solution can be easily detected when checking the local path validity, and the singular configuration

(marked by the small rectangle in Fig. 4) connecting  $\mathcal{M}_1$  and  $\mathcal{M}_2$  can then be identified.

Let us consider now a more difficult case where  $\mathcal{M}_1$  and  $\mathcal{M}_2$  do not intersect along a line but meet at a point. None of the three possible parameterizations would allow to identify such a singular point exactly. The difference in this case is that the singular set has dimension  $\rho - 2$  instead of  $\rho - 1$ . In theory, sets of kinematically singular configurations can have dimension from  $\rho - 1$  to zero. Using sampling techniques for generating configurations on  $\mathcal{C}$  and steering methods on subsets of configuration parameters  $q^a$ , our approach has only the guarantee (if we do not admit a tolerance) to find connections through singular sets of dimension  $\rho - 1$ . The other singular sets, from dimension  $\rho - 2$  to isolated singularities, must be identified by other methods. The general treatment of such singularities goes beyond the scope of this paper. As far as we know, techniques able to globally characterize singular configurations have been proposed only for particular classes of mechanisms (e.g. [10,28]).

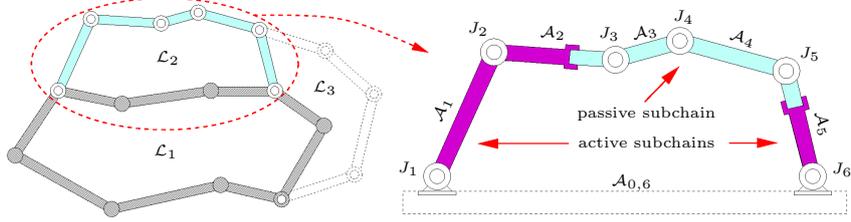
## 5 The RLG Algorithm

We have developed an algorithm, that we call Random Loop Generator (RLG), for sampling configurations of closed-chain mechanisms. The general approach was presented in [5]. Then, a variant that treats more efficiently parallel mechanisms was introduced in [6]. More details can be found in [7]. In this section, we present an overview containing main ideas.

### 5.1 Mechanical System Decomposition

**General Case:** RLG is based on a decomposition of the mechanism into open kinematic chains. In the general case, single loops are handled separately, in a determined order. For each loop, sets of active and passive joint variables are defined consecutively such that they correspond to segments of the kinematic chain. We call *passive subchain* the segment involving the passive variables and *active subchains* to the other segments. There can be one or two active subchains depending on the placement of the passive subchain. The passive subchain is a non-redundant mechanism whose end-frame can span full-rank subsets of the workspace. In general, this requires three joint variables for a planar mechanism and six for a spatial mechanism. Efficient methods to solve inverse kinematics problems for such mechanisms are usually available [22].

In multi-loops, some joint variables are involved in the configuration of several individual loops. Their value is computed for the single loop treated first, and then, these common portions of the mechanism become rigid bodies when treating the other loops. Figure 5 shows an example of a planar multi-loop mechanism. The individual loops are designated by  $\mathcal{L}_i$ , where the index



**Fig. 5.** Decomposition of a multi-loop in single loops, and a possible choice of passive and active subchains in loop  $\mathcal{L}_2$ .

$i$  indicates the order for the treatment. The figure also illustrates the decomposition of  $\mathcal{L}_2$ . This 6R planar linkage has mobility  $\rho = 3$ . Thus,  $q^a$  and  $q^p$  contain three joint variables each. In this illustration we have chosen  $\theta_3$ ,  $\theta_4$  and  $\theta_5$  (the variables associated with joints  $J_3$ ,  $J_4$  and  $J_5$ ) to be the passive variables. Then, active variables can be seen as configuration parameters of two open chains rooted at a (fictive) link  $\mathcal{A}_{0,6}$ .

**Parallel Mechanisms:** A parallel mechanism is an articulated multi-loop structure in which a solid, the end-effector or platform  $\mathcal{P}$ , is connected to the base  $\mathcal{A}_0$  by at least two independent kinematic chains  $\mathcal{K}_i$ . The pose of  $\mathcal{P}$  is defined by a vector  $q_{\mathcal{P}} = \{x_{\mathcal{P}}, y_{\mathcal{P}}, z_{\mathcal{P}}, \gamma_{\mathcal{P}}, \beta_{\mathcal{P}}, \alpha_{\mathcal{P}}\}$ . The three first elements represent the position of  $F_{\mathcal{P}}$  relative to  $F_{\mathcal{A}_0}$ , the frames associated with  $\mathcal{P}$  and  $\mathcal{A}_0$  respectively. The orientation is given by three consecutive rotations around the coordinate axes of  $F_{\mathcal{P}}$ <sup>1</sup>. We consider the configuration  $q$  of a parallel mechanism is defined by the platform pose and the configuration of the chains  $\mathcal{K}_i$ :  $q = \{q_{\mathcal{P}}, q_{\mathcal{K}_1}, \dots, q_{\mathcal{K}_{n_k}}\}$ . The parameters defining the platform pose  $q_{\mathcal{P}}$  are selected as active variables. For a given platform pose, the base-frame and the end-frame of each chains  $\mathcal{K}_i$  have fixed relative location. Thus, they have to be treated as closed kinematic chains. If a chain  $\mathcal{K}_i$  is redundant, it is decomposed in active and passive subchains as explained above for a single loop. Thus, we can separate joint variables as follows:  $q_{\mathcal{K}_i} = \{q_{\mathcal{K}_i}^a, q_{\mathcal{K}_i}^p\}$ . Hence, variables defining the configuration  $q$  of a parallel mechanism are divided into active and passive such that:

$$q^a = \{q_{\mathcal{P}}, q_{\mathcal{K}_1}^a, \dots, q_{\mathcal{K}_{n_k}}^a\}, \quad q^p = \{q_{\mathcal{K}_1}^p, \dots, q_{\mathcal{K}_{n_k}}^p\}$$

## 5.2 Sampling process

RLG performs a “guided”-random sampling for  $q^a$  that notably increases the probability of obtaining real solutions for  $q^p$  when solving the loop-closure equations. Figure 6 illustrates the process on a planar 6R linkage (the loop  $\mathcal{L}_2$  in Fig. 5). The active joint variables are computed sequentially by the function `SAMPLE- $q^a$`  detailed in Algorithm 1. The two active subchains are

<sup>1</sup> RLG could also handle other parameterizations (e.g. Euler angles).

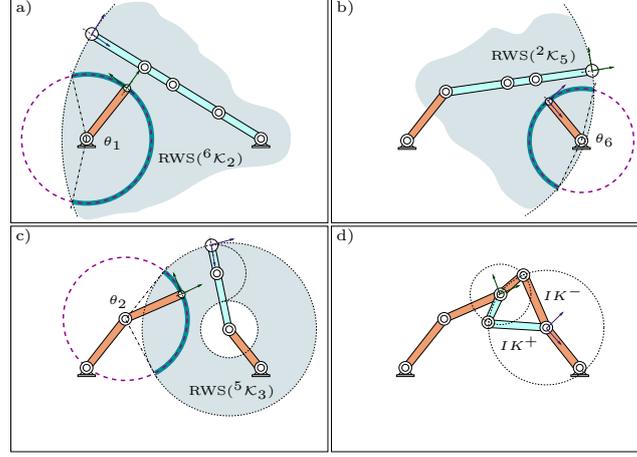


Fig. 6. Steps of the RLG algorithm performing on a 6R planar linkage.

---

**Algorithm 1:** SAMPLE- $q^a$ 


---

```

input   : the loop  $\mathcal{L}$ 
output  : the parameters  $q^a$ 
begin
1   $(J_b, J_e) \leftarrow \text{INITSAMPLER}(\mathcal{L});$ 
   while not ENDACTIVECHAIN( $\mathcal{L}, J_b$ ) do
      $I_c \leftarrow \text{COMPUTECLOSURERANGE}(\mathcal{L}, J_b, J_e);$ 
     if  $I_c = \emptyset$  then goto line 1;
     SETJOINTVALUE( $J_b$ , RANDOM( $I_c$ ));
      $J_b \leftarrow \text{NEXTJOINT}(\mathcal{L}, J_b);$ 
     if not ENDACTIVECHAIN( $\mathcal{L}, J_e$ ) then SWITCH( $J_b, J_e$ );
   end
end

```

---

treated alternately. The idea of the algorithm is to progressively decrease the complexity of the closed chain until only the configuration of the passive subchain,  $q^p$ , remains to be solved (by inverse kinematics).

At each iteration, the function COMPUTECLOSURERANGE returns a set of intervals  $I_c$  which approximate the closure range of a joint variable, for a fixed configuration of the portions of loop previously generated. The approximation must be conservative in the sense that no region of  $\mathcal{C}$  is excluded for the sampling. This is required in order to guarantee any form of sampling-based completeness (e.g. probabilistic completeness) of motion planning algorithms. The value of the joint variable is randomly sampled inside the intervals  $I_c$ .

The problem of computing the closure range of a joint variable can be formulated as follows. Given a closed kinematic chain  ${}^b\mathcal{K}_e$  involving joints from

$J_b$  to  $J_e$  (we consider  $b < e$  in this explanation), two open kinematic chains are obtained by breaking the link  $\mathcal{A}_b$  between  $J_b$  and  $J_{b+1}$ . A suitable break-point is the physical placement of  $J_{b+1}$ , but any other point can be chosen. A frame  $F_c$  associated with this break-point can be seen as the end-frame of both open chains. The closure range of the joint variable corresponding to  $J_b$  is the subset of values making  $F_c$  reachable by the chain  ${}^e\mathcal{K}_{b+1}$ . Solving such a problem requires to represent the workspace of this chain, which is in general very complicated. For our purpose, a simple and fast method is preferred versus a more accurate but slower one. RLG handles simple volumes that bound the reachable workspace (i.e. only considering positional reachability). They are denoted by RWS in Fig. 6. In general, a reasonable choice for the RWS is a spherical shell with external and internal radii corresponding respectively to an upper bound of the maximum extension and a lower bound of the minimum extensions of the chain. Once defined  $\text{RWS}({}^e\mathcal{K}_{b+1})$ , computing the approximation of the closure intervals for  $J_b$  is very simple. If  $J_b$  is a revolute joint, then the origin of  $F_c$  describes a circle around its axis. If  $J_b$  is a prismatic joint, the origin of  $F_c$  moves on a straight-line segment. Then,  $I_c$  is obtained from the intersection of a circle or a line with a simple volume RWS.

The function to sample  $q_{\mathcal{P}}$  for a parallel mechanism is based on the same principle that `SAMPLE_` $q^a$ . The parameters are sampled progressively from the computed closure range approximations. The main difference is that the closure range depends now on the satisfaction of closure constraints that simultaneously involve several individual loops.

## 6 Performance of RLG

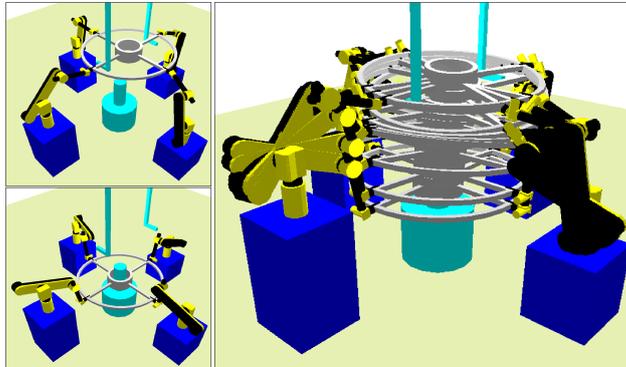
RLG is a general technique, applicable to any sampling-based planner. In this section, we show examples of motion planning problems solved by PRM and RRT planners extended to handle loop-closure constraints. The goal of the experiments is to compare the performance of the planners with and without incorporating RLG. The tests have been made with the software Move3D [25], in which our algorithms have been implemented.

### 6.1 Results with a PRM-based Planner

PRM-based algorithms build a graph (the roadmap) whose nodes are randomly sampled configurations that satisfy motion constraints. For a closed-chain mechanism, only the parameters  $q^a$  are sampled, and loop-closure equations have to be solved to determine if the sample yields a valid configuration or not. We have compared the performance of the planner using an uniform random sampling <sup>2</sup> or RLG for sampling  $q^a$ .

Figure 7 shows four manipulators handling an object. The whole system can be seen as an articulated parallel structure. If the grasps are modeled as

<sup>2</sup> Implemented using the `rand()` function of the GNU C Library.



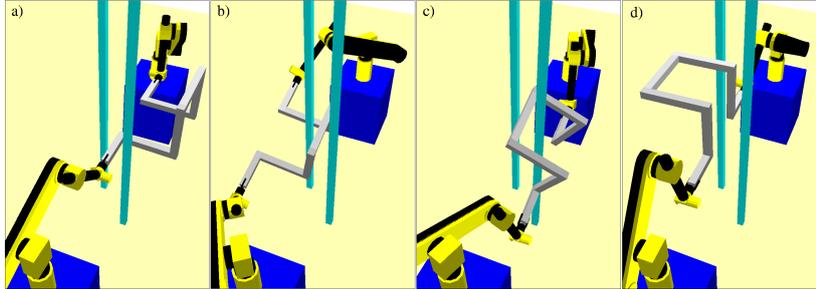
**Fig. 7.** Motion planning problem for four robotic arms manipulating an object. Start and goal configurations (left) and trace of the solution path (right).

fixed attachments, the composed mechanism involves  $m = 30$  joint variables (6 for each manipulator and 6 for the movable object), and the mobility is  $\rho = 6$ . Since the manipulators are non-redundant,  $q^a = q_{\mathcal{P}}$ : the parameters defining the location of the object. Almost all (more than 90%) the configurations  $q_{\mathcal{P}}$  generated by RLG make the object simultaneously reachable by the four manipulators. Using a uniform random sampling, the bounds of the parameters  $q_{\mathcal{P}}$  have first to be adjusted “by hand” with relation to the workspace of the manipulators (this is not necessary for RLG). Even with a good setting of the bounds, less than 0.05% of the samples yield valid (closed) configurations. Let us see now the repercussion when solving motion planning problems using the PRM approach. In the problem illustrated in the figure, the manipulators have to unhook an object and to insert it into the cylindrical axis. For computing a roadmap containing the solution path to this problem, millions of poses generated by uniform random sampling were necessary and the process took more than 20 minutes<sup>3</sup>. Using RLG for sampling configurations, less than 500 random platform poses were generated and the roadmap was built in less than 20 seconds. This result show that RLG avoids an enormous number of futile operations (i.e. calls to inverse kinematics functions) which drop off the performance of the planner.

## 6.2 Results with an RRT-based Planner

The principle of RRT is to use uniformly distributed configurations  $q_{rand}$ , sampled at random in  $\mathcal{C}$ , to bias the expansion of search trees in  $\mathcal{C}_{free}$ . Under loop-closure constraints, the use of samples in  $\mathcal{Q}$  was proposed in [29] in order to evade the cost of sampling  $\mathcal{C}$ . We have compared the performance of a uniform random sampling in  $\mathcal{Q}$  versus RLG for generating  $q_{rand}^a$ .

<sup>3</sup> Tests were performed using a Sun Blade 100 Workstation with a 500-MHz UltraSPARC-IIe processor.



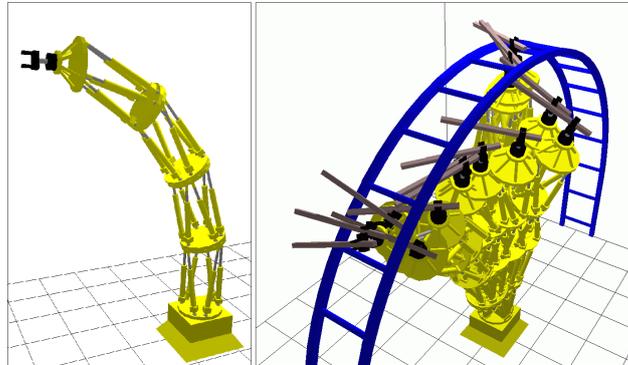
**Fig. 8.** Sequence of the solution to a difficult motion planning problem for two robotic arms manipulating an object among obstacles.

In the example illustrated in Fig. 8, two robotic arms coordinate for manipulating a twisted bar among two vertical bars that restrict its motion. The goal is to solve a motion planning query between configurations in Fig. 8.a and Fig. 8.d. The figure shows a sequence of intermediate configurations of the solution of this puzzle-like problem. The difficulty of this problem depends on the distance between the vertical bars  $d_{bars}$ . We have made tests with three settings:  $d_{bars} = \{150, 175, 200\}$ . In this example,  $q^a$  corresponds to the configuration of one of the arms grasping the bar. The next table shows averaged results of tests. **N** is the number of iterations for expanding the search trees. **T** is the computing time. These results show the importance of an appropriate sampling considering the presence of kinematic loops. The gain obtained using the RLG sampling technique increases with the difficulty of the motion planning problem.

$d_{bars}$	Uniform		With RLG			
	N	T	N	T	gain	T
200	2719	42.26	118	3.23	×	13
175	4995	102.92	424	6.30	×	16
150	9761	312.76	615	9.14	×	34

## 7 Applications of Closed-Chain Motion Planning

We have studied some of the possible applications of algorithms for motion planning under loop-closure constraints. One of them, coordinated manipulation planning, has been illustrated with the two examples in Sec. 6 and the problem shown in Fig. 1. This last problem combines several types of difficulty. First, the virtual structure composed by the three mobile manipulators grasping the piano can be seen as a parallel mechanism with redundant legs (the chains  $\mathcal{K}_i$  correspond to the mobile manipulators). And second, the geometric complexity of the scene makes collision checking very hard. Besides, obstacles are strategically placed in order to hinder the motion of robots for changing the orientation of the piano. A roadmap that permits to solve most



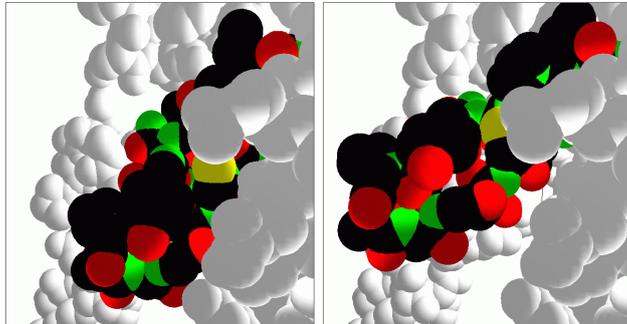
**Fig. 9.** Model of the robot Logabex-LX4, composed of four Gough-Stewart platforms connected in series, and trace of a collision-free path.

possible queries in this scene was computed using the an extended version of the Visibility-PRM algorithm [24] in 5 minutes.

Applied to parallel robots, motion planning algorithms can help designers of these mechanisms, or can provide useful data for real-time trajectory planning. Our work on parallel mechanisms [6] represents the first effective application of sampling-based planners to this kind of articulated structures. The generality of our approach is demonstrated by the complexity of the systems that it is able to treat, such as the model of the Logabex-LX4 (Fig. 9), whose configuration-space is a 25-dimensional variety embedded in a 97-dimensional joint-space. Planning queries for moving the manipulator with the grasped bar from one to another opening of the bridge were solved by the extended RRT-based algorithm in less than one minute.

The above expounded closed-chain planners can also be used as a key component of a novel manipulation planning approach described in [26]. The clever idea is to explore the connectivity of the subset where the manipulation sub-paths (i.e. *transit* and *transfer* paths) meet via a virtual closed-chain mechanism consisting of the robot grasping the movable object placed at a stable position. Our manipulation planner automatically generates, among continuous sets, the grasps and the intermediate placements of the movable object required to solve complicated problems. It is the first general manipulation planner with this capability.

We also began to investigate applications out of the field of robotics. Motion planning techniques can be used as new tools to help the resolution of important open problems in computational biology [9]. We discuss in [8] the application of closed-chain planning techniques to the structural analysis of protein loops. The algorithms that we have presented can act as efficient filters for conformational search methods by making a geometric treatment of strong energetic constraints: maintaining the backbone integrity (i.e. loop-closure) and avoiding steric clashes. We have developed a conformational sampling algorithm that provides random conformations achieving such ge-



**Fig. 10.** Simulated conformational gating of loop 7 in amylosucrase. Crystallographic conformation (left) and geometrically feasible change (right).

ometric constraints. The algorithm combines RLG for generating the backbone conformation with other sampling and collision detection techniques to obtain feasible conformations of the side-chains. We also propose a new conformational search technique, inspired by the RRT approach, for studying geometrically feasible loop motions. Preliminary results are promising. They demonstrate the capacity of the techniques to handle protein models with long loops that involve several dozens of degrees of freedom in the backbone and in the side-chains. Figure 10 illustrates an example of such a long loop in the model of amylosucrase from *Neisseria polysaccharea*.

## 8 Conclusions and Future Work

We have introduced sampling-based motion planning algorithms into an extended formulation of the motion planning problem under kinematic loop-closure constraints. We have tried to give general directives without focusing on a particular implementation. The RLG algorithm allows to overcome the challenge of sampling random configurations for general closed-chain mechanisms. The results obtained, in different domains of application, when solving difficult problems with PRM-based and RRT-based extended planners, demonstrate the efficacy and the generality of the approach.

Several points remain for future research. Some of them concern the RLG sampling algorithm. RLG provided good results in all our experiments. Nevertheless, a deeper analytical work is necessary in order to characterize its performance. Also, studying new forms of sampling, using quasi-random sequences or multi-resolution grids recently proposed [17], seems to be an interesting way to follow. Another improvement involves the selection of active and passive subchains. A general automatic method, based on an analysis of kinematic diagrams of mechanisms, remains to be devised.

A general methodology for the treatment of singularities within sampling-based motion planning algorithms remains an open topic to be further investigated. Recent interval methods [21,23] appear to be another matter to

study for the improvement of closed-chain motion planning techniques. Interval methods can provide a complete approximated representation of the whole set of configurations satisfying loop closure equations. They compute a set of boxes that contain the continuum of solutions. Such a representation is very suitable for the application of sampling-based motion planning algorithms.

Another direction for future research concerns the application of closed-chain motion planning to computational biology. We intend to further study problems in structural biology that require the development of new techniques for the conformational analysis of protein loops. A first goal is to get efficient algorithms to capture the whole subset of the geometrically feasible conformations of one or several loops in the same protein. The next goal is to handle several proteins that interact while changing the conformation of loops on their surfaces. Applications to structural biology are attractive for the algorithmic development because of the kinematic complexity of molecular models. Obviously, robotic applications could also benefit from improvements achieved in this field.

#### Acknowledgments.

This work has been supported by the European project IST-37185 MOVIE and the project BioMove3D of the CNRS Bioinformatics Research Program.

#### References

1. Basu S., Pollack R., Roy M.-F. (2000). Computing Roadmaps of Semi-algebraic Sets on a Variety. *J. Am. Math. Soc.* **13**(1), 55–82
2. Burdick J.W. (1988). Kinematic Analysis and Design of Redundant Robot Manipulators. PhD Thesis, Stanford University.
3. Burdick J.W. (1989). On the Inverse Kinematics of Redundant Manipulators: Characterization of the Self-Motion Manifold. *Proc. IEEE Int. Conf. Rob. & Autom.*, 264–270
4. Canny J.F. (1988). *The Complexity of Robot Motion Planning*. MIT Press
5. Cortés J., Siméon T., Laumond J.-P. (2002). A Random Loop Generator for Planning the Motions of Closed Kinematic Chains using PRM Methods. *Proc. IEEE Int. Conf. Rob. & Autom.*, 2141–2146
6. Cortés J., Siméon T. (2003). Probabilistic Motion Planning for Parallel Mechanisms. *Proc. IEEE Int. Conf. Rob. & Autom.*, 4354–4359
7. Cortés J. (2003). Motion Planning Algorithms for General Closed-Chain Mechanisms. PhD Thesis, Institut National Polytechnique de Toulouse
8. Cortés J., Siméon T., Remaud-Siméon M., Tran V. (2004). Geometric Algorithms for the Conformational Analysis of Long Protein Loops. *J. Comp. Chem.*, *in press*
9. Finn P.W., Kavraki L.E., Latombe J.-C. et al. (1998). RAPID: Randomized Pharmacophore Identification. *Comp. Geom.: Theory & Appl.* **10**(4), 263–272
10. Gosselin C. (1988). Kinematic Analysis, Optimization and Programming of Parallel Robotic Manipulators. PhD Thesis, McGill University

11. Han L., Amato N.M. (2001). A Kinematics-Based Probabilistic Roadmap Method for Closed Kinematic Chains. In: Donald B.R. et al. (Eds.) *Algorithmic and Computational Robotics: New Directions*. A.K. Peters, 233–245
12. Kavraki L.E., Svestka P., Latombe J.-C., Overmars M.H. (1996). Probabilistic Roadmaps for Path Planning in High-Dimensional Configuration Spaces. *IEEE Trans. Rob. & Autom.* **12(4)**, 566–580
13. Latombe J.-C. (1991). *Robot Motion Planning*. Kluwer Academic Publishers
14. Latombe J.-C. (1999). Motion Planning: A Journey of Robots, Molecules, Digital Actors, and Other Artifacts. *Int. J. Rob. Res.* **18(11)**, 1119–1128
15. LaValle S.M., Yakey J.H., Kavraki L.E. (1999). A Probabilistic Roadmap Approach for Systems with Closed Kinematic Chains. *Proc. IEEE Int. Conf. Rob. & Autom.*, 473–479
16. LaValle S.M., Kuffner, J.J. (2001). Rapidly-Exploring Random Trees: Progress and Prospects. In: Donald B.R. et al. (Eds.) *Algorithmic and Computational Robotics: New Directions*. A.K. Peters, 293–308
17. LaValle S.M., Branicky M.S. (2004). On the Relationship Between Classical Grid Search and Probabilistic Roadmaps. In: Boissonnat J.-D. et al. (Eds.). *Algorithmic Foundations of Roadmaps V*. Springer-Verlag, 59–75
18. Lin M.C., Manocha D. (2004). Collision and Proximity Queries. In: *Handbook of Discrete and Computational Geometry: Collision Detection*. To appear. Available at <http://www.cs.unc.edu/~dm/>
19. Lozano-Pérez, T. (1983). Spatial Planning: A Configuration Space Approach. *IEEE Trans. Computers* **32(2)**, 108–120
20. McCarthy J.M. (2000). *Geometric Design of Linkages*. Springer-Verlag
21. Merlet J.-P. (2001). A Parser for the Interval Evaluation of Analytical Functions and its Applications to Engineering Problems. *J. Symb. Comp.* **31**, 475–486
22. Nielsen J., Roth B. (1997). Formulation and Solution for the Direct and Inverse Kinematics Problem for Mechanisms and Mechatronic Systems. *Proc. NATO Adv. Study Inst. on Comp. Meth. in Mech.* **1**, 233–252
23. Porta J.M., Ros L., Thomas F., Torras C. (2003). A Branch-and-Prune Algorithm for Solving Systems of Distance Constraint. *Proc. IEEE Int. Conf. Rob. & Autom.*, 342–348
24. Siméon T., Laumond J.-P., Nissoux C. (2000). Visibility-Based Probabilistic Roadmaps for Motion Planning. *Adv. Rob. J.* **14(6)**, 477–494
25. Siméon T., Laumond J.-P., Lamiraux F. (2001). Move3D: a Generic Platform for Path Planning. *Proc. IEEE Int. Symp. Assembly & Task Planning*, 25–30
26. Siméon T., Laumond J.-P., Cortés J., Sahbani A. (2004). Manipulation Planning with Probabilistic Roadmaps. *Int. J. Rob. Res.*, *in press*
27. Thomas F. (1993). The Self-Motion Manifold of the N-bar Mechanism. In: Angeles J. et al. (Eds.) *Computational Kinematics*. Kluwer Academic Publishers, 95–107
28. Trinkle J.C., Milgram R.J. (2002). Complete Path Planning for Closed Kinematic Chains with Spherical Joints. *Int. J. Rob. Res.* **21(9)**, 773–789
29. Yakey J.H. (2000). *Randomized Path Planning for Linkages with Closed Kinematic Chains*. MA Thesis, Iowa State University
30. Yakey J.H., LaValle S.M., Kavraki L.E. (2001). Randomized Path Planning for Linkages with Closed Kinematic Chains. *IEEE Trans. Rob. & Autom.* **17(6)**, 951–958

# Sampling-Based Path Planning on Configuration-Space Costmaps

Léonard Jaillet, Juan Cortés, and Thierry Siméon

**Abstract**—This paper addresses path planning to consider a cost function defined over the configuration space. The proposed planner computes low-cost paths that follow valleys and saddle points of the configuration-space costmap. It combines the exploratory strength of the Rapidly exploring Random Tree (RRT) algorithm with transition tests used in stochastic optimization methods to accept or to reject new potential states. The planner is analyzed and shown to compute low-cost solutions with respect to a path-quality criterion based on the notion of mechanical work. A large set of experimental results is provided to demonstrate the effectiveness of the method. Current limitations and possible extensions are also discussed.

**Index Terms**—Costmap planning, path quality, sampling-based motion planning.

## I. INTRODUCTION

**S**AMPLING-BASED path planning has proven to be an effective framework that is suitable for a large class of problems in domains, such as robotics, manufacturing, computer animation, and computational biology (see [1] and [2] for a survey). These techniques handle complex problems in high-dimensional spaces but usually operate in a binary world, which aims to find out collision-free solutions rather than the optimal path.

Specific path-planning methods have been developed in field robotics for outdoor navigation, where the goal is to find optimal paths according to a cost function, which is usually computed from a model of the terrain. Classical grid-based methods, such as  $A^*$  or  $D^*$  [3] can be used to compute resolution-optimal paths over a costmap. However, compared with sampling-based algorithms, these methods are limited to problems which involve low-dimensional spaces that can be discretized and searched using grid-search techniques.

Some recent works [4]–[8] have tried to bridge the gap between sampling-based planners and grid-based costmap planners. They mainly rely on the Rapidly exploring Random

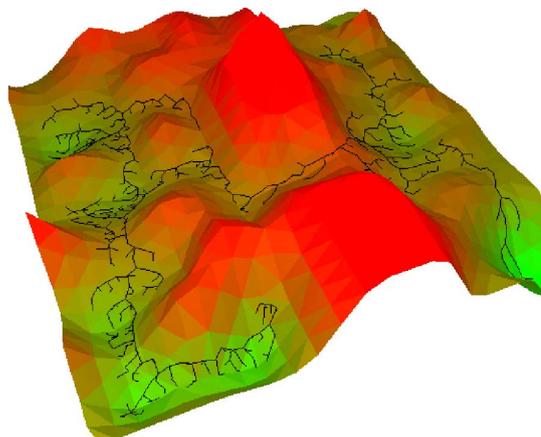


Fig. 1. Transition-based RRT on a 2-D costmap (the elevation corresponds to the costs). The exploration favors the expansion in valleys and saddle points, which connect to low-cost regions.

Tree (RRT) algorithm [9] and are generally focused on specific applications (e.g., real-time problems [7], [10] or statistical learning of feasible paths [8]) in the context of 2-D robot navigation problems.

This paper presents a general algorithm, called Transition-based RRT (T-RRRT),<sup>1</sup> for path planning on configuration-space costmaps. The algorithm considers a user-given cost function defined over the configuration space as an additional input to the standard path-planning problem, and it produces solution paths that are not only feasible (e.g., collision free), but also have a good quality with respect to the input costmap. For instance, the costmap may correspond in outdoor navigation problems to the elevation map of the terrain in order to compute motions that minimize climbing of high-slope regions. In addition, in robotic-manipulation problems, the cost function may be defined from distances to be maximized between the robot and some objects, in order to find high-clearance solution paths. Finally, in computational biology applications, the costmap can be viewed as the energy landscape of the conformational space to be considered for the simulation of low-energy molecular motions.

The proposed algorithm combines the exploratory strength of RRTs with the efficiency of stochastic-optimization methods (e.g., Monte Carlo optimization and simulated annealing) that use transition tests to accept or reject new potential states. The filtering of the transition test relies on the gradient of cost function along the local motion to connect a given state to the RRT tree that results in an expansion biased to follow the valleys and the saddle points of the configuration-space costmap (see Fig. 1). Solution paths computed by T-RRRT fulfill a quality

Manuscript received November 16, 2009; revised March 10, 2010; accepted April 27, 2010. Date of publication June 3, 2010; date of current version August 10, 2010. This paper was recommended for publication by Associate Editor O. Brock and Editor K. Lynch upon evaluation of the reviewers' comments. This work was supported in part by the French National Research Agency under Project "Glucodesign," in part by the European Community under FP7-ICT Project 216239 "DEXMART," and in part by the Spanish Ministry of Science and Innovation under Project DPI2007-60858. The work of L. Jaillet was supported by Consejo Superior de Investigaciones Científicas under the JAE-Doc Fellowship.

L. Jaillet is with the Institut de Robòtica i Informàtica Industrial, Consejo Superior de Investigaciones Científicas, Universitat Politècnica de Catalunya, Barcelona 08028, Spain (e-mail: ljaillet@iri.upc.edu).

J. Cortés and T. Siméon are with CNRS; LAAS, F-31077 Toulouse, France, and also with Université de Toulouse; UPS, INSA, INP, ISAE; LAAS; F-31077 Toulouse, France (e-mail: jcortes@laas.fr; nic@laas.fr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TRO.2010.2049527

<sup>1</sup>The T-RRRT planner was introduced in a shorter version published in [11].

property based on the notion of mechanical work, which is also introduced in this paper as an effective criterion to evaluate path quality for costmap planning.

This paper is organized as follows. After a brief presentation of related work (see Section II), we introduce and discuss the notion of Minimal Work (MW) paths (see Section III). These paths are optimal according to a new criterion called mechanical work that is used to evaluate path quality. Comparison with other existing criteria shows the advantage of this criterion that may be more suitable in many situations, since it yields better paths following the low-cost valleys of the costmap. Additional properties of MW are also presented for a deep understanding of this notion. Section IV describes the T-RRT algorithm and explains the methods for self-tuning of parameters and for the expansion-rate control. Section V shows how T-RRT implicitly computes MW paths and discusses its probabilistic completeness. An experimental validation of the planner is conducted in Section VI. The overall efficacy of T-RRT is shown on different problems and positively compared with other existing techniques [4], [6]. Section VI also analyzes the influence of the intrinsic parameters of the algorithm on the overall performance, and results indicate that no specific tuning is actually needed. Section VII presents some extensions of T-RRT, and finally, conclusions are outlined in Section VIII.

## II. RELATED WORK

Early potential field methods [12], as well as their combination with strategies to escape local minima, e.g., the randomized planner described in [13], rely on some numerical field defined over the configuration space that may be viewed as a specific kind of costmap. Note, however, that the artificial potential field of these methods is only defined as a way to plan collision-free paths, without considering path optimality. Thus, these methods do not address the problem considered here to compute low-cost, feasible paths from an arbitrarily complex costmap given as input to the planner.

Recent sampling-based planners have proven to be very effective to find feasible solutions that can be locally optimized in a postprocessing stage. Local path-optimization methods, such as the shortcut algorithm [14] are generally used to improve path quality with respect to simple criteria, e.g., path length, clearance, or a combination of both [15]. These smoothing methods only aim to locally improve a solution path, as opposed to the global exploration algorithm proposed in this paper. Moreover, their extension to arbitrary cost functions has not yet been addressed, and the resulting efficacy of such an extension remains to be further evaluated.

Only few papers consider sampling-based path planning on arbitrary cost spaces. An adaptation of the RRT-connect planner is used to find low-cost paths for rough terrain navigation in [4]. The idea is to keep new configurations only if their cost is under a given threshold, first initialized to a low value, and then iteratively increased during the search. One limitation of this technique comes from the nondecreasing threshold, which limits the efficiency of low-cost search to the vicinity of the root nodes. To overcome this issue, the extension proposed in [5]

considers multiple RRTs grown from randomly sampled root configurations. However, this solution still expects an appropriate number of initial samples in order to get enough low-cost seeds among the space. Moreover, it requires a manual tuning of the parameter that controls the cost threshold-growth rate. This tuning is highly problem-dependent.

In [6], the heuristically guided RRT (hRRT) biases the search by using a quality measure based on the integral of the cost along the path from the root node and an estimation of the optimal cost to the goal. Such an approach, inspired from graph-search techniques, can also be found in the context of real-time applications [7], [10] and statistical learning of feasible paths [8]. However, with these techniques, the estimated *cost to goal* is heuristic, and tends to bias the search straight toward the goal at the expense of lower quality solution paths. Moreover, the aforementioned methods have only been demonstrated on simple low-dimensional examples with discrete cost states (invalid, low cost, and high cost, respectively). Their scalability and performance for problems which involve complex cost spaces in higher dimensions have yet to be established.

The T-RRT algorithm introduced in the following is inspired by Monte Carlo optimization techniques. Developed in order to find global optima in very complex spaces [16], these techniques introduce randomness as a means to avoid local minima traps. Many variants have been developed (e.g., random walk and simulated annealing [17]). The basic exploration process typically relies on successive transition tests using the Metropolis criterion (see Section IV-B). Note also that the probabilistic conformational roadmap [18] developed to explore molecular-energy landscapes in computational biology applications integrates a similar transition test in the Probabilistic RoadMap (PRM) framework [19].

## III. MINIMAL WORK PATHS

This section introduces the mechanical work criterion to measure path quality in a space that is mapped by a given cost function. Paths that are optimal according to this criterion are called Minimal Work (MW) paths. The T-RRT algorithm presented in the next section tends to produce such MW paths, as shown by the theoretical analysis and the experimental results in Sections V and VI, respectively. First, we introduce the notion of MW paths and illustrate how this criterion generally yields more natural solution paths (i.e., paths following well the low-cost valleys of the costmap) compared with other existing path-quality measures.

### A. Notation

Let us consider a system with a configuration space  $\mathcal{C}$ , possibly constrained by “binary” obstacle regions. Let us also consider a cost function  $c : \mathcal{C} \rightarrow \mathbb{R}_+^*$  mapping this space, i.e., a cost  $c(q) > 0$  can be computed for each  $q \in \mathcal{C}$ . This cost function  $c$  is assumed to be continuous. A path  $\mathcal{P}$  of length  $l$  is represented by a unit-speed parametric function<sup>2</sup>  $\tau : [0, l] \rightarrow \mathcal{C}$

<sup>2</sup>This representation assumes that the parameterized curve that represents the path is regular, which simplifies the notation.

with  $\tau(s) = q_s \in \mathcal{P}$ . Then, we define the parametric cost function  $v : [0, l] \rightarrow \mathbb{R}_+^*$  of a path as  $v(s) = c \circ \tau(s) = c(q_s)$ .

### B. Classical Path-Quality Measures

Several criteria have been proposed to evaluate the quality of a path from its parametric cost function, e.g., maximal cost [5], average cost [5]–[7], or costs sum over discrete path configurations [5], [8] (as a way to approximate the integral of the cost along the path). The maximal cost criterion is the most limited one, since it only relies on a point value of the parameterized cost function. The average cost can also be misleading, since it does not account for path length (a path that involves many detours inside a low-cost region will have an average cost smaller than a path that goes straight through this region). Thus, the integral of the cost along a path appears to be a more reliable criterion. It is mathematically defined as follows:

$$S(\mathcal{P}) = \int_0^l v(s) ds.$$

A discrete approximation of the integral leads to

$$S(\mathcal{P}) \sim \frac{l}{n} \sum_{k=0}^{n-1} v(s_k), \text{ with } s_k = \left(\frac{k}{n-1}\right)l.$$

In what follows, optimal paths according to the Integral of the Cost criterion are called IC paths. The next section introduces an alternative way to measure path quality based on the notion of mechanical work. This alternative technique will then be compared with IC criterion in Section III-D.

### C. Mechanical Work of a Path

The key idea is that positive variations of the parametric cost function can be seen as forces acting against motion, and thus, producing mechanical work. We propose to use this loss of “energy” induced by the mechanical work to measure the quality of a path. In the case of negative variation of costs, the system loses no energy. Then, a small penalty proportional to the distance is added in order to favor shortest paths of equal mechanical energy. Based on this principle, the mechanical work of a path is defined as follows:

$$W(\mathcal{P}) = \int_{\mathcal{P}^+} \frac{\partial v}{\partial s} ds + \epsilon \int_{\mathcal{P}} ds \quad (1)$$

where  $\mathcal{P}^+$  represents the portions of path with positive slopes (i.e., where the parametric cost function is strictly increasing), and  $\epsilon$  is assumed to be very low compared with cost values.

The continuous expression of  $W$  in (1) can be transformed into a discrete formulation expressed from the local extrema values along the path:

$$\begin{aligned} W(\mathcal{P}) &= \sum_i (v(\beta_i) - v(\alpha_i)) + \epsilon l \\ &= \sum_i \Delta v_i^+ + \epsilon l \end{aligned} \quad (2)$$

where  $\alpha_i$  and  $\beta_i$  are consecutive minima and maxima of the costs along the paths, and  $\Delta v_i^+ = v(\beta_i) - v(\alpha_i)$  are the positive

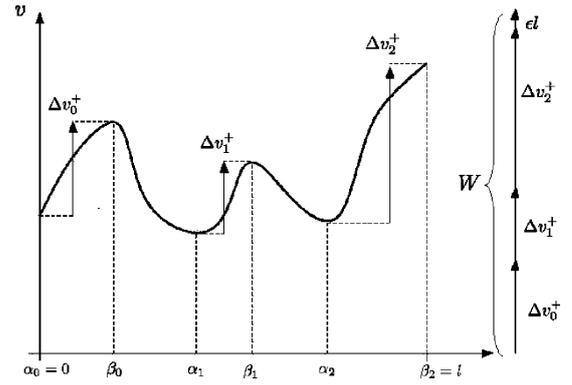


Fig. 2. Decomposition of a path into portions of monotonic cost variation.  $\alpha_i$  and  $\beta_i$  correspond to local minima and maxima, respectively. (Right) Mechanical work is the sum of positive cost variations between consecutive extrema plus a small value  $\epsilon l$  proportional to the path length.

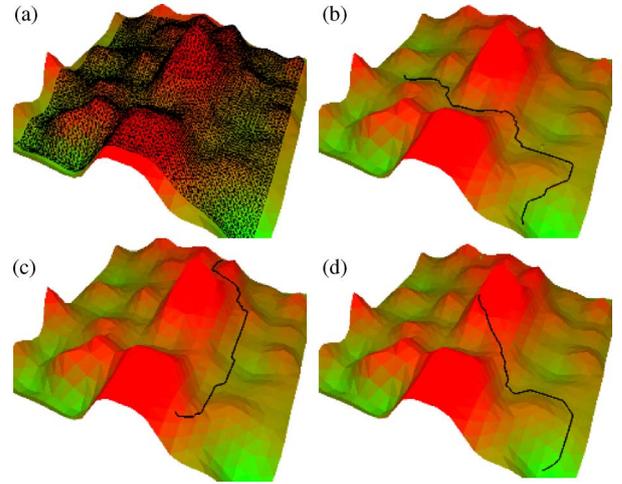


Fig. 3. MW solution paths. (a) Paths are computed by using the  $A^*$  algorithm within a 2-D grid discretizing the space. The examples illustrate (b) down-to-top, (c) top-to-top, and (d) top-to-down queries, respectively.

variations between two consecutive extrema (see Fig. 2). The mechanical work of a path is simply obtained by summing up the positive differences between extrema of its parameterized cost function and adding  $\epsilon l$  in order to favor shortest paths among the ones having equally positive cost variations. Paths that minimize the mechanical work for a given query are called MW paths.

Fig. 3 shows examples of MW paths for several queries on a 2-D hilly costmap. The paths were computed by using a standard  $A^*$  search performed on a grid discretizing the 2-D landscape. As one can see, the shapes of the MW paths appear to be suitable in the sense that they follow as much as possible the low-cost regions of the space. In order to better state the pertinence of the mechanical work criterion, we first compare it with the IC criterion discussed in the next section. Then, we state some interesting properties of the MW criterion in Section III-E.

### D. Minimal Work versus Integral of the Cost

This section compares the optimal solutions for the integral of the cost  $S$  (IC paths) and for the mechanical work  $W$  (MW paths) on representative cost spaces.

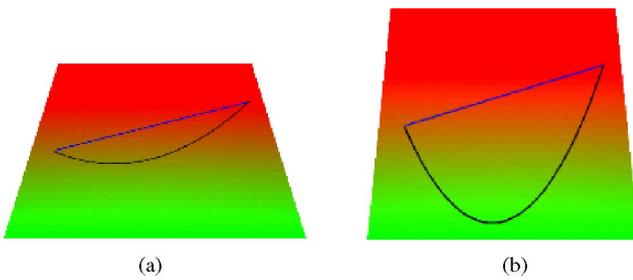


Fig. 4. Straight-line MW path (blue) and two different IC paths for two different inclinations of the plane that represent the cost function.

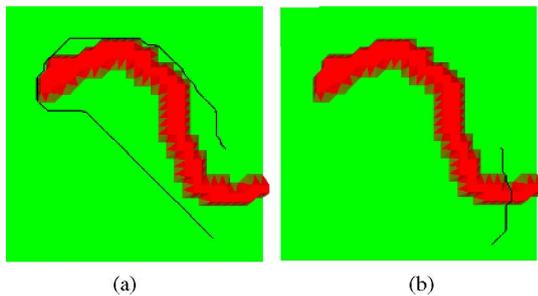


Fig. 5. High-cost barrier problem. (a) MW path. (b) IC path.

1) *Constant Slope*: Let us first consider the example of a planar landscape with a constant slope. In this simple case, IC solutions can be numerically characterized from calculus of variations. As shown in Fig. 4, the solutions obtained for two different slopes show that IC paths (in black) are not intuitive and, moreover, depend on the plane inclination. In contrast, in both cases, the MW path is the trivial straight-line path (in blue). Indeed, the cost of MW paths is always lower bounded by the cost variation between the initial and final configurations. In situations for which the query configurations can be connected through a set of paths having a monotonic cost variation (as for the specific case of a constant slope landscape), the MW path will be the shortest one among the set of minimal cost variation paths. This yields a straight-line solution for the planar slope example.

2) *High-Cost Barrier*: This example corresponds to a flat cost surface with a high-cost barrier that should be preferably avoided (see Fig. 5). In this case, the MW path is the shortest path to get around the barrier [see Fig. 5(a)], while the IC solution is a direct path that crosses the barrier [see Fig. 5(b)]. This example highlights another possibly negative feature of the integral of the cost criterion that may favor undesirable paths with short high-cost portions.

3) *Hilly Costmap*: In this more complex example, solution paths have to go through a saddle point to link the query configurations located at two opposite corners of the hilly landscape (see Fig. 6). The MW path makes necessary detours to follow low-cost valleys of the space. In contrast, the IC solution prefers shortest paths at the expense of local high costs [circled in blue in Fig. 6(b)]. As can be seen in the parameterized cost functions of the two kinds of optimal paths (see Fig. 7), the cost profile of the IC path (red) is globally much higher than the one of the

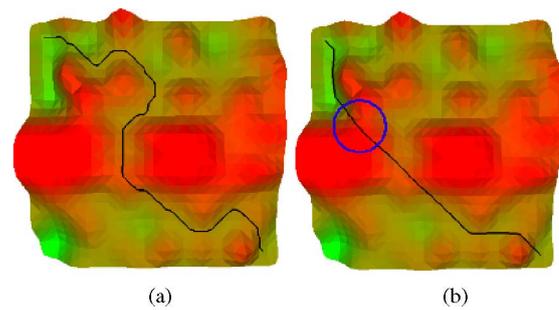


Fig. 6. Hilly costmap problem. (a) MW path. (b) IC path.

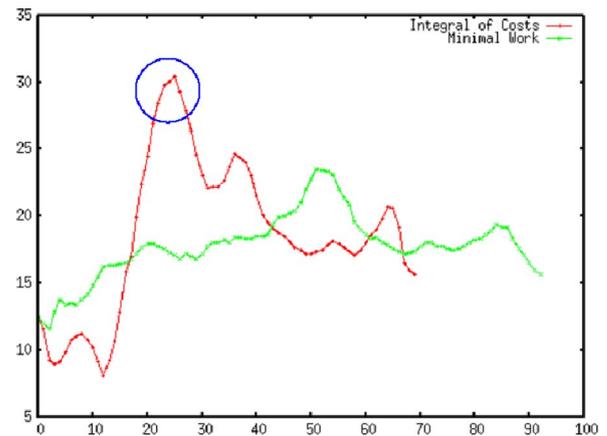


Fig. 7. Parameterized cost functions of the MW path (green) and the IC path (red) shown in Fig. 6(a) and (b).

TABLE I  
MW AND IC OPTIMAL PATHS OF FIG. 6 COMPARED WITH  
A REFERENCE STRAIGHT-LINE SOLUTION

	$Length$	$C_{ave}$	$C_{max}$	$S$	$W$
MW path	186	<b>17.9</b>	<b>23.6</b>	3324	<b>15.9</b>
IC path	139	18.6	30.4	2592	32.5
Str. Line	127	22.6	39.6	2877	41.0

MW path (green). This observation is particularly true when the IC path goes through the high-cost region avoided by the MW path.

Finally, Table I compares the costs of the two solutions with respect to various path-quality measures. It shows that both the average and maximum costs are better for the MW path than for the IC path. Indeed, the IC path characteristics are intermediate between the ones of the MW path and of a simple straight-line path, not biased to avoid high-cost regions.

These results highlight some interesting features of the MW criterion. Compared with IC paths, MW paths avoid steep variations of the cost function. This may be particularly important in applications, such as outdoor navigation (to avoid high-slope motions), or computational biology (to minimize the crossing of high-energy barriers). Besides, in the presented cases, MW paths look more natural. In the next section, we present some additional properties of MW paths for a deep understanding of the mechanical work criterion.

### E. Minimal Work Path Properties

1) *Negative Slopes Minimization*: Section III-C states that to minimize the mechanical work means to minimize the amount of positive cost variations. A first property is that, between two given configurations, it also leads to minimize the negative cost variations. Indeed, the total amount of cost variations along the path can be expressed as follows:

$$v(l) - v(0) = \sum_i \Delta v_i^+ + \sum_j \Delta v_j^-$$

where  $\Delta v_i^+$  and  $\Delta v_j^-$  are the intervals of positive and negative cost variation, respectively. Using (2), we obtain

$$W(\mathcal{P}) = v(l) - v(0) + \epsilon l + \sum_j |\Delta v_j^-|. \quad (3)$$

Because  $v(l)$  and  $v(0)$  are constants and  $\epsilon l$  is small relative to cost values, (3) states that to minimize  $W$  is equivalent to minimize the last term in the right-hand side of the equation, that is, the total amount of negative cost variations.

2) *Cost Variations Minimization*: Since the MW path  $\mathcal{P}$  minimizes both positive and negative cost variations,  $\mathcal{P}$  is indeed the path that minimizes any cost variation between two given configurations. Let  $V(\mathcal{P})$  be a function that sums positive and negative variations

$$V(\mathcal{P}) = \sum_i \Delta v_i^+ + \sum_j |\Delta v_j^-|.$$

Using (2) and (3), we get

$$V(\mathcal{P}) = 2W(\mathcal{P}) - (v(l) - v(0) + 2\epsilon l). \quad (4)$$

Thus, the ordering of the paths remains the same regardless of the criterion ( $V$  or  $W$ ), which, indeed, means that they are equivalent. However, we will keep the formulation of MW path, since this notion facilitates the analysis of the T-RRT algorithm.

3) *Reversibility of Minimal Work Paths*: Let  ${}^{-1}\mathcal{P}$  be the reverse path of  $\mathcal{P}$ . Since the parametric cost functions  $v$  and  ${}^{-1}v$  have opposed variations, i.e.,  $\Delta {}^{-1}v^+ = |\Delta v^-|$ , we have

$$W({}^{-1}\mathcal{P}) = \sum_j |\Delta v_j^-| + \epsilon l$$

and using (3), we get

$$W({}^{-1}\mathcal{P}) = W(\mathcal{P}) + v(0) - v(l). \quad (5)$$

Consequently, the mechanical work of a path is equal to the mechanical work of its inverse, except for a constant. This property allows us to speak about the MW path between two configurations with no need to orient the path.

## IV. TRANSITION-BASED RAPIDLY EXPLORING RANDOM TREE

### A. Main Algorithm

The T-RRT algorithm combines the advantages of two methods. First, it benefits from the exploratory strength of RRT-like algorithms, which result from their expansion bias toward large Voronoi regions of the space. Additionally, it integrates features of stochastic optimization methods developed to compute global

---

### Algorithm 1: Transition-based RRT

---

```

input   : the configuration space  $\mathcal{C}$ ;
           : the cost function  $c : \mathcal{C} \rightarrow \mathbb{R}_+^*$ ;
           : the root  $q_{\text{init}}$  and the goal  $q_{\text{goal}}$ ;

output  : the tree  $\mathcal{T}$ ;

begin
   $\mathcal{T} \leftarrow \text{InitTree}(q_{\text{init}})$ ;
  while not StopCondition( $\mathcal{T}$ ,  $q_{\text{goal}}$ ) do
     $q_{\text{rand}} \leftarrow \text{SampleConf}(\mathcal{C})$ ;
     $q_{\text{near}} \leftarrow \text{NearestNeighbor}(q_{\text{rand}}, \mathcal{T})$ ;
     $q_{\text{new}} \leftarrow \text{Extend}(\mathcal{T}, q_{\text{rand}}, q_{\text{near}})$ ;
    if  $q_{\text{new}} \neq \text{NULL}$ 
      and TransitionTest( $c(q_{\text{near}})$ ,  $c(q_{\text{new}})$ ,  $d_{\text{near-new}}$ )
      and MinExpandControl( $\mathcal{T}$ ,  $q_{\text{near}}$ ,  $q_{\text{rand}}$ ) then
        AddNewNode( $\mathcal{T}$ ,  $q_{\text{new}}$ );
        AddNewEdge( $\mathcal{T}$ ,  $q_{\text{near}}$ ,  $q_{\text{new}}$ );
  end

```

---

minima in complex spaces: It uses transition tests to accept or reject potential states.

Algorithm 1 shows the pseudocode of the T-RRT planner. Similar to the *Extend* version of the basic RRT algorithm [20], a randomly sampled configuration  $q_{\text{rand}}$  is used to determine both the nearest tree node  $q_{\text{near}}$  to be extended and the extension direction. The extension from  $q_{\text{near}}$  is performed toward  $q_{\text{rand}}$  with an increment step  $\delta$ . In the case of T-RRT,  $\delta$  has to be small enough to avoid cost picks to be missed by the linear interpolation between  $q_{\text{near}}$  and  $q_{\text{new}}$ . This stage also integrates collision detections in the presence of “binary obstacles.” Thus, if the new portion of path leads to a collision, a null configuration is returned and the extension fails, independently of the associated costs. This extension process ensures the bias toward unexplored free regions of the space. The goal of the second stage is to filter irrelevant configurations regarding the search of low-cost paths before inserting  $q_{\text{new}}$  in the tree. Such filtering is performed by the *TransitionTest* function. It relies on the Metropolis criterion commonly used in stochastic-optimization methods. This test integrates a self-tuning technique in order to automatically control its filtering strength and, thus, to ensure continuous growth of the tree. Finally, the *MinExpandControl* function forces the planner to maintain a minimal rate of expansion toward unexplored regions of the space and avoids possible blocking situations during the search. The following sections detail the *TransitionTest* and *MinExpandControl* functions.

### B. Transition Test

The *TransitionTest* function is presented in Algorithm 2. First, configurations with a higher cost than the maximum cost threshold  $c_{\text{max}}$  are filtered. The probability of acceptance of a new configuration is defined by comparing its cost  $c_j$  relatively to the cost  $c_i$  of its parent in the tree. This test is based on the Metropolis criterion initially introduced in statistical physics and molecular modeling. The transition probability

**Algorithm 2:** TransitionTest( $c_i, c_j, d_{ij}$ )

---

```

begin
  nFail = GetCurrentNFail();
  if  $c_j > c_{max}$  then return False;
  if  $c_j < c_i$  then return True;
   $p = \exp\left(\frac{-(c_j - c_i)/d_{ij}}{K \cdot T}\right)$ ;
  if Rand(0, 1) < p then
     $T = T/\alpha$ ;
    nFail = 0;
    return True;
  else
    if nFail > nFailmax then
       $T = T \cdot \alpha$ ;
      nFail = 0;
    else
      nFail = nFail + 1;
    return False;
end

```

---

$p_{ij}$  is defined as follows:

$$p_{ij} = \begin{cases} \exp\left(-\frac{\Delta c_{ij}}{KT}\right), & \text{if } \Delta c_{ij} > 0 \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

where we have the following:

- 1)  $\Delta c_{ij} = (c_j - c_i)/d_{ij}$  is the slope of the cost, i.e., the cost variation divided by the distances between the configurations.<sup>3</sup>
- 2)  $K$  is a constant value used to normalize the expression. It is based on the order of magnitude of the considered costs.  $K$  is taken as the average cost of the query configurations, since they are the only cost values known at the beginning of the search process.
- 3)  $T$  is a parameter called *temperature* that is used to control the difficulty level of transition tests, as further explained in the following. Note that the term temperature is employed in analogy with methods in statistical physics, but in our case, it does not have any physical meaning.

Using this transition probability, downhill transitions are automatically accepted, whereas for uphill transitions, the chance of acceptance decreases exponentially with the cost increment.

1) *Temperature Parameter:*  $T$  is a key parameter of the algorithm, since it defines the level of difficulty of a transition for a given cost increment. Low temperatures limit the expansion to slightly positive slopes. In contrast, higher temperatures enable to climb the steeper slopes. Within methods that involve the Metropolis criterion, the temperature is usually kept constant (e.g., Monte Carlo search) or decreases gradually as the search progresses (e.g., simulated annealing). In our algorithm, this parameter is dynamically tuned according to the information acquired during the exploration.

2) *Adaptive Tuning:* The TransitionTest function performs an adaptive tuning of the temperature during the search

<sup>3</sup>Contrarily to classical Monte Carlo methods, the cost variation is normalized by the distance to the previous state, since this distance is not necessarily constant.

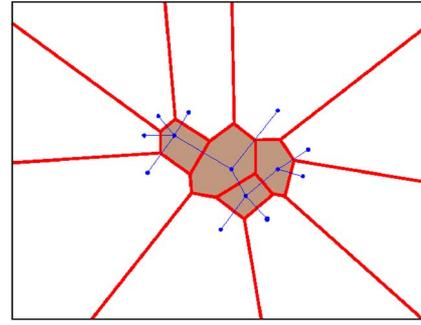


Fig. 8. Frontier nodes (in white regions) have a Voronoi region bounded by the space limits. On the contrary, the Voronoi region of nonfrontier nodes is bounded by the Voronoi region of other nodes (in brown/gray regions).

process (second stage of Algorithm 2). At the initialization,  $T$  is set to a very low value (e.g.,  $10^{-6}$ ) in order to only authorize very easy positive slopes (and negative ones). Then, during the exploration, the number  $nFail$  of consecutive times the Metropolis criterion discards a configuration is recorded and used for temperature tuning. When the T-RRT search reaches a maximal number of rejections  $nFail_{max}$ , the temperature is multiplied by a given factor  $\alpha$ . Each time an uphill transition test succeeds, the temperature is divided by the same factor  $\alpha$ . Thus, the temperature automatically adapts itself, such that an extension that corresponds to a positive cost variation is performed in average every  $nFail_{max}$  times. The influence of parameters  $\alpha$  and  $nFail_{max}$  is analyzed in Section VI.

### C. Minimal Expansion Control

The adaptive temperature tuning introduced earlier ensures a given success rate of positive slope transitions. A possible side effect may appear when the tree expansion toward unexplored regions remains slow, and the new nodes contribute only to refine already explored regions. We discuss in the following this issue and explain how the MinExpandControl function overcomes this problem.

1) *Exploration Versus Refinement:* The behavior of the RRT expansion can be explained by distinguishing two types of nodes [21]. Frontier nodes are the external nodes of the tree with a Voronoi region bounded by the space limits, whereas nonfrontier nodes are the internal ones, whose Voronoi region is entirely bounded by the Voronoi region of the other nodes (see Fig. 8). Thus, the extension of a frontier node tends to explore new regions of the space, and the extension of a nonfrontier node only refines the existing tree. The problem of unbalanced refinement and exploration modes was addressed in [21] and [22] for standard RRTs. However, for T-RRT, the interaction between these two kinds of extensions is more subtle than for the basic RRT. Indeed, situations occur where the temperature is stabilized by new nonfrontier nodes, which refine the tree in easier regions of the space; however, the expansion toward new regions requires the development of frontier nodes. Fig. 9(a) illustrates this issue with an example of a tree, whose expansion has been slowed down by the too-frequent insertion of nonfrontier nodes.

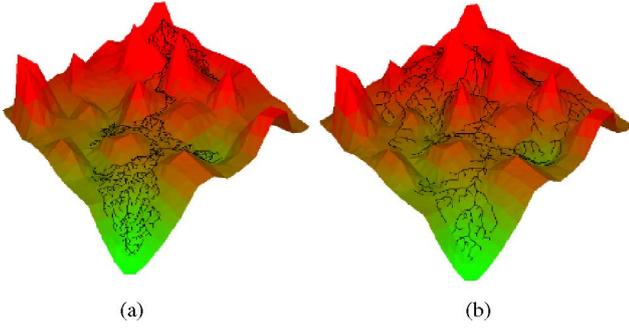


Fig. 9. Impact of the minimal expansion control on the T-RRT algorithm. (a) Without control, the insertion of nonfrontier nodes tends to slow down the exploration by decreasing the temperature. (b) With control, the planner is forced to keep to explore new regions of the space.

---

**Algorithm 3:** `MinExpandControl( $\mathcal{T}$ ,  $q_{near}$ ,  $q_{rand}$ )`


---

```

begin
  if Distance( $q_{near}$ ,  $q_{rand}$ ) >  $\delta$  then
    UpdateNbNodeTree( $\mathcal{T}$ );
    return True;
  else
    if  $\frac{NbRefineNodeTree(\mathcal{T})+1}{NbNodeTree(\mathcal{T})+1} > \rho$  then return False;
    else
      UpdateNbRefineNodeTree( $\mathcal{T}$ );
      UpdateNbNodeTree( $\mathcal{T}$ );
      return True;
end
    
```

---

Fig. 9(b) shows the tree obtained by using the minimal expansion control detailed in the following.

2) *Minimal Exploration Rate:* The proposed solution is to force the planner to explore new regions by controlling the ratio between exploration and refinement steps. Note that as long as the tree coverage remains limited compared with the size of the space, nonfrontier nodes have a Voronoi region that is much smaller than the one of frontier nodes. Hence, extension steps can be estimated as refinements or expansions, which depend on the distance between  $q_{near}$  and  $q_{rand}$ . For a large distance value,  $q_{near}$  has greater chances to be a frontier node, whereas a small distance value corresponds most probably to the case of a non-frontier node extension. The control of minimal exploration rate is performed by the `MinExpandControl` function presented in Algorithm 3. If the distance  $q_{near} - q_{rand}$  is greater than the expansion step  $\delta$ ,  $q_{new}$  is considered to participate in the tree expansion, and it is inserted in the data structure. Otherwise,  $q_{new}$  is considered to participate in the tree refinement. The configuration is not inserted in the tree if it makes the ratio of nonfrontier nodes be greater than a given maximal value  $\rho$ . The influence of this parameter is further discussed in Section VI.

## V. THEORETICAL ANALYSIS OF T-RRT

### A. T-RRT and Minimal Work Path

This section analyzes the relationship between T-RRT and the notion of MW path introduced in Section III. An important property is obtained first for the simplified case of a discrete

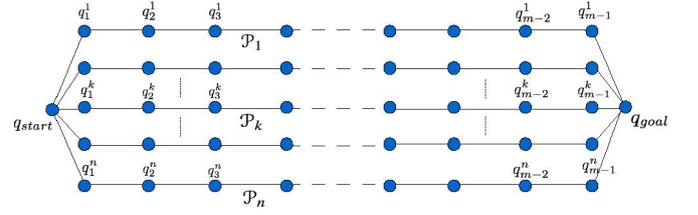


Fig. 10. Case of  $n$  equal-length paths. With T-RRT, the branches with the lowest mechanical work have the highest chances to reach the goal first.

search process. Then, we discuss the extension of this result to the general case of the T-RRT search.

1) *Simplified Case:* Let us consider a path search within a discrete set of  $n$  equal length possible paths, each one defined by a sequence of  $m$  edges and  $m + 1$  nodes (see Fig. 10). Using a T-RRT scheme, each expansion of a given path requires the path to be selected and the associated transition test to succeed. Thus, the probability  $P_k$  of a given path  $\mathcal{P}_k$  to be entirely developed in  $m$  iterations is equal to

$$P_k = \prod_{i \in [1, m]} e p_i^k = \prod_{i \in [1, m]} ({}^s p_i^k) ({}^t p_i^k)$$

where  ${}^e p$  denotes the probability for a given node to be *extended*,  ${}^s p$  is the probability to be *selected*, and  ${}^t p$  is the probability to have an accepted *transition*. In addition, we assume that the paths have equal chances of being extended at each step (i.e., the node-selection process is not biased by Voronoi regions), i.e.,

$$P_k = \frac{1}{n^m} \prod_{i \in [1, m]} {}^t p_i^k.$$

If the transition probability depends only on the transition tests (i.e., the `MinExpandControl` is omitted), we get

$$P_k = \frac{1}{n^m} \prod_j e^{(-\Delta v_j^{k+})/KT_j}.$$

Moreover, if we assume that the temperature remains constant during the expansion, we have

$$P_k = \frac{1}{n^m} e^{1/KT} e^{-\sum_j \Delta v_j^{k+}}$$

where  $\Delta v_j^{k+}$  are summed over the positive variations of cost along the path  $k$ . Finally, since  $\epsilon l$  is negligible in the mechanical work expression, we get

$$P_k = \frac{1}{n^m} e^{1/KT} e^{-W(\mathcal{P}_k)}. \quad (7)$$

Since  $(1/n^m) e^{1/KT}$  is the same for all the paths, we obtain an important property for this simplified version: The paths with the lowest mechanical work have the highest probability of reaching the goal first.

2) *General Case:* One first assumption made in the analysis earlier is that each branch has an equal chance of being chosen for the expansion. In practice, the various paths developed by the T-RRT algorithm (from the root node to each leaf) are not spatially independent. Each branch expansion tends to increase

its global Voronoi region and, thus, increases the chance for its nodes to be selected at the next iteration. This process reinforces the extension of the paths with the most favorable mechanical work and increases the convergence of the planner toward lower cost solutions.

The simplified version also assumed that the temperature is constant. This parameter affects each path in the same way. Thus, we can argue that the property remains valid, even when  $T$  varies during the search.

Finally, whereas the aforementioned property is established for the discrete case of equal-length solution paths, T-RRT search is performed among an infinite number of variable-length paths. Since shortest paths require less expansion steps to connect the queries, it is not possible to guarantee that paths of lower cost have always better chances to reach the goal first. However, as one can see from (7), the mechanical work of a path affects exponentially its chances of success. This reveals how strongly the T-RRT exploration is implicitly biased toward solution paths of low mechanical work.

### B. Probabilistic Completeness

The T-RRT algorithm is a probabilistically complete planner [19]. This property is directly inherited from the probabilistic completeness of the RRT planner (see [9, Sec. IV]). The only difference is that in the present case, the extension steps can be rejected because of the transition tests, even in the case of a convex, open,  $n$ -dimensional subset of an  $n$ -dimensional configuration space. However, we argue that the success probability of the transitions is always strictly positive, since the cost function takes finite values in this subset, and thus, the cost variations are bounded. As a result, the planner converges eventually toward an entire coverage of the considered subset, and the transition tests affect only the convergence rate of the algorithm.

## VI. EXPERIMENTAL RESULTS

A large set of experiments has been conducted to evaluate the performance of the planner. First, the general behavior of the method is presented on various problems. Second, its performance is compared with that of the existing methods to highlight the good quality of the T-RRT solutions. Finally, we investigate the influence of some intrinsic parameters on the overall efficacy of the method. All the algorithms have been implemented within the path-planning software *Move3D* [23]. The performance results summarized in the tables are values averaged over ten runs.

### A. General Performance

A variety of problems are proposed to illustrate the generality of the method. The examples vary not only in the geometrical complexity and the configuration space dimensionality but in the nature of the cost function as well. Two settings of T-RRT are considered: A *greedy* version of the planner referred to as T-RRT<sub>g</sub> that takes  $nFail_{max} = 10$ , and a *tempered* version, which is referred to as T-RRT<sub>t</sub>, with  $nFail_{max} = 100$ . The latter leads to higher quality solution paths but is more computationally expensive. We used  $\alpha = 2$  in all the examples. The results obtained

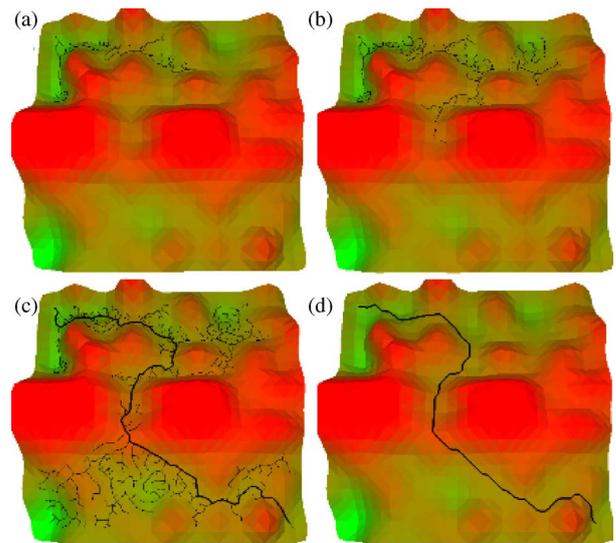


Fig. 11. (a) and (b) Construction process of the transition-based RRT planner. The solution path (c) is close to the optimal one (d) computed from a space discretization.

TABLE II  
COMPARATIVE RESULTS FOR THE COSTMAP PROBLEM

	<i>Length</i>	<i>C<sub>ave</sub></i>	<i>C<sub>max</sub></i>	$S \times 10^{-3}$	<i>W</i>	<i>Time</i>
RRT	148	22.6	41.2	3.3	45.1 (36.9)	0.1
T-RRT <sub>g</sub>	182	18.2	25	3.9	28.0 (19.7)	0.9
T-RRT <sub>t</sub>	214	17.1	23.2	4.0	23.1 (16.9)	11.0
MW path	186	17.9	23.6	3.3	15.9	-
IC path	139	18.6	39.6	2.6	41.0	-
Thresh.	157-192	17.8-20.9	23.7-32.7	3.3-3.4	27-30 (21-24)	0.3-329
hRRT	155	20.9	33.3	3.2	41.4 (32.9)	0.6

with the basic RRT planner are given as references. The tables also present comparative results with two existing cost-based methods that will be discussed later.

The first set of experiments is performed on the 2-D cost space shown in Fig. 1. In this example, the solution paths have to go through a saddle point to link the query configurations located at two opposite corners of the landscape. Fig. 11 shows snapshots of the exploration tree and the solution path found [see Fig. 11(c)], which is close to the optimal one [see Fig. 11(d)]. Table II presents the characteristics of the paths obtained with each planner.<sup>4</sup> It also provides values for the MW and IC optimal paths (computed with an  $A^*$  search within a  $128 \times 128$  grid discretizing the landscape).

The mechanical work of solutions obtained by the different methods is reported in the  $W$  column. The numbers in parentheses integrate the effect of some local smoothing of the solution path with a simple procedure based on the shortcut algorithm [14]. As one can see from Table II, the mechanical work of the reference RRT path is almost three times higher than for the optimal MW solution, and smoothing does not successfully get close to the optimal value (36.9 versus 15.9). In comparison, the mechanical work of the path obtained with the tempered version of T-RRT is only 45% higher than for the MW path, and it

<sup>4</sup>In the case of RRT, since there is no obstacle in the scene, connection attempts to the goal are only performed when  $d(q_{new}, q_{goal}) < 15\delta$  to avoid getting a trivial straight-line solution.

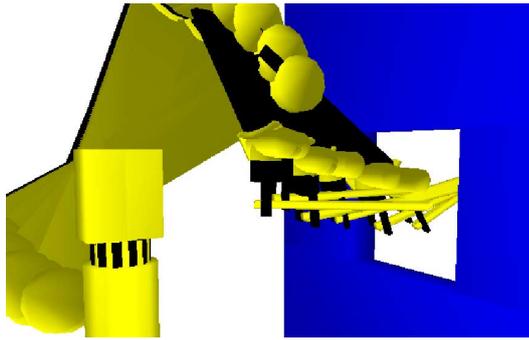


Fig. 12. Stick-extraction problem. A 6-DOF manipulator arm has to extract a stick from a hole. The T-RRT solution path keeps the stick horizontal to maximize its distance to the obstacles.

TABLE III  
COMPARATIVE RESULTS FOR THE STICK-EXTRACTION PROBLEM

	<i>Length</i>	$C_{ave}$	$C_{max}$	<i>S</i>	<i>W</i>	<i>Time</i>
RRT	55.3	6.6	377	363	1196	1.5
T-RRT <sub>g</sub>	53.9	0.4	1.0	21.8	1.9	7.4
T-RRT <sub>t</sub>	51.5	0.3	0.9	17.4	1.1	32.8
Thresh.	53.2-54.9	0.6-3.2	1.6-40.1	32.0-172	4.2-196	1.6-2.7
hRRT	53.5	5.0	341	265	786	3.0

becomes only 6% higher than the optimal value after smoothing. Most important, the overall shape of the T-RRT solution is very close to the optimal-MW path and follows the same low-cost regions. In addition, note that the relatively slight loss of path quality of the greedy version is compensated by a much smaller computing time (0.9 s versus 11.0 s). Comparative results obtained with other existing costmap planners (Thresh. and hRRT rows in Table II) are discussed in Section VI-B.

In the next experiment, a six-degrees-of-freedom (DOF) manipulator arm is carrying a stick in a 3-D workspace with obstacles (see Fig. 12). Here, the goal is to extract the stick from a hole, while keeping the stick as far as possible from the obstacles. Thus, the cost function considered here is the inverse of the distance between the stick and the obstacles. Results are presented in Table III.

The costs of the T-RRT solution paths are considerably lower than the ones of RRT. This shows the effectiveness of the planner to find low-MW paths in higher dimensional spaces. T-RRT solutions tend to keep the stick horizontal during its extraction from the hole in order to remain as far as possible from the obstacles, whereas the basic RRT planner produces erratic paths. Once again, the slight loss of path quality of the greedy version of the T-RRT (1.9 versus 1.1) is compensated by a significant speed up (7.4 s versus 32.8 s).

The third scenario involves the same manipulator arm that carries a sensor with a spherical extremity for the inspection of the surface of a car part. The goal here is to keep the sensor close to the surface of the car part during the motion, in order to satisfy the requirements for the surface following task (see Fig. 13 and Table IV). Note that for such a scenario, where the robot is subject to task-space constraints, specific path-planning schemes also exist (e.g., [24]).

As to be expected, the T-RRT computing time is higher than the one of RRT because to compute a collision-free path with

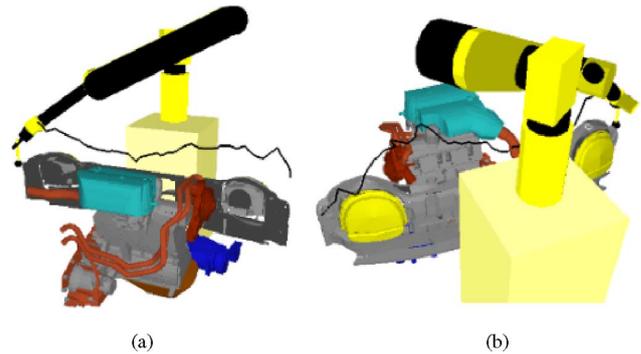


Fig. 13. Car-part-inspection problem. The path for a 6-DOF arm that manipulates a sensor (black sphere), which needs to remain close to the surface during the inspection task is shown.

TABLE IV  
COMPARATIVE RESULTS FOR THE CAR-PART-INSPECTION PROBLEM

	<i>Length</i>	$C_{ave}$	$C_{max}$	$S \times 10^{-3}$	<i>W</i>	<i>Time</i>
RRT	3785	457	875	1730	1434	0.98
T-RRT <sub>g</sub>	3515	24.4	87.5	85.8	400	16.4
T-RRT <sub>t</sub>	3396	4.3	28.1	14.6	187	206
Thresh.	2848-3942	36-186	79-436	116-529	447-635	4.3-19
hRRT	3127	253	452	792	943	18.0

RRT and with no cost consideration is a much easier problem than to obtain a solution that minimizes the distance to the inspected surface. However, with regard to paths quality, the mechanical work of T-RRT<sub>g</sub> and T-RRT<sub>t</sub> are 3.6 times and 7.7 times lower than the one of RRT, respectively. The average and maximal costs reported in Table IV are interesting indicators to get a better idea of the quality of the results, since they correspond directly to the average and maximal distances between the sensor and the part, respectively. For a distance reference, the diameter of the black sphere at the extremity of the sensor is 40 mm. For T-RRT<sub>g</sub>, the maximal cost corresponds approximately to twice this value, whereas the average distance is close to the sensor radius. In the case of T-RRT, solution paths follow the surface of the part so well that the maximal distance never exceeds the size of the sphere and the average one is about one tenth of this diameter.

Finally, the last scenario corresponds to a molecular model shown in Fig. 14. The task is to compute the pathway to extract the ligand (small molecule in red/dark) from the active site located inside a protein. This problem can be seen as a mechanical disassembly path-planning problem for the free-flying ligand [25]. Energetic constraints are translated into geometric ones by considering a steric model of the molecule, and a collision-detection algorithm [26] is applied as a geometric filter that rejects conformations with prohibitively high van der Waals (VdW) energy. The cost function considered for this problem is the inverse of the distance between the ligand and the protein. The interest of this molecular model is to provide a simple way to quantify the quality of the computed solution path. The ligand-free space can be simply dilated by shrinking the atoms radii. The results reported in Table V correspond to both RRT and T-RRT algorithms applied on the shrunk model shown in Fig. 14(b) (25% of VdW radii).

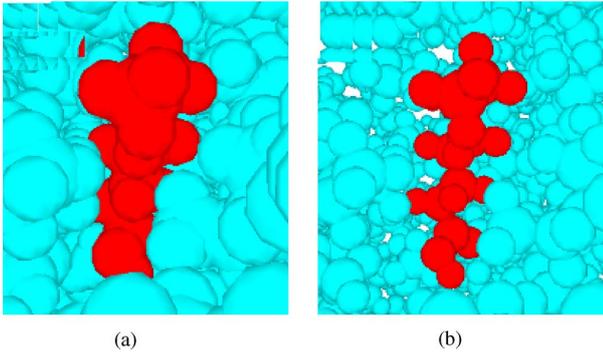


Fig. 14. Two representations of the same ligand–protein “disassembly” problem, with different VdW radii. (a) Maximal radius. (b) Shrunk radius. The goal is to compute paths that maximize the clearance and, thus, remain valid for large VdW radii.

TABLE V  
COMPARATIVE RESULTS FOR THE LIGAND–PROTEIN PROBLEM

	<i>Length</i>	<i>C<sub>ave</sub></i>	<i>C<sub>max</sub></i>	<i>S</i>	<i>W</i>	<i>Time</i>	<i>VdW<sub>m</sub></i>
RRT	59	14	2236	826	471	0.7	25
T-RRT <sub>g</sub>	62	0.4	1.0	22	1.25	8.39	65
T-RRT <sub>t</sub>	64	0.3	0.9	22	1.0	426	69
Thresh.	58–62	4.7	1.1–3.7	25–42	1.4–4.3	4.8–73.5	34–62
hRRT	59	8.3	733	490	353	15.2	26

The T-RRT solution paths have a much lower cost compared with the one computed by RRT. The higher clearance of the T-RRTs solutions are also quantified by the maximal VdW ratio indicated in the last row of the table. This maximal ratio was obtained by testing solution paths by increasing the VdW radii until a collision was detected between the ligand and the protein. While no growing is possible for the RRT solution, the T-RRT paths (computed with a 25% ratio) remain valid up to 65% and 69% growing, depending on the variant. These values are close to the maximal radius that allows the ligand to exit (80%). The high clearance of T-RRT paths reflects their good quality with respect to the considered distance-based cost.

### B. T-RRT versus Existing Methods

T-RRT has been compared with two existing cost-based planners: the maximal *Threshold* technique proposed by Ettl and Bleuler [4] and the *hRRT* of Urmsion and Simmons [6]. Results obtained for the set of experiments with these planners are reported in the last two rows of Tables II–V. In the case of the *threshold* method, results are highly sensitive to the threshold-growing speed, and thus, reported data correspond to the extremal values obtained when varying this parameter in the range (1–100).

With regard to the mechanical work criterion, results show that T-RRT<sub>t</sub> provides significantly better solutions than existing methods in all tests. Remarkably, T-RRT solutions are also better with respect to the IC criterion in the three more difficult problems, which involve a 6-D cost space.

The overall bad performance of the *hRRT* method is due to the strong bias introduced by the heuristic that steers the exploration toward the goal, resulting in a poor exploratory ability, which makes it unable to circumvent high-cost regions and find higher quality paths. Comparatively, the threshold technique can pro-

TABLE VI  
INFLUENCE OF THE  $\alpha$  AND  $nFail_{max}$  PARAMETERS

Hilly costmap						
$\alpha$	Time			W		
	$nFail_{max}$			$nFail_{max}$		
	10	100	1000	10	100	1000
2	<b>0.9</b>	<b>11.0</b>	121	<b>28.0</b>	<b>23.1</b>	23.4
10	0.7	7.1	93.4	32.1	26.6	24.4
50	0.7	7.0	85.5	32.0	28.7	24.8
Stick extraction						
$\alpha$	Time			W		
	$nFail_{max}$			$nFail_{max}$		
	10	100	1000	10	100	1000
2	<b>7.4</b>	<b>32.8</b>	226	<b>1.9</b>	<b>1.1</b>	1.1
10	8.2	31.0	218	5.5	2.9	2.4
50	7.0	29.8	226	3.5	3.3	1.2
Car part inspection						
$\alpha$	Time			W		
	$nFail_{max}$			$nFail_{max}$		
	10	100	1000	10	100	1000
2	<b>16.4</b>	<b>206</b>	2012	<b>400</b>	<b>187</b>	166
10	14.3	171	1697	439	188	165
50	15.9	167	1583	507	212	203

TABLE VII  
INFLUENCE OF THE  $\rho$  PARAMETER

$\rho$	Hilly costmap		Stick extraction		Car part inspection	
	Time	W	Time	W	Time	W
1	420	19.6	30.3	1.1	201	192
1/2	16.7	23.7	33.4	1.2	198	202
1/10	<b>11.0</b>	<b>23.1</b>	<b>32.8</b>	<b>1.1</b>	<b>206</b>	<b>187</b>
1/100	9.7	23.8	32.0	1.2	269	263

vide solution paths whose quality is close to the one of T-RRT<sub>g</sub>, but its performance is highly sensitive to the parameter that regulates the variation speed of the threshold. Depending on the value of this parameter, the running time increases up to 1000 times for the costmap problem, the mechanical work increases up to 47-fold for the stick-extraction problem, and both the running time and the mechanical work are notably affected by the threshold-speed value for the car-part-inspection problem. Furthermore, this sensitive parameter is problem-dependent and has to be tuned for each application, whereas T-RRT parameters remain robust to problem changes, as shown next.

### C. Influence of Intrinsic Parameters

We now analyze the influence of the main parameters of the T-RRT algorithm. Experiments are performed on three problems that correspond to three different types of cost functions: the hilly costmap (see Fig. 11), the stick-extraction problem (see Fig. 12), and the car-part-inspection problem (see Fig. 13). The results are presented in Tables VI and VII. Bold values are the default settings used in previous tests.

1) *Temperature Variation Control*:  $nFail_{max}$  and  $\alpha$  are the two parameters that control the derivative of the temperature, and hence, the selectivity of the transition test (as explained in Section IV-B).

Table VI shows that  $nFail_{max}$  is an important parameter that determines the appropriate balance between time performance and solution-path quality. In the costmap problem, when  $nFail_{max}$  is increased by a factor of ten, the running time also increases 9–13-fold. Its influence on the runtime performance is less direct on the two manipulator problems (due to the

additional cost of collision checking), even though the tendency is the same. Finally, note that higher values of  $nFail_{max}$  improve path quality but only up to a point: The quality increases when  $nFail_{max}$  varies from 10 to 100 but remains approximately constant from 100 to 1000.

With regard to the  $\alpha$  parameter, results show that it affects only slightly the behavior of the algorithm, even if higher values tend to increase the time performance, while decreasing the path quality. Overall, values  $nFail_{max} = 100$  and  $\alpha = 2$  provide the best results for the three examples and are used as default setting for all tests.

2) *Expansion Versus Refinement Control*: Table VII presents result for various values of the  $\rho$  parameter used in the `MinExpandControl` function to set the maximal ratio of refinement nodes.

In the first line of the table,  $\rho = 1$  means that the `MinExpandControl` function is inactive. The results for the 2-D hilly costmap highlight the importance of this function, the computing time being much higher when  $\rho = 1$ . This example illustrates cases in which the refinement process slows down the exploration by decreasing the temperature. This effect is less visible in the two other examples, where refinement steps are less likely to happen because of the large size of the space. Results for the other settings (i.e.,  $\rho \neq 1$ ) are quite similar, which means that  $\rho$  does not require to be tuned precisely. In all experiments, the default setting  $\rho = 1/10$  appears to be a good compromise between computing time and path quality.

## VII. EXTENSIONS

### A. Bidirectional T-RRT

Similar to the bidirectional version of the RRT planner [9], a bidirectional T-RRT can be envisaged. However, a naive approach that would use the same transition test for both trees would lead to poor quality solutions. It would tend to create paths with consecutive downhill and uphill cost variations, which correspond to branches expanded from the *init*-tree and *goal*-tree, respectively, and may fail to find a more flat solution path of lower MW cost. A better alternative, using the property of Section III-E2, which states that the MW paths minimize any cost variations, is to modify transition tests in order to filter both positive and negative cost variations when expanding the two trees. This can be achieved easily by replacing the transition probability  $p_{ij}$  of (6) by the expression  $p_{ij} = \exp(-(|\Delta c_{ij}|)/KT)$ . Preliminary results show that this approach performs well in problems where positive and negative cost variations for the best cost paths are globally of the same amplitude. However, in problems where the profile of the cost between query nodes is asymmetric, it turns out to reject too many configurations during the transition test, which degrades the performance. In this case, a method based on a more sophisticated transition test should be designed.

### B. Toward a Greedy Anytime T-RRT

In this section, we discuss a possible extension of T-RRT for performance improvement in tricky situations, such as the one illustrated in Fig. 15. In this example, the large low-cost region has to be fully explored 1) for determining the need to cross

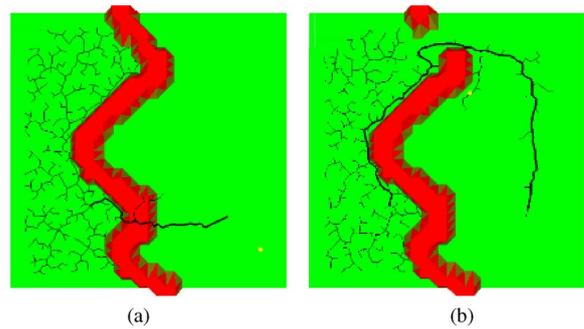


Fig. 15. Tricky problem for T-RRT. A large low-cost region has to be explored before deciding to cross the high-cost barrier. (a) Useless in or (b) leading to a better solution.

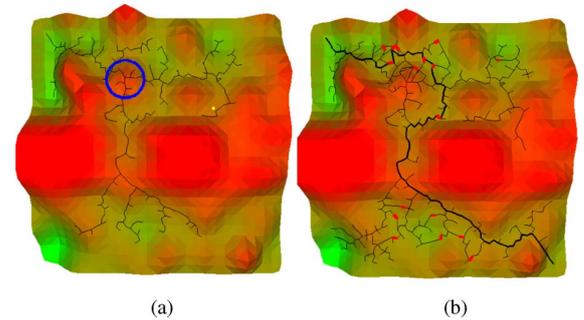


Fig. 16. (a) Initial tree built using a greedy T-RRT version. (b) Addition of cycles (in red) leads to higher quality paths.

the higher cost barrier or 2) to discover the low-cost passage that yields a better solution. In both cases, the greedy T-RRT<sub>g</sub> version may rapidly cross the barrier and, thus, speed up the computation compared with the tempered T-RRT<sub>t</sub>. However, it may miss the preferred detour path in problem 2) for which a longer exploration is needed to find the passage. To keep the performance of an aggressive exploration, while avoiding this issue, we propose to combine the greedy version of the planner with a cycle-addition mechanism. The idea is to create cycles in the tree when good paths initially missed during the search are discovered afterward. The idea has been tested by using the technique described in [27] for cycle addition. Fig. 16 shows an initial tree built using a greedy version of T-RRT that goes through a medium cost region [circled in blue on Fig. 16(a)] that could have been avoided. The addition of cycles provides alternative paths and yields higher quality solutions [see Fig. 16(b)].

## VIII. CONCLUSION AND FUTURE WORK

We have presented a sampling-based algorithm to compute paths in problems which involve high-dimensional cost spaces. The proposed method combines the exploratory strength of RRTs, with the efficiency of stochastic-optimization methods. It integrates an adaptive mechanism that helps to ensure a good performance for a large set of problems.

The notion of MW path has been proposed to quantify the quality of solution paths. By design, the proposed T-RRT algorithm computes paths that tend to satisfy such a quality property.

A large set of experiments were performed to show the efficacy of the T-RRT planner.

Experimental results have shown that the planner is general enough to be applied, at least, to 6-D spaces constrained by obstacles. Future work concerns the application of T-RRT to new classes of problems, such as the integration of human-robot interaction constraints within path planning or the exploration of energy landscapes in computational biology problems. Extensions discussed in the previous section also need to be further explored for performance improvement. Furthermore, another direction is to incorporate in the planner other methods inspired by Monte Carlo optimization techniques, such as stochastic tunneling [28] or parallel tempering [29]. Finally, it would be interesting to test our approach on benchmark problems of the stochastic optimization community, since T-RRT could be used as a generic optimization tool and, in principle, applied to any metric cost space.

#### REFERENCES

- [1] H. Choset, K. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L. Kavraki, and S. Thrun, *Principles of Robot Motion: Theory, Algorithms, and Implementations*. Cambridge, MA: MIT Press, 2005.
- [2] S. LaValle, *Planning Algorithms*. New York: Cambridge Univ. Press, 2006.
- [3] A. Stentz, "Optimal and efficient path planning for partially-known environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1994, pp. 3310–3317.
- [4] A. Ettlín and H. Bleuler, "Rough-terrain robot motion planning based on obstacleness," in *Proc. Int. Conf. Control, Autom., Robot. Vis.*, 2006, pp. 1–6.
- [5] A. Ettlín and H. Bleuler, "Randomised rough-terrain robot motion planning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2006, pp. 5798–5803.
- [6] C. Urmson and R. Simmons, "Approaches for heuristically biasing RRT growth," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2003, pp. 1178–1183.
- [7] D. Ferguson and A. Stentz, "Anytime RRTs," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2006, pp. 5369–5375.
- [8] R. Diankov and J. Kuffner, "Randomized statistical path planning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 1–6.
- [9] J. Kuffner and S. LaValle, "RRT-connect: An efficient approach to single-query path planning," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2000, pp. 995–1001.
- [10] J. Lee, C. Pippin, and T. Balch, "Cost based planning with RRT in outdoor environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 684–689.
- [11] L. Jaillet, J. Cortés, and T. Siméon, "Transition-based RRT for path planning in continuous cost spaces," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 2145–2150.
- [12] J.-C. Latombe, *Robot Motion Planning*. Boston, MA: Kluwer, 1991.
- [13] J. Barraquand and J.-C. Latombe, "A Monte-Carlo algorithm for path planning with many degrees of freedom," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1990, pp. 1712–1717.
- [14] S. Sekhavat, P. Svestka, and M. Overmars, "Multi-level path planning for nonholonomic robots using semi-holonomic subsystems," *Int. J. Robot. Res.*, vol. 17, no. 8, pp. 840–857, 1998.
- [15] R. Geraerts and M. H. Overmars, "Creating high-quality paths for motion planning," *Int. J. Robot. Res.*, vol. 26, pp. 845–863, 2007.
- [16] J. Spall, *Introduction to Stochastic Search and Optimization*. New York: Wiley, 2003.
- [17] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
- [18] M. Apaydin, A. Singh, D. Brutlag, and J.-C. Latombe, "Capturing molecular energy landscapes with probabilistic conformational roadmaps," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2001, pp. 932–939.
- [19] L. Kavraki, P. Svestka, J.-C. Latombe, and M. Overmars, "Probabilistic roadmaps for path planning in high-dimensional configuration spaces," *IEEE Trans. Robot. Autom.*, vol. 12, no. 4, pp. 566–580, Aug. 1996.
- [20] S. LaValle, "Rapidly-exploring random trees: A new tool for path planning," Comput. Sci. Dept., Iowa State Univ., Ames, IA, Tech. Rep. TR 98-11, 1998.
- [21] A. Yershova, L. Jaillet, T. Siméon, and S. LaValle, "Dynamic-domain RRTs: Efficient exploration by controlling the sampling domain," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2005, pp. 3867–3872.
- [22] L. Jaillet, A. Yershova, S. LaValle, and T. Siméon, "Adaptive tuning of the sampling domain for dynamic-domain RRTs," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2005, pp. 4086–4091.
- [23] T. Siméon, J.-P. Laumond, and F. Lamiraux, "Move3D: A generic platform for path planning," in *Proc. IEEE Int. Symp. Assem. Task Plann.*, 2001, pp. 25–30.
- [24] M. Stilman, "Task constrained motion planning in robot joint space," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 3074–3081.
- [25] J. Cortés, L. Jaillet, and T. Siméon, "Disassembly path planning for complex articulated objects," *IEEE Trans. Robot. Autom.*, vol. 24, no. 2, pp. 475–481, Apr. 2008.
- [26] V. Ruiz de Angulo, J. Cortés, and T. Siméon, "BioCD: An efficient algorithm for self-collision and distance computation between highly articulated molecular models," in *Robotics: Science and Systems*, S. T. and G. Sukhatme, S. Schaal, and O. Brock, Eds. Cambridge, MA: MIT Press, 2005, pp. 6–11.
- [27] D. Nieuwenhuisen and M. Overmars, "Useful cycles in probabilistic roadmap graphs," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2004, pp. 446–452.
- [28] K. Hamacher and W. Wenzel, "Scaling behavior of stochastic minimization algorithms in a perfect funnel landscape," *Phys. Rev. E*, vol. 59, no. 1, pp. 938–941, Jan. 1999.
- [29] D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications, and new perspectives," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 3910–3916, 2005.



**Léonard Jaillet** received the engineering degree in mechanical engineering from the Institut Supérieur de Mécanique de Paris, Paris, France, and the Ph.D. degree in robotics from the University of Toulouse, Toulouse, France, in 2001 and 2005, respectively.

Since 2008, he has been a Postdoctoral Fellow with the Institut de Robòtica i Informàtica Industrial, Spanish National Research Council, Barcelona, Spain. His current research interests include motion planning for complex robotic systems and molecular simulations for structural biology.



**Juan Cortés** received the engineering degree in control and robotics from the Universidad de Zaragoza, Zaragoza, Spain, in 2000 and the Ph.D. degree in robotics from the Institut National Polytechnique de Toulouse, Toulouse, France, in 2003.

Since 2004, he is a Researcher with the LAAS, Centre National de la Recherche Scientifique, Toulouse. His current research interests include the development of algorithms for computing and analyzing the motion of complex systems in robotics and structural biology.

Dr. Cortés is a Co-Chair of the IEEE-RAS TC on Algorithms for Planning and Control of Robot Motion.



**Thierry Siméon** received the engineering degree in computer science from the Institut National des Sciences Appliquées, Toulouse, France, and the Ph.D. degree in robotics from the University of Toulouse in 1985 and 1989, respectively.

Since 1990, he has been with the LAAS, Centre National de la Recherche Scientifique, Toulouse. He is the author or coauthor of more than 100 papers. He was a Co-Founder of the LAAS spin-off Kineo Company. His current research interests include robot motion planning and applications to structural bioinformatics.

Dr. Siméon is currently an Associate Editor of the IEEE TRANSACTIONS ON ROBOTICS.

---

**Thierry Siméon**  
**Jean-Paul Laumond**  
**Juan Cortés**  
**Anis Sahbani**

LAAS-CNRS  
Toulouse, France

# Manipulation Planning with Probabilistic Roadmaps

## Abstract

*This paper deals with motion planning for robots manipulating movable objects among obstacles. We propose a general manipulation planning approach capable of addressing continuous sets for modeling both the possible grasps and the stable placements of the movable object, rather than discrete sets generally assumed by the previous approaches. The proposed algorithm relies on a topological property that characterizes the existence of solutions in the subspace of configurations where the robot grasps the object placed at a stable position. It allows us to devise a manipulation planner that captures in a probabilistic roadmap the connectivity of sub-dimensional manifolds of the composite configuration space. Experiments conducted with the planner in simulated environments demonstrate its efficacy to solve complex manipulation problems.*

**KEY WORDS**—manipulation task planning, path planning, probabilistic roadmaps

## 1. Introduction

Manipulation planning concerns the automatic generation of robot motion sequences allowing the manipulation of movable objects among obstacles. The presence of movable objects, i.e., objects that can only move when grasped by a robot, leads to a more general and computationally complex version of the classical motion planning problem (Latombe 1991). Indeed, the robot has the ability to modify the structure of its configuration space depending on how the movable object is grasped and where it is released in the environment. Also, movable objects cannot move by themselves; either they are transported by robots or they must lie at some stable placement. Motion planning in this context appears as a constrained instance of the coordinated motion planning problem. The solution of a manipulation planning problem (see, for example, Alami, Siméon, and Laumond 1989; Latombe 1991) consists

in a sequence of subpaths satisfying these motion restrictions. Motions of the robot holding the object at a fixed grasp are called “transfer paths”, and motions of the robot while the object stays at a stable placement are called “transit paths”.

Let us consider the manipulation planning example illustrated by Figure 1. The manipulator arm has to get a movable object (the bar) out of the cage, and place it on the other side of the environment. Solving this problem requires the automatic production of the sequence of transfer/transit paths separated by grasps/ungrasps operations, allowing one extremity of the bar out of the cage; the manipulator can then re-grasp the object by the extremity that was made accessible by the previous motions, perform a transfer path to extract the bar from the cage, and finally reach the specified goal position. In particular, the motion shown on the second image illustrating the solution requires itself four re-grasping operations to obtain a sufficient sliding motion of the bar. This example shows that a manipulation task possibly leads to a complex sequence of motions including several re-grasping operations. A challenging aspect of manipulation planning is to consider the automatic task decomposition into such elementary collision-free motions.

Most existing algorithms (e.g., Ahuactzin, Gupta, and Mazer 1998; Alami, Siméon, and Laumond 1989; Barraquand and Ferbach 1994; Koga and Latombe 1994; Nielsen and Kavraki 2000) assume that a finite set of stable placements and of possible grasps of the movable object are given in the definition of the problem. Consequently, a part of the task decomposition is thus solved by the user since the initial knowledge provided with these finite sets has to contain the grasps and the intermediate placements required to solve the problem. Referring back to the example, getting the bar out of the cage would require a large number of grasps and placements to be given as input data.

In this paper, we describe a general approach based on recent results presented in Siméon et al. (2002) and Sahbani, Cortés, and Siméon (2002). We propose a manipulation planner that automatically generates grasps and intermediate placements solving complicated manipulation problems such as illustrated in Figure 1. The main contribution is the ability to

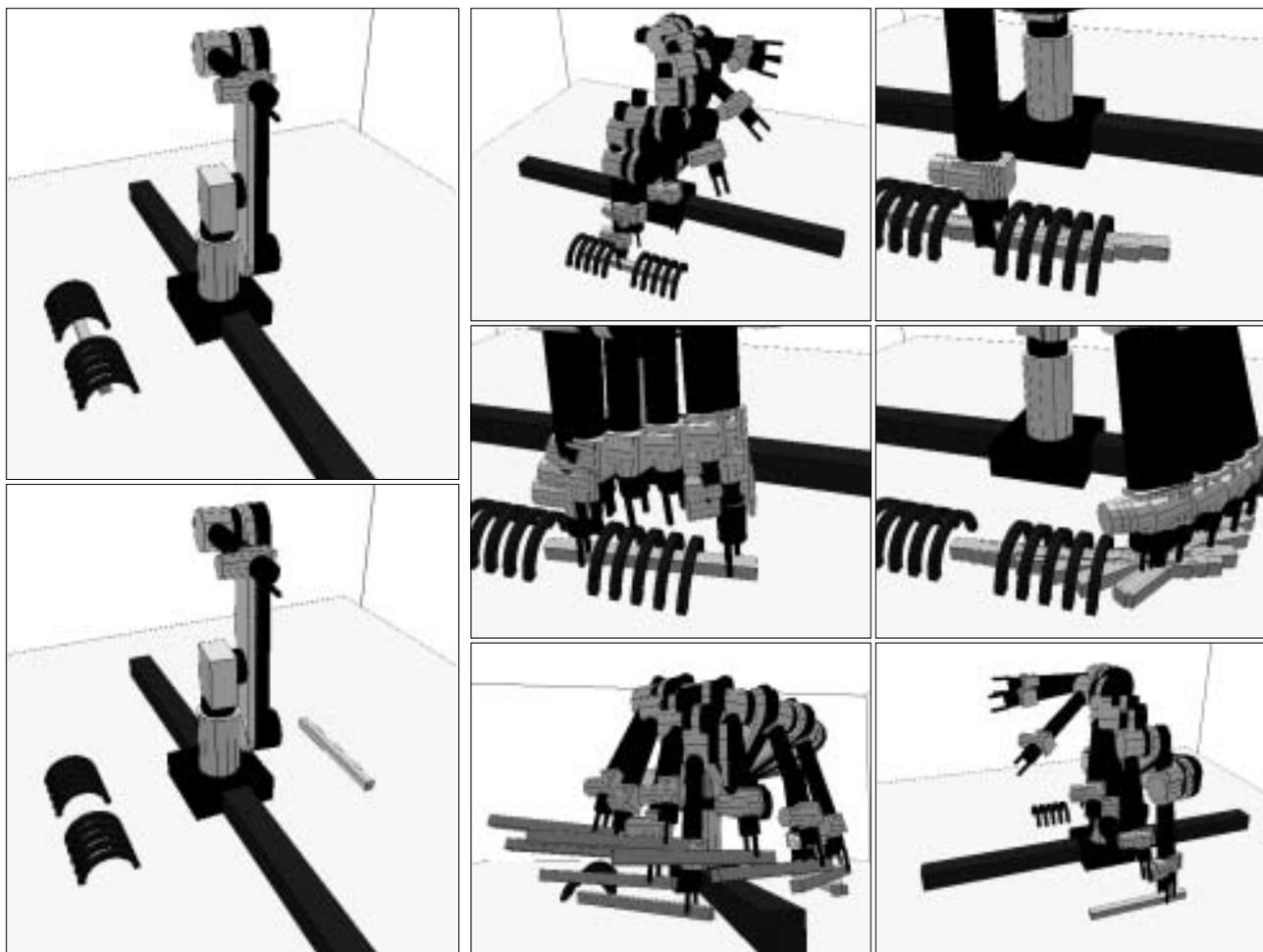


Fig. 1. How can we manipulate the bar from its initial position (top left) to the goal (bottom left)? The solution (right) requires several pick and place operations.

deal with continuous sets in the definition of the manipulation problem, while covering the scope of the previous proposed approaches (Section 2). The approach relies on a topological property first established in Alami, Laumond, and Siméon (1994) and recalled in Section 3. This property allows us to reduce the problem by characterizing the existence of a solution in the lower-dimensional subspace of configurations where the robot grasps the movable object placed at a stable position.

Section 4 describes the proposed approach and shows how the connected components of this subspace can be captured in a probabilistic roadmap (PRM) computed for a virtual closed-chain system. Section 5 details the planning techniques developed to implement the approach. Using the Visibility-PRM algorithm (Siméon, Laumond, and Nissoux 2000) extended to deal with such closed systems (Cortés, Siméon, and Laumond 2002), we first capture the connectivity of the search space

into a small roadmap composed of a low number of connected components (Section 5.1). Connections between these components using transit or transfer motions are then computed by solving a limited number of point-to-point path planning problems (Section 5.2). The details of an implemented planner interleaving both stages in an efficient way are described in Section 5.3. Finally, Sections 6 and 7 presents some experiments and comments on the performance of the planner.

## 2. Related Work

One of the challenging issues of manipulation planning is to integrate the additional difficulty of planning the grasping and re-grasping operations to the path planning problem. This interdependency between path planning and grasp planning was first touched upon by work done in the 1980s for

the development of automatic robot programming systems. In particular, the Handey system (Lozano-Pérez et al. 1992) integrated both planning levels and was capable to plan simple pick-and-place operations including some re-grasping capabilities. The geometric formulation of manipulation planning (Alami, Siméon, and Laumond 1989; Latombe 1991), seen as an instance of the motion planning problem extended by the presence of movable objects, provided a unified framework allowing us to better tackle the interdependency issues between both planning levels.

Motion planning in the presence of movable objects is first addressed as such in Wilfong (1988). In this work, an exact cell decomposition algorithm is proposed for the particular case of a polygonal robot and of one movable object translating in a polygonal workspace, assuming a finite grasp set of the movable object.

The manipulation graph concept is introduced in Alami, Siméon, and Laumond (1989) for the case of one robot and several movable objects manipulated with discrete grasps and placements. In this case, the nodes of the manipulation graph correspond to discrete configurations and the edges are constructed by searching for transfer (or transit) paths between nodes sharing the same grasp (or placement) of the movable object(s). Following this general framework, the approach was implemented for a translating polygon (Alami, Siméon, and Laumond 1989) and a three-degrees-of-freedom (3-DoF) planar manipulator (Laumond and Alami 1989). An exact cell decomposition algorithm is also proposed in Alami, Laumond, and Siméon (1994) for the specific case of a translating polygonal robot capable of manipulating one movable polygon with an infinite set of grasps.

The manipulation planning framework is extended in Koga and Latombe (1992, 1994) to multi-arm manipulation where several robots cooperate to carry a single movable object among obstacles. In this work, the number of legal grasps of the objects is finite and the movable object has to be held at least by one robot at any time during a re-grasp operation. The planner proposed in Koga and Latombe (1994) first plans the motions of the movable object using an adapted version of a randomized potential field planner (Barraquand and Latombe 1991), and then finds the sequence of re-grasp operations of the arms to move the object along the computed path. This planner relies on several simplifications, but it can deal with complex and realistic problems.

Another heuristic planning approach proposed in Barraquand and Ferbach (1994) is to iteratively deform a coordinated path first generated in the composite configuration space using a variational dynamic programming technique that progressively enforces the manipulation constraints.

Variants of the manipulation planning problem have been investigated. In Lynch and Mason (1994), grasping is replaced by pushing and the space of stable pushing directions imposes a set of non-holonomic constraints that introduce some controllability issues to the problem. The heuristic algorithm de-

scribed in Chen and Hwang (1991) considers a problem where all the obstacles can be moved by a circular robot in order to find its way to the goal.

Two other contributions extend recent planning techniques to manipulation planning. In Ahuactzin, Gupta, and Mazer (1998), the Ariadne's Clew algorithm (Bessiere et al. 1993) is applied to a redundant robot manipulating a single object in a three-dimensional (3D) workspace. The method assumes discrete grasps of the movable object; it is, however, capable in realistic situations of dealing with redundant manipulators (Ahuactzin and Gupta 1999) for which each grasp possibly corresponds to an infinite number of robot configurations. Finally, Nielsen and Kavraki (2000) propose a practical manipulation planner based on the extension of the PRM framework (Kavraki and Latombe 1994; Overmars and Švestka 1994). The planner constructs a manipulation graph between discrete configurations; connections are computed using a fuzzy PRM planner that builds a roadmap with edges annotated by a probability of collision-freeness. Computing such roadmaps improves the efficiency of the planner for solving the possibly high number of path planning queries (in changing environments) required to compute the connections.

### 2.1. Contribution

The manipulation planning techniques described above mostly address the discrete instance of the problem. Only the algorithms in Alami, Laumond, and Siméon (1994) and Ahuactzin, Gupta, and Mazer (1998) consider more difficult instances for which the nodes of the manipulation graph (i.e., the places where the connections between the feasible transit and transfer paths have to be searched) correspond to a collection of submanifolds of the composite configuration space, as opposed to discrete configurations. Such manifolds arise when considering infinite grasps and continuous placements of the object. This continuous formulation is only addressed in Alami, Laumond, and Siméon (1994) for the specific case of a translating robot in a polygonal world. Manifolds also arise in Ahuactzin, Gupta, and Mazer (1998) because of the redundancy of the robot although the planner assumes a set of predefined discrete grasps.

In this paper, we propose a general approach for dealing with such continuous settings of the manipulation planning problem. Our planning approach considers continuous placements and grasps, and it is also able to handle redundant robots. It relies on a structuring of the search space allowing us to efficiently capture the connectivity of the submanifolds in a probabilistic roadmap computed for virtual closed-chain mechanisms. The resulting planner is general and practical for solving complicated manipulation planning problems in constrained 3D environments.

For example, one can describe the set of stable placements by constraining the movable object to be placed on top of some horizontal faces of the static obstacles. Such placement

constraints define a 3D submanifold of the object's configuration space (two translations in the horizontal plane and one rotation around the vertical axis). Also, one can consider sets of continuous grasping domains such that the jaws of a parallel gripper have a contact with two given faces of the object. Such grasp constraints also define a 3D domain (two translations parallel to the grasped faces and one rotation around the axis perpendicular to the faces).

### 3. Manipulation Planning

#### 3.1. Notations

We consider a 3D workspace with a robot  $\mathcal{R}$  and a movable object  $\mathcal{M}$  moving among static obstacles. The robot has  $n$  DoF and  $\mathcal{M}$  is a rigid object with six DoF that can only move when it is grasped by the robot. Let  $CS_{rob}$  and  $CS_{obj}$  be the configuration spaces of the robot and the object, respectively. The composite configuration space of the system is  $CS = CS_{rob} \times CS_{obj}$  and we call  $CS_{free}$  the subset in  $CS$  of all admissible configurations, i.e., configurations where the moving bodies do not intersect together or with the static obstacles. The domain in  $CS$  corresponding to valid placements of  $\mathcal{M}$  (i.e., stable placements where the object can rest when ungrasped by the robot) is denoted by  $CP$ . The domain in  $CS$  corresponding to valid grasps configurations of  $\mathcal{M}$  by the robot  $\mathcal{R}$  is denoted by  $CG$ . Both  $CP$  and  $CG$  are sub-dimensional manifolds in  $CS$ .

#### 3.2. Manipulation Constraints

A solution to a manipulation planning problem corresponds to a constrained path in  $CS_{free}$ . Such a solution path is an alternate sequence of two types of subpaths verifying the specific constraints of the manipulation problem, and separated by grasp/ungrasp operations.

- *Transit paths* where the robot moves alone while the object  $\mathcal{M}$  remains stationary in a stable position. The configuration parameters of  $\mathcal{M}$  remain constant along a transit path. Such motions allow us to place the robot at a configuration where it can grasp the object. They are also involved when changing the grasp of the object. Transit paths lie in  $CP$ . However, a path in  $CP$  is not generally a transit path since such a path has to belong to the submanifold corresponding to a fixed placement of  $\mathcal{M}$ . Transit paths induce a foliation<sup>1</sup> of  $CP$  (Figure 2(a)).

1. A *foliation* (Ito 1987) of an  $n$ -dimensional manifold  $M$  is an indexed family  $L_\alpha$  of arc-wise connected  $m$ -dimensional submanifolds ( $m < n$ ), called *leaves* of  $M$ , such that

- $L_\alpha \cap L_{\alpha'} = \emptyset$  if  $\alpha \neq \alpha'$
- $\cup_\alpha L_\alpha = M$
- every point in  $M$  has a local coordinate system such that  $n - m$  coordinates are constant

- *Transfer paths* where the robot moves while holding  $\mathcal{M}$  with the same grasp. Along a transfer path, the configuration of  $\mathcal{M}$  changes according to the grasp mapping induced by the forward kinematics of the robot:  $q_{obj} = \mathcal{G}(q_{rob})$ . Transfer paths lie in  $CG$ . They induce a foliation of  $CG$  (Figure 2(b)).

#### 3.3. Problem

Consider the two sets of constraints defining the stable placements and feasible grasps. A manipulation planning problem is to find a manipulation path (i.e., an alternate sequence of transit and transfer paths) connecting two given configurations  $q_i$  and  $q_f$  in  $CG \cup CP$  (Figure 2(c)). Manipulation planning then consists of searching for transit and transfer paths in a collection of submanifolds corresponding to particular grasps or stable placements of the movable object. Note that the intersection  $CG \cap CP$  between the submanifolds<sup>2</sup> defines the places where transit paths and transfer paths should be connected. The manipulation planning problem appears as a constrained path planning problem inside and between the various connected components of  $CG \cap CP$  (Figure 2(d)).

#### 3.4. Reduction Property

Two foliation structures are defined in  $CG \cap CP$ : the first is induced by the transit paths; the second is induced by the transfer paths. As a consequence, any path lying in a connected component of  $CG \cap CP$  can be transformed into a finite sequence of transit and transfer paths (the proof of this property<sup>3</sup> appears in Alami, Laumond, and Siméon 1994). Therefore, two configurations which are in a same connected component of  $CG \cap CP$  can be connected by a manipulation path.

It is then sufficient to study the  $CG \cap CP$  component's connectivity by transit and transfer paths. Let us consider a transit (or transfer) path whose endpoints belong to two distinct connected components  $(CG \cap CP)_i$  and  $(CG \cap CP)_j$  of  $CG \cap CP$ . From the reduction property above one may deduce that any configuration in  $(CG \cap CP)_i$  can be connected to any configuration in  $(CG \cap CP)_j$  along a manipulation path.

#### 3.5. Manipulation Graph

It is then possible to build a graph  $MG$  whose nodes are the various connected components of  $CG \cap CP$  while an edge between two nodes  $(CG \cap CP)_i$  and  $(CG \cap CP)_j$  indicates the existence of a transit (or transfer) path whose endpoints belong respectively to  $(CG \cap CP)_i$  and  $(CG \cap CP)_j$ .

2. The intersection  $CG \cap CP$  is also a submanifold. Note, however, that  $CG \cup P$  is not a submanifold.

3. Note that this property holds for a single movable object under the hypothesis that the robot does not touch the static obstacles.

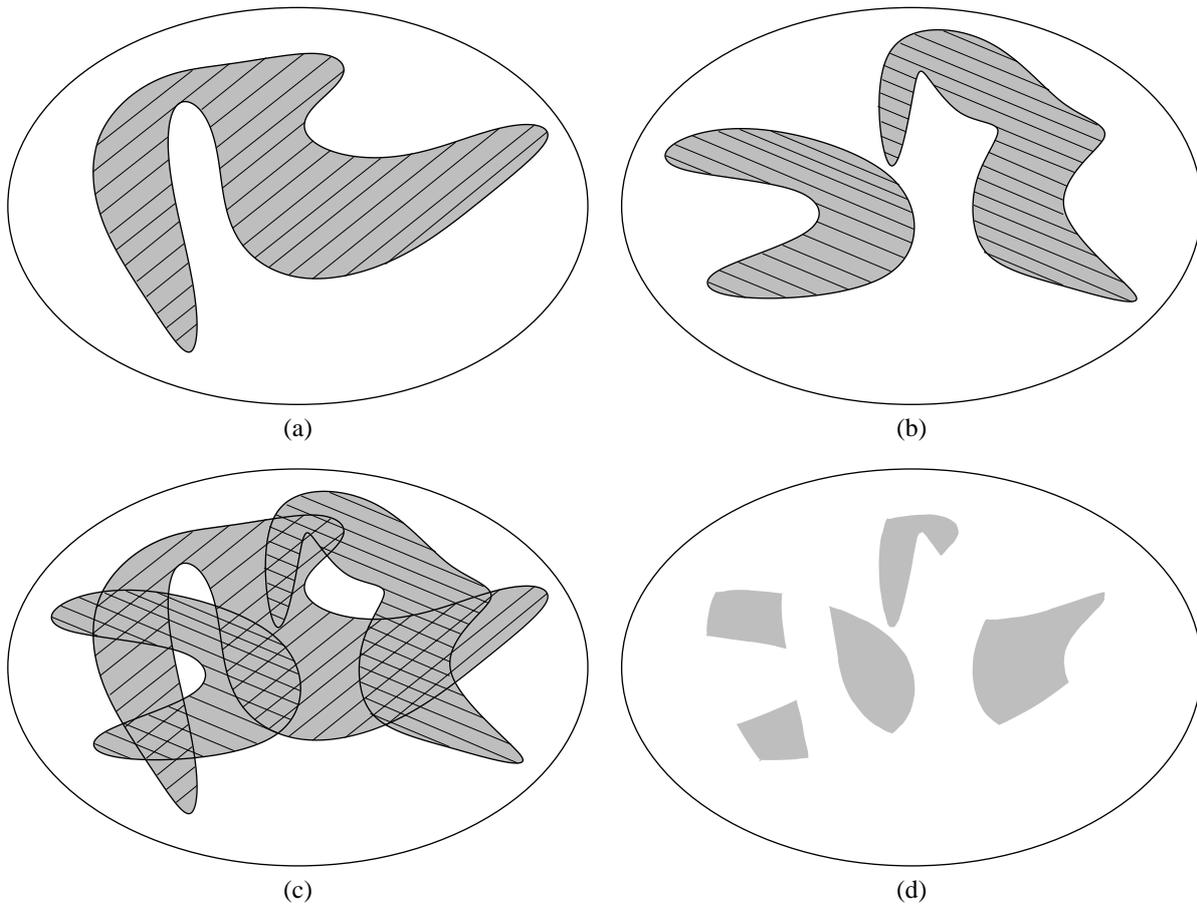


Fig. 2. Moving along transit (transfer) paths induces a foliation of the placement (grasp) space. Both foliations intersect themselves in  $CG \cap CP$ . (a) The placement space  $CP$ ; (b) the grasp space  $CG$ ; (c)  $CG \cup CP$ ; (d)  $CG \cap CP$  has five connected components.

Figure 3 illustrates the graph structure for the example introduced in Figure 2. Examples of manipulation paths are shown in the bottom-left picture;  $q_1$  is not a valid configuration for the manipulation problem (it does not belong to  $CP \cup CG$ ). Configuration  $q_6$  is in  $CG$ ; nevertheless it cannot escape from its leaf in  $CG$ . A manipulation path exists between  $q_3$  and  $q_5$  and between  $q_2$  and  $q_4$ . No manipulation path exists between  $q_5$  and  $q_4$ .

Let  $q_i$  and  $q_f$  be two configurations in  $CG \cup CP$ . There exists a manipulation path between  $q_i$  and  $q_f$  iff there exist two nodes  $(CG \cap CP)_i$  and  $(CG \cap CP)_f$  in  $MG$ , called the manipulation graph, such as:

- there exists a transit (or transfer) path from  $q_i$  to some point in  $(CG \cap CP)_i$  ;
- there exists a transit (or transfer) path from some point in  $(CG \cap CP)_f$  to  $q_f$ ;

- $(CG \cap CP)_i$  and  $(CG \cap CP)_f$  belong to a same connected component of  $MG$ .

### 3.6. Combinatorial Issues

How can we capture the connected components of  $CG \cap CP$ ? How can we capture their adjacency by transit and transfer paths? These are the two key issues in manipulation task planning. All the techniques overviewed above fall into this general framework.

## 4. A General Approach to Manipulation Planning

We now describe our approach for solving manipulation problems in the general setting of continuous grasp and placement constraints. The proposed approach relies on the structure of

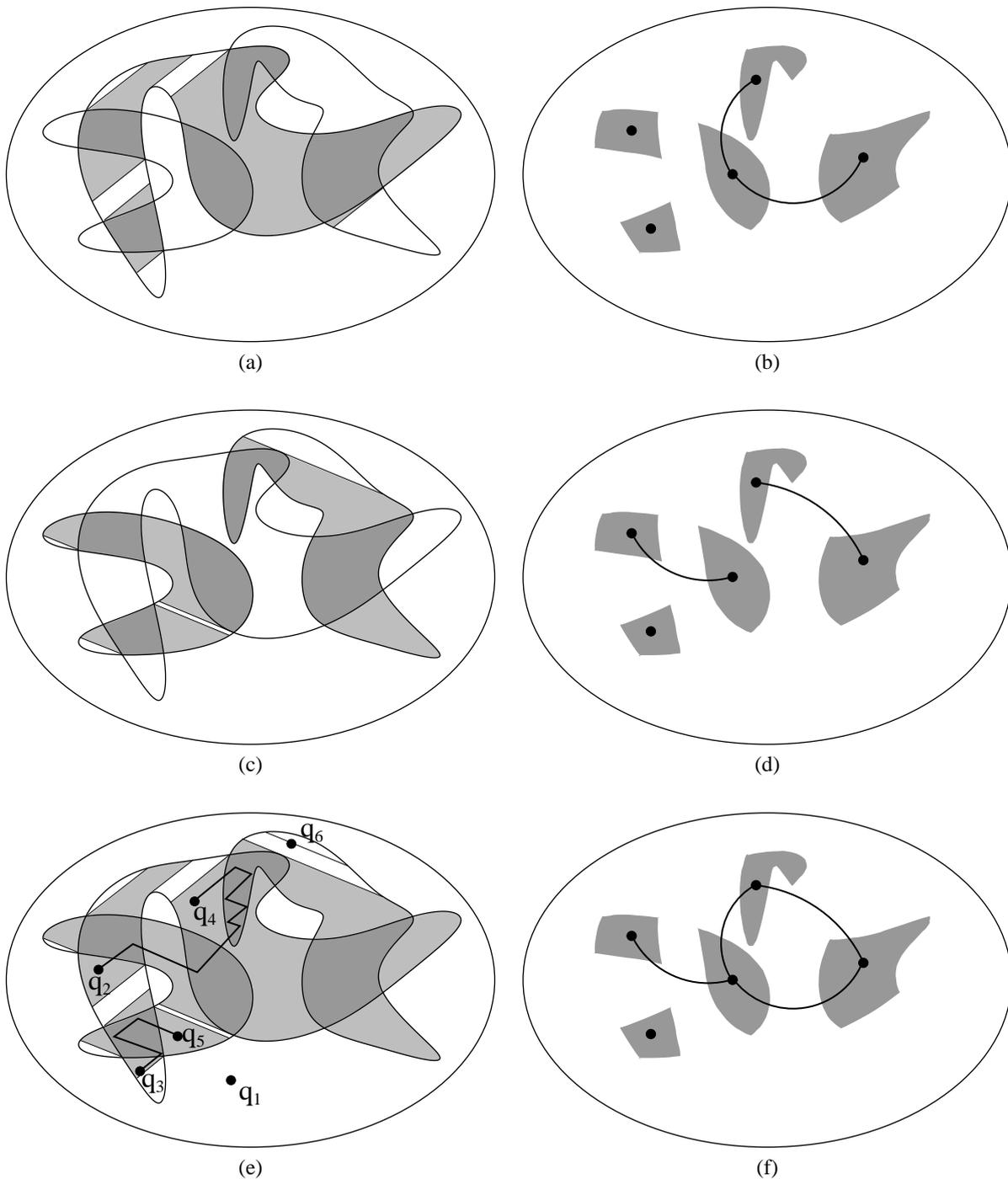


Fig. 3. The topology of  $CS$  induced by the manipulation problem constraints can be captured by a so-called manipulation graph. (a) Set of configurations reachable by a transit path starting at a configuration in  $CG \cap CP$ . (b) Adjacency of  $CG \cap CP$  components via transit paths. (c) Set of configurations reachable by a transfer path starting at a configuration in  $CG \cap CP$ . (d) Adjacency of  $CG \cap CP$  components via transfer paths. (e) Examples of manipulation paths. (f) The manipulation graph.

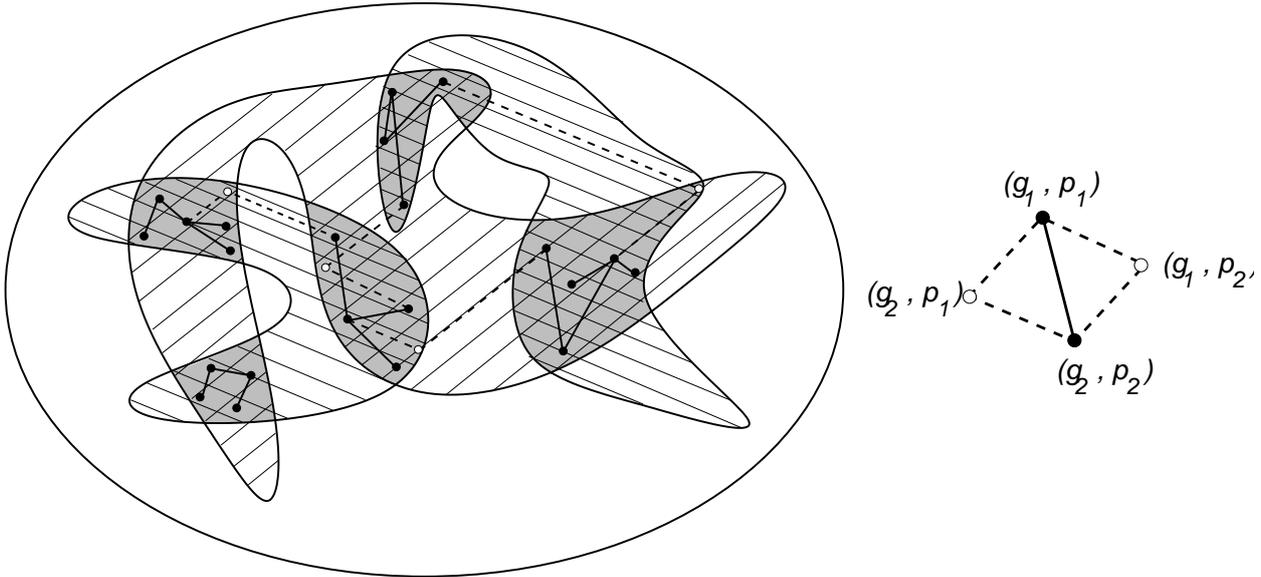


Fig. 4. A probabilistic roadmap as a manipulation graph; nodes belong to  $CG \cap CP$  while edges model paths belonging to either  $CG \cap CP$ ,  $CP$  or  $CG$ . Two types of adjacency are considered: direct  $CG \cap CP$  paths (plain segments) or elementary sequences of transit–transfer (or transfer–transit) paths (dashed segments).

$CG \cap CP$  discussed in the previous section. The main idea is to exploit the reduction property of Section to decompose the construction of the manipulation graph at two levels:

- compute the connected components of  $CG \cap CP$ ;
- determine the connectivity of  $CG \cap CP$  components using transit and transfer paths.

#### 4.1. A Two-Level Probabilistic Manipulation Roadmap

The manipulation graph is computed as in Nielsen and Kavraki (2000) using a probabilistic technique (Kavraki and Latombe 1994; Overmars and Švestka 1994), but our construction of the manipulation roadmap integrates a specific step allowing us to directly capture the connectivity of the submanifold  $CG \cap CP$  inside the roadmap. The structure of a manipulation roadmap computed using this approach is illustrated by Figure 4.

The roadmap is composed by a small number of nodes (the connected components of  $CG \cap CP$ ) connected together with transit or transfer paths. Each  $CG \cap CP$  component is captured into a subroadmap computed using a local planner that generates feasible  $CG \cap CP$  motions (the black edges in Figure 4) between nodes (in black) randomly sampled in  $CG \cap CP$ . These subroadmaps are connected via transit and transfer paths (the dotted edges) using some intermediate nodes (in white). The intermediate nodes are defined as follows. Consider two configurations in  $CG \cap CP$  that cannot be directly

connected by a collision-free path in  $CG \cap CP$  (i.e., configurations that do not belong to the same connected component of  $CG \cap CP$ ). These configurations correspond to fixed grasps and placements of the movable object, noted  $(g_i, p_i)_{i=1,2}$ . Using motions outside  $CG \cap CP$ , they can only be connected by following the particular leaves of  $CP$  and  $CG$  issued from both configurations. We then define the intermediate nodes as  $(g_1, p_2)$  and  $(g_2, p_1)$ . An edge between  $(g_1, p_1)$  and  $(g_2, p_2)$  is added if at least one of the intermediate nodes  $(g_1, p_2)$  and  $(g_2, p_1)$  belongs to  $CG \cap CP$  and the node is reachable from  $(g_1, p_1)$  and  $(g_2, p_2)$  by a collision-free transit/transfer path. The connection between two randomly sampled configurations of  $CG \cap CP$  is then possible if one of the three types of adjacency (Figure 4) exists:

- **Type1**—a direct path from  $(g_1, p_1)$  to  $(g_2, p_2)$  lying inside  $CG \cap CP$  is collision-free;
- **Type2a**—a transfer path from  $(g_1, p_1)$  to  $(g_1, p_2)$  followed by a transit path from  $(g_1, p_2)$  to  $(g_2, p_2)$  are both collision-free;
- **Type2b**—a transit path from  $(g_1, p_1)$  to  $(g_2, p_1)$  followed by a transfer path from  $(g_2, p_1)$  to  $(g_2, p_2)$  are both collision-free.

Once the manipulation roadmap is computed, queries are solved by searching for a path inside  $MG$ . The obtained solution alternates elementary manipulation paths (i.e.,

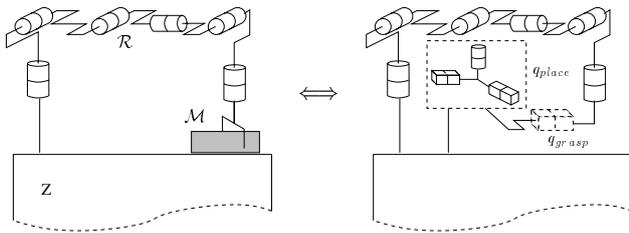


Fig. 5. Closed-chain system (right) formed in  $CG \cap CP$  by the robot and the movable object (left).

transfer/transit paths computed when traversing edges of  $MG$  using `Type2` adjacencies) with  $CG \cap CP$  paths (i.e., paths computed inside the nodes of  $MG$  using `Type1` adjacencies). Note that the direct  $CG \cap CP$  paths correspond to simultaneous changes of grasp and placement; they are therefore not feasible from the manipulation point of view. However, thanks to the reduction property, any such `Type1` paths can be transformed in a post-processing stage into a finite sequence of `Type2` transit and transfer paths.

#### 4.2. Capturing $CG \cap CP$ Topology via Closed-Chain Systems

The main critical issue of the approach is to capture into a probabilistic roadmap the topology of  $CG \cap CP$  which is a submanifold of the global configuration space  $CS$  with a lower dimension. The idea here is to explore  $CG \cap CP$  as such. For this, we consider that  $CG \cap CP$  is the configuration space of a single system consisting of the robot together with the movable object placed at a stable position. Maintaining the stable placement while the object is grasped by the robot induces a closed chain for the global system (Figure 5).

We now explain how the closed chain used for the exploration of  $CG \cap CP$  is defined. A fixed grasp of the movable object corresponds to a transformation matrix  $T_g$  positioning the end-frame of the robot with respect to the coordinate frame of the object. The set of continuous grasps can then be defined by a transform matrix  $T_g(q_{grasp})$  where  $q_{grasp}$  denotes a set of varying parameters. The  $CG$  subspace corresponds to the set of free configurations  $(q_{rob}, q_{obj})$  for which the configuration  $q_{obj}$  of  $\mathcal{M}$  changes according to the grasp mapping induced by the forward kinematics of the robot and by the grasp of the object:  $q_{obj} = \mathcal{G}(q_{rob}, q_{grasp})$ .  $CG$  is therefore parametrized by the configuration vector  $(q_{rob}, q_{grasp})$  associated with a composite robot obtained by adding virtual joints induced by  $q_{grasp}$  between the last link of  $\mathcal{R}$  and the object  $\mathcal{M}$ . On the other hand, the set of stable placements is defined by a transformation matrix  $T_p(q_{place})$  relating the object's frame to the world frame, where  $q_{place}$  denotes the set of varying placements parameters. The  $CP$  submanifold corresponds to configurations where  $q_{obj}$  changes according to the

mapping  $q_{obj} = \mathcal{P}(q_{place})$ . Then, the  $CG \cap CP$  space can be parametrized as the set of configurations  $(q_{rob}, q_{grasp}, q_{place})$  satisfying the closure constraints  $\mathcal{G}(q_{rob}, q_{grasp}) = \mathcal{P}(q_{place})$ .

Facing such sub-dimensional manifolds is a challenging problem for motion planning. In particular, applying a purely randomized PRM framework (Kavraki and Latombe 1994; Overmars and Švestka 1994) to closed-chain mechanisms is prohibited by the fact that the probability to choose a configuration at random on a given sub-dimensional manifold is null (LaValle, Yakey, and Kavraki 1999). However, several recent contributions (LaValle, Yakey, and Kavraki 1999; Han and Amato 2001; Cortés, Siméon, and Laumond 2002) have extended the PRM framework to face this issue. In Section 5 we describe the planning technique used in our implementation.

#### 4.3. Connections with Transit and Transfer Paths

Computing such connections requires that we solve multiple point-to-point path planning problems, as for the case of discrete grasps and placements. Here, the issue is to provide efficient solutions for searching such collision-free transit (or transfer) paths in the various leaves of  $CP$  (or  $CG$ ). For example, the fuzzy roadmap technique (Nielsen and Kavraki 2000) could be used to gain efficiency by limiting the number of collision tests performed when solving the queries. Our implemented planner, however, uses another kind of speed-up. It relies on a simple technique sharing a similar idea with the kinematic roadmaps (Han and Amato 2001). It exploits the fact that each planning problem has to be performed in a partially modified environment to re-use a precomputed static roadmap that is dynamically updated when solving the planning queries. This planning technique is also further explained in the section below.

### 5. Planning Techniques

We now detail the planning techniques developed to implement the approach. The two basic primitives required for computing the `Type1` and `Type2` motions are respectively described in Sections 5.1 and 5.2. Then, we explain how both primitives are combined by the algorithm used to build the manipulation roadmap.

#### 5.1. Closed-Chain Planner for `Type1` Motions

As explained above, our approach requires the application of planning techniques for closed-chain systems in order to capture the topology of  $CG \cap CP$ . Several recent contributions extended the PRM framework to deal with closure constraints (LaValle, Yakey, and Kavraki 1999; Han and Amato 2001; Cortés, Siméon, and Laumond 2002). In particular, we use the Random Loop Generator (RLG) algorithm (Cortés, Siméon, and Laumond 2002) that demonstrates good performance on

complex 3D closed chains involving more than 20 DoF. As initially proposed in Han and Amato (2001), the loop is broken into two open subchains, called the active and passive subchains. Using the RLG algorithm, the random closure configurations (i.e., valid nodes) are obtained by combining random sampling techniques with simple geometrical operations that compute approximated reachable workspaces of various subchains to iteratively generate the configuration for the active chain. Then, it performs inverse kinematics for the remaining passive part of the loop in order to force the closure constraint. The advantage of the RLG algorithm is to produce random samples for the active chain that have a high probability to be reachable by the passive part. This significantly decreases the cost of computing and connecting closure configurations. The roadmap edges are computed using a local planner limited to act on the active joints, while the passive part of the loop follows the motion of the rest of the chain. The practical efficacy of our approach results from the good performance reached today by these closed-chain extensions of the PRM framework.

The  $CG \cap CP$  roadmap is then computed using Visibility-PRM (Laumond and Siméon 2000; Siméon, Laumond, and Nissoux 2000). This technique keeps the roadmap as small as possible by only adding two types of useful samples: guards that correspond to samples not already “seen” by the current roadmap, and connectors allowing to merge several connected components. Its interest is first to control the quality of the roadmap in term of coverage and, second, to capture the connectivity of possibly complex spaces into a small data structure. We believe that the small size of the visibility roadmaps, combined with the proposed structuring of  $CG \cap CP$ , contributes to the overall efficiency of our approach by limiting the number of costly path planning queries to be performed during the second stage, when searching the connections with collision-free transfer or transit paths.

Figure 6 shows the closed-chain system formed by the 6-DoF arm manipulating the long bar for the manipulation example of Figure 1. The bar moves in contact with the floor while sliding within the gripper. The sliding motion of the gripper results from the additional DoF  $q_{grasp}$  introduced in the system to characterize the infinite set of grasps. In this example,  $q_{grasp}$  is chosen to allow a translation of the parallel jaw gripper along the bar. Similarly, the set of stable placements corresponds to the planar motions parametrized by a 3D vector  $q_{place}$  (two horizontal translations and a vertical rotation), that maintain the contact of the bar with the floor. The motion shown in the right image of Figure 6 is a feasible motion in  $CG \cap CP$ . It is not admissible from the manipulation problem point of view. However, thanks to the reduction property it can be transformed into a finite sequence of feasible transit and transfer paths.

Figure 7 shows the visibility roadmap computed by the algorithm in  $CG \cap CP$  for the example of Figure 1. While the collision-free configuration space of the arm alone is con-

nected,  $CG \cap CP$  is not. The computed roadmap has four connected components: two main components separated by the long static obstacle, and two other small components that correspond to placements of the movable object inside the cage obstacle while it is grasped by the arm through the open passage in the middle of the cage. These two small components (inside the dashed circle of the left image) correspond to the same position of the bar with two different orientations  $180^\circ$  apart. The associated placement of the system is shown in the top-right image. The bottom-right image corresponds to a node of the main component with the bar placed at the same position, but using a different grasp. Connecting this node to the small component is not possible because of the cage obstacle that limits the continuous change of grasp. Such re-grasping requires the computation of collision-free paths outside  $CG \cap CP$  as explained below.

## 5.2. Connection Planner for Type2 Motions

Computing Type2 connections requires a basic routine to find elementary collision-free transit and transfer paths. Each of the planning problems corresponds to a particular grasp or placement of the movable object. Then, the queries have to be performed in a partially modified environment. The motivation of the two-stage method used by the connection planner is simply to reduce the cost of dealing with such partial changes by re-using at each query some of the paths pre-computed during the first stage regardless of the movable object.

First, we compute a roadmap for the robot and the static obstacles, without considering the presence of the movable object. Then, before solving a given (transit or transfer) path query, the roadmap is updated by checking whether each edge is collision-free in respect with the current position of the movable object. Colliding edges are labeled as blocked in the roadmap.

The search for a given path is then performed within the labeled roadmap. As illustrated by Figure 8, three cases possibly occur. When the search fails, this means that no path exists even in the absence of the movable object; the problem has no solution. Similarly, when the computed path does not contain any blocked edge (dashed edges in Figure 8) then a solution is found. Now let us consider the intermediate situation where the solution path necessarily contains blocked edges. In such case, the algorithm tries to solve the problem locally using a rapidly-exploring random tree planner (Kuffner and LaValle 2000) to connect the endpoints of the blocked edges. The principle of the bidirectional RRT-Connect algorithm (see Kuffner and LaValle 2000) used in our connection planner consists in incrementally building two random trees rooted at the start and goal configurations, such that both trees explore the space around them and advance toward each other through the use of a simple heuristic. This algorithm was originally designed to efficiently process single-query path planning problems. The main interest of RRT is to perform well locally. Its complexity

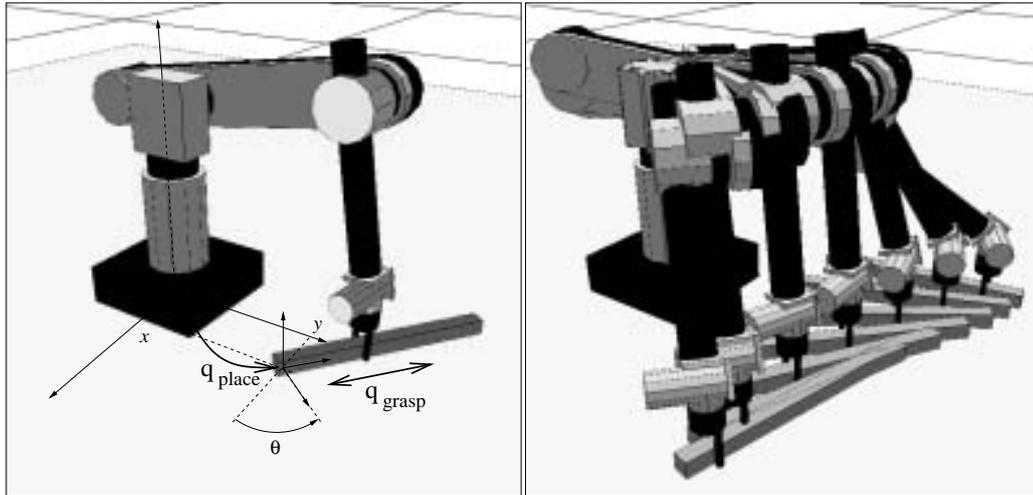


Fig. 6. Virtual closed-chain system and a feasible  $CG \cap CP$  motion (the bar moves on the floor while sliding into the gripper's jaws).

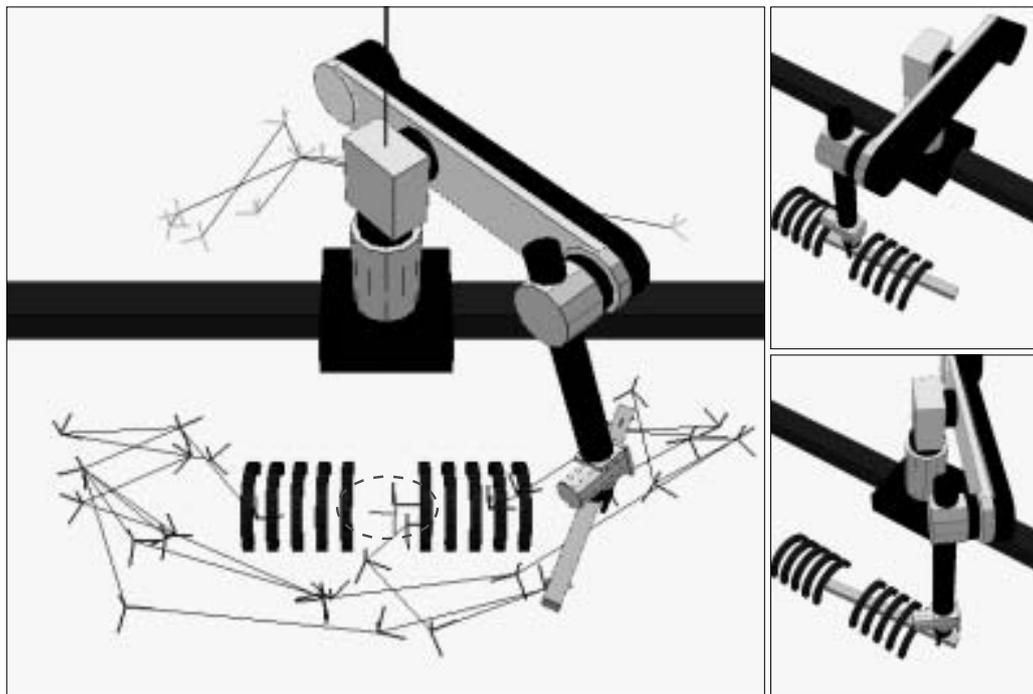


Fig. 7. A visibility roadmap computed in  $CG \cap CP$  (left) and two placements of the system inside two different connected components of  $CG \cap CP$  (right).

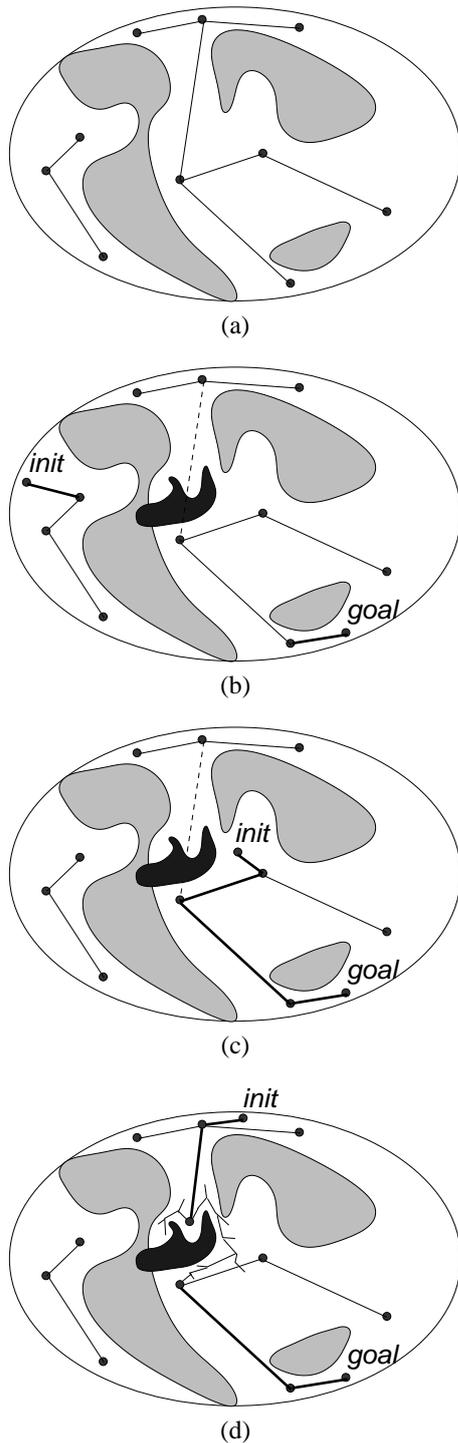


Fig. 8. (a) A static roadmap is computed in the configuration space of the robot. During queries, it is labeled according to collisions with  $\mathcal{M}$ . If the query fails, there is no solution (b). Otherwise, either there exists a solution path in the roadmap avoiding labeled edges (c) or not (d). In the latter case, the colliding part of the path is locally updated using an RRT-like technique.

depends on the length of the solution path. This means that the approach quickly finds easy solutions. It may be viewed as a dynamic updating of the roadmaps.

Figure 9 shows the connecting paths computed by the planner for linking the connected components of the  $CG \cap CP$  roadmap shown in Figure 7. The transfer path (left) is used to connect the two main components of  $CG \cap CP$ , while the transit path connects the small component (inside the dashed circle of Figure 7) to the main one.

### 5.3. Manipulation Planning Algorithm

The algorithm incrementally constructs the manipulation roadmap  $MG$  by interleaving the two steps of the approach: computing  $CG \cap CP$  connected components (Type1 adjacency) and linking them (Type2a-b adjacencies). Following the principle of Visibility-PRM, the algorithm stops when it is not able to expand the graph after a given number of tries. This number of failures is related to an estimated coverage of the search space (Siméon, Laumond, and Nissoux 2000; in our case, the  $CG \cap CP$  space). The function EXPAND\_GRAPH performs one expansion step of  $MG$ . Candidate nodes are first sampled in  $CG \cap CP$  and the different types of connections to the graph are then tested.

---

```

EXPAND_MANIP_GRAPH( $MG$ )
 $q \leftarrow$  NEW_CONFIG( $MG$ )
 $Type \leftarrow$  ADJACENCY_CHOICE( $MG$ )
 $n_{linked\ comp.} \leftarrow$  TEST_CONNECTIONS( $MG, q, Type$ )
if  $n_{linked\ comp.} \neq 1$  then
    ADD_NODE( $q, MG, Type$ )
    UPDATE_GRAPH( $MG$ )
    return TRUE
else
    return FALSE
    
```

---

#### 5.3.1. Node Generation

Our algorithm possibly considers several classes of continuous grasps (placements), each defined by a transformation matrix  $T_{g_i}(q_{grasp})$  ( $T_{p_j}(q_{place})$ ) with  $q_{grasp}$  ( $q_{place}$ ) as varying parameters. Therefore, each couple  $(T_{g_i}, T_{p_j})$  induces a particular closed-chain system. A candidate node is generated as follows by the function NEW\_CONFIG; it first randomly selects one couple  $(i, j)$  of grasps and placement classes. The grasp and the stable placement of the movable object is then chosen by randomly sampling the parameters of vectors  $q_{grasp}$  and  $q_{place}$  inside their variation interval. The candidate node  $N$  is generated when the sampled grasp and placement are collision-free and feasible for the virtual closed system induced by the couple  $(T_{g_i}, T_{p_j})$ .

#### 5.3.2. Adjacency Selection

Following the discussion in Section 4.2, the desired behavior of the roadmap builder is to start by constructing portions of

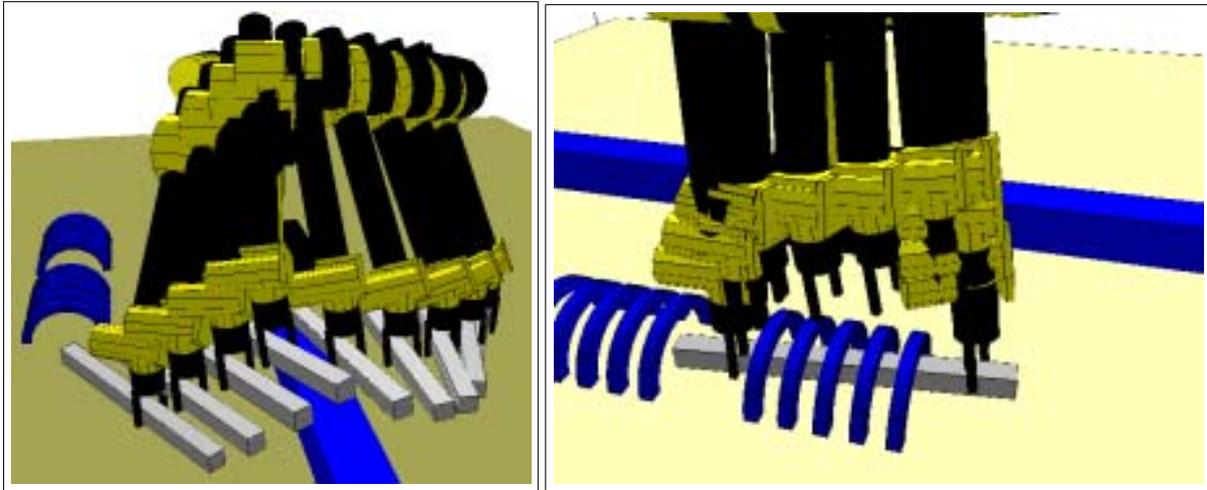


Fig. 9. The transfer path (left) and the transit path (right) computed for connecting  $CG \cap CP$  components shown in Figure 7.

the roadmap inside  $CG \cap CP$  components using `Type1` adjacency, and then to determine connections of the components using `Type2` adjacencies. Rather than considering separately the two stages, the algorithm uses a more sophisticated way to interleave both phases. Function `ADJACENCY_CHOICE` performs a biased random choice  $\{\text{Type1}, \text{Type2}\}$  that depends on the evolution of the size of  $MG$ : the first expansion steps start with a low probability to return a `Type2` choice; when the roadmap grows, this probability increases as the percentage of the coverage  $cov$  estimated by the fraction  $(1 - \frac{1}{n_{try}})$  (see Siméon, Laumond, and Nissoux 2000 for details).

A tuning parameter  $\alpha \in [0, 1[$  is used to put more or less weight between expanding the  $CG \cap CP$  components and connecting them using transit/transfer paths: the probability of choosing the  $CG \cap CP$  expansion is determined by  $Prob(\text{Type1}) = \alpha \cdot (1 - cov)$  and  $Prob(\text{Type2}) = 1 - Prob(\text{Type1})$ . With  $\alpha$  set to zero, the roadmap builder only considers connections of  $MG$  nodes with transit/transfer paths. When  $\alpha$  tends toward 1, the algorithm rarely selects such `Type2` connections before a sufficient coverage of  $CG \cap CP$  has been reached. The effect of  $\alpha$  on the performance of the algorithm when solving the manipulation problem of Figure 1 is further discussed in Section 6.

---

```

TEST_CONNECTIONS( $MG, q, Type$ )
 $n_{linked\ comp.} \leftarrow 0$ 
for  $k = 1$  to  $N\_COMP(MG)$  do
if LINKED_TO_COMP( $q, C_k, Type$ ) then
     $n_{linked\ comp.} = n_{linked\ comp.} + 1$ 
return  $n_{linked\ comp.}$ 

```

---

### 5.3.3. Edge Generation

The function `TEST_CONNECTIONS` checks the connection between the candidate node and each connected component  $C_k$  of  $MG$  using the type of adjacency selected by function `ADJACENCY_CHOICE`. When the expansion step is performed using `Type1` motions, connections are computed using the closed-chain planner of Section 5.1. In this case, it can be noted that the connection of the candidate node to the roadmap is only possible with nodes computed for the same classes of grasps and placements  $(T_{g_i}, T_{p_j})$ . For each component  $C_k$ , nodes with such characteristics are tested until a connection is found feasible for the closed-chain mechanism induced by  $(T_{g_i}, T_{p_j})$ . When the expansion is performed using `Type2` motions, function `TEST_CONNECTIONS` stops checking the component  $C_k$  as soon as valid connection is found using the planning technique of Section 5.2. Following the visibility principle, the candidate node is added to the graph only if the random sample  $q$  was linked to none or to more than one connected component. In the second case, the linked components are merged.

### 5.3.4. Solving Manipulation Queries

Once the manipulation roadmap is built, queries can be performed using the three following steps. First, the start and goal configurations are connected to  $MG$  using the `TEST_CONNECTIONS` function called with a `Type2` adjacency choice, and the manipulation graph is searched for a path between both configurations. The second step is necessary to transform  $CG \cap CP$  portions of the solution path into a finite sequence of transfer/transit paths. This is done by a

dichotomic procedure that iteratively splits the  $CG \cap CP$  paths into pieces whose endpoints can be connected by a composition of two collision-free transit/transfer paths. The operation of the algorithm is very simple. It begins by computing the `Type2a` path and the `Type2b` path which connect the initial and final configurations of the `Type1` portion (see Figure 4). If one of the paths is collision-free, the algorithm stops and returns the collision-free path. If both paths are colliding, the configuration halfway along the  $CG \cap CP$  portion is generated and the algorithm is recursively applied to two subpaths connecting this intermediate configuration to the initial and the final ones. When all the necessary subdivisions are completed, the concatenation of all elementary subpaths is collision-free and respects the manipulation constraints. The process is guaranteed to converge. Finally, the solution is smoothed by a procedure that eliminates unnecessary motions.

## 6. Performance Analysis

### 6.1. Performance of the Approach

The purpose of the proposed approach is first to reduce the complexity of the problem since the  $CG \cap CP$  submanifold is a lower-dimensional space compared to the leaves of the placements and grasps spaces. Let us illustrate this by detailing the dimension of the various spaces for the problem of Figure 1. Here, we have  $\dim(CS_{rob}) = 6$ ,  $\dim(CS_{obj}) = 6$  and  $\dim(CS) = 12$ . Placements of the object are allowed only when the bar is placed on the table (three DoF). For a fixed placement of the bar, the robot can freely move its six DoF. Then, the dimension of the placement space is  $\dim(CP) = 9$ . The bar is grasped by the robot by allowing a (one DoF) translating motion along its length; we then have  $\dim(CG) = 7$ . In this example, the leaves in both  $CP$  and  $CG$  have dimension 6 while  $\dim(CG \cap CP) = 4$ .

The other rationale is also to enlarge the size of the solution space when searching inside  $CG \cap CP$ . Once a solution path (including `Type1` sliding motions) is found, it is always possible to approximate it by a feasible manipulation path. Such an additional transformation step is preferable to other approaches that would directly take into account the manipulation constraints during the search. In particular, for solving the problem of Figure 1, the sliding motion allowing us to get the bar out the cage (see Figure 11) is obtained much more easily inside  $CG \cap CP$  than the resulting sequence of transit/transfer paths that would be computed by the existing planners (e.g., Ahuactzin, Gupta, and Mazer 1998; Nielsen and Kavraki 2000) after discretizing the continuous grasps and placements.

### 6.2. Influence of the $\alpha$ Parameter

Let us now discuss the performance of the planner according to  $\alpha$  which is the major parameter of our planner. The

curve displayed in Figure 10 plots the time<sup>4</sup> spent by the algorithm to build the manipulation roadmap allowing to solve the illustrative problem of Figure 1. As explained above, the role of the parameter  $\alpha$  is to control the rate of connections searched inside  $CG \cap CP$  (`Type1` adjacencies) compared to connections searched outside  $CG \cap CP$  along the leaves of the  $CG$  and  $CP$  spaces (`Type2` adjacencies). When  $\alpha = 0$ , the roadmap builder only considers collision-free transit and transfer paths to connect the random samples generated in  $CG \cap CP$ . In this case the algorithm behaves as the discrete approaches.

Increasing  $\alpha$  allows us to favor the construction of the  $CG \cap CP$  connected components using `Type1` adjacencies before trying possible connections along leaves with `Type2` adjacencies. The curve shows that the computation time significantly decreases for runs performed with higher values of  $\alpha$ . This increased performance can be explained by the fact that many searches of collision-free motions along the leaves of  $CP$  and  $CG$  are avoided thanks to the direct exploration of the  $CG \cap CP$  submanifold. Note, however, that when  $\alpha$  tends towards 1, the probability of selecting `Type2` adjacencies remains very low until a sufficient coverage of  $CG \cap CP$  with `Type1` adjacencies has been reached. Since `Type2` adjacencies are required to link the  $CG \cap CP$  connected components, the performance decreases again when  $\alpha \rightarrow 1$ . The reason is that the algorithm spares time to reach such good coverage inside  $CG \cap CP$  instead of trying connections outside  $CG \cap CP$ . In all the experiments performed with the planner, this degradation of performance was observed to become significant for values of  $\alpha$  closed to 1. The experimental study conducted on the difficult manipulation problem of Figure 1 tends to show that when the problem is rather constrained, it is qualitatively advantageous to spend time on the connectivity of the  $CG \cap CP$  submanifold before checking connections with feasible manipulation paths. As shown by the curve, the gain can be very important in such constrained situations. It is however observed to be less significant on simpler problems like the two other examples presented in the next section. As often with the probabilistic methods, the choice of the best value for this parameter remains an issue that would need to be further investigated. In our experiments with the planner, runs are generally performed with a value of  $\alpha$  set to .9.

## 7. Experimental Results

The manipulation planner was implemented within the software platform Move3D (Siméon et al. 2001) developed at the Laboratory for Analysis and Architecture of Systems (LAAS). Several environments have been used as test-bed of the planner. In this section, we present the results obtained with three of them. The computation times correspond to experiments conducted on a 330 MHz Sparc Ultra 10 workstation.

4. Each time value was averaged over ten runs performed using different seeds to initialize the random generator.

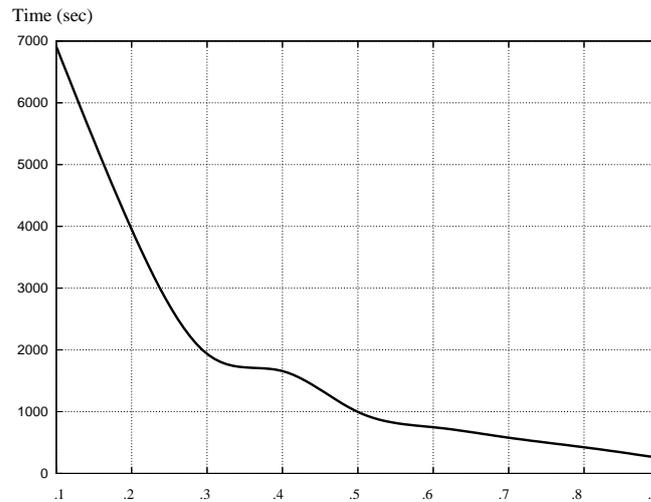


Fig. 10. Performance of the algorithm depending on the percentage of  $CG \cap CP$  exploration (Type1 paths) with respect to transit-transfer (Type2 paths) used to build the manipulation roadmap for the example of Figure 1. The abscissa corresponds to the parameter  $\alpha$ .

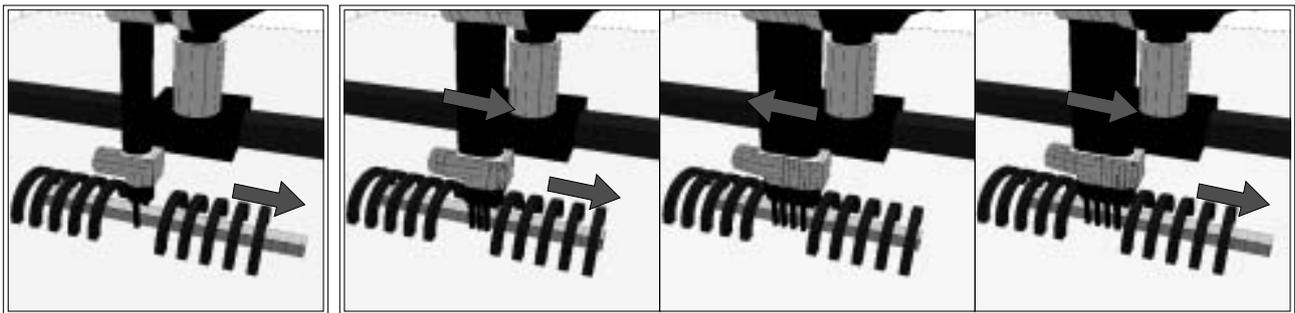


Fig. 11. A  $CG \cap CP$  path with a sliding motion of the bar (left) transformed into a sequence of three feasible transit-transfer manipulation paths (right).

The first example corresponds to the problem of Figure 1. We refer to it as the Cage example. Two other scenes are shown in Figure 12: the left image illustrates a problem (MulGP) involving the same arm manipulating a more complicated U-shaped object. Manipulating this object requires to consider multiple classes of grasps and of placements depending on the contact faces used to grasp/release the object. The right example (MobM) corresponds to a problem involving a manipulator arm mounted on a mobile platform.

The difficulty with the Cage example is the complexity of the manipulation task. Several consecutive re-grasping motions through the middle of the cage obstacle are necessary to move the bar to a position where it can be regrasped by its

extremity. The planner automatically computes the required configurations from only one continuous placement domain (the floor) and one grasping zone all along the bar. The path to get the bar out of the cage is found in the  $CG \cap CP$  manifold, and then transformed during the post-processing step in a sequence of transit and transfer paths (see Figure 11). The final path contains 20 elementary paths with eight re-grasping of the movable object. This difficult manipulation problem was solved in less than 2 min, which demonstrates the efficacy of the proposed approach.

In the example MulGP, the manipulation problem is to re-orient the U-shaped movable object, starting from an initial placement where it is trapped by the mechanical device lying

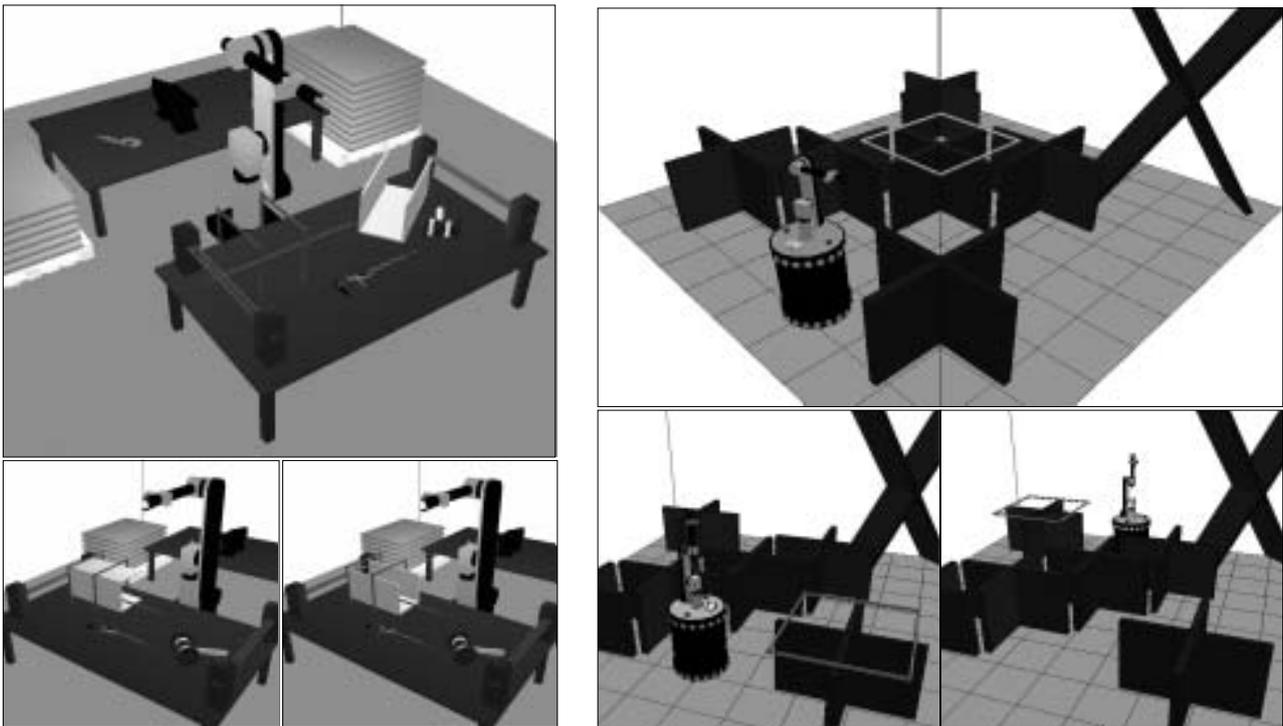


Fig. 12. Scenes and start/goal positions for the MulGP and MobM examples.

at the left of the workplan. This problem was solved by considering eight grasp classes, each corresponding to a continuous grasp along one of the eight thin faces defined by the U-shaped form of the object. Also, the set of stable placements corresponds to positions where one of the three large faces contact with the workplan. We then consider three classes of placements according to the orientation of the movable object when placed on the table. Figure 13 shows the manipulation solution computed by the planner. Here, the presence of several grasp/placements classes, and the larger size of the movable object (which results in more RRT calls during the connection stage) increase the overall cost of manipulation planner (see Table 1).

In the example MobM, the mobile manipulator (nine DoF) can only pass from one side of the scene to the other side through the passage under the X-shaped obstacle. However, this passage is too narrow for the movable object (the square frame). A continuous grasping set is defined all around this object. The frame can be placed on the central obstacles. Figure 14 shows the manipulation solution computed by the planner. Here, the manipulation task is simpler compared to the previous examples; fewer re-grasps are needed to solve the problem. The difficulty illustrated by the example is to deal with a redundant system. An infinite set of solutions exists to achieve the same grasp. Redundancy is a challenge

**Table 1. Numerical Results**

	Examples		
	Cage	MulGR	RobM
Total time	96 s	330 s	293 s
Computing $CG \cap CP$	23 s	10 s	6 s
Transit/transfer paths	70 s	306 s	284 s
Total collision checks	43,518	187,342	102,241
Local $CG \cap CP$ paths	3689	1104	190
Transit/transfer paths	54	168	147
Dynamic updates	15	91	17
No. of manipulation nodes	32	36	21
No. of manipulation paths	29	30	21

when treating closed-chain mechanisms. The exploration of the  $CG \cap CP$  manifold for such systems is efficiently performed using the RLG-based closed-chain planner (Cortés, Siméon, and Laumond 2002).

Table 1 shows, for the three examples, numerical results that illustrate the good performance of the planner. All problems were solved with  $\alpha = 0.9$  after less than 5 min of computation. Most of the computation time is spent checking connections with transit and transfer paths; this shows the

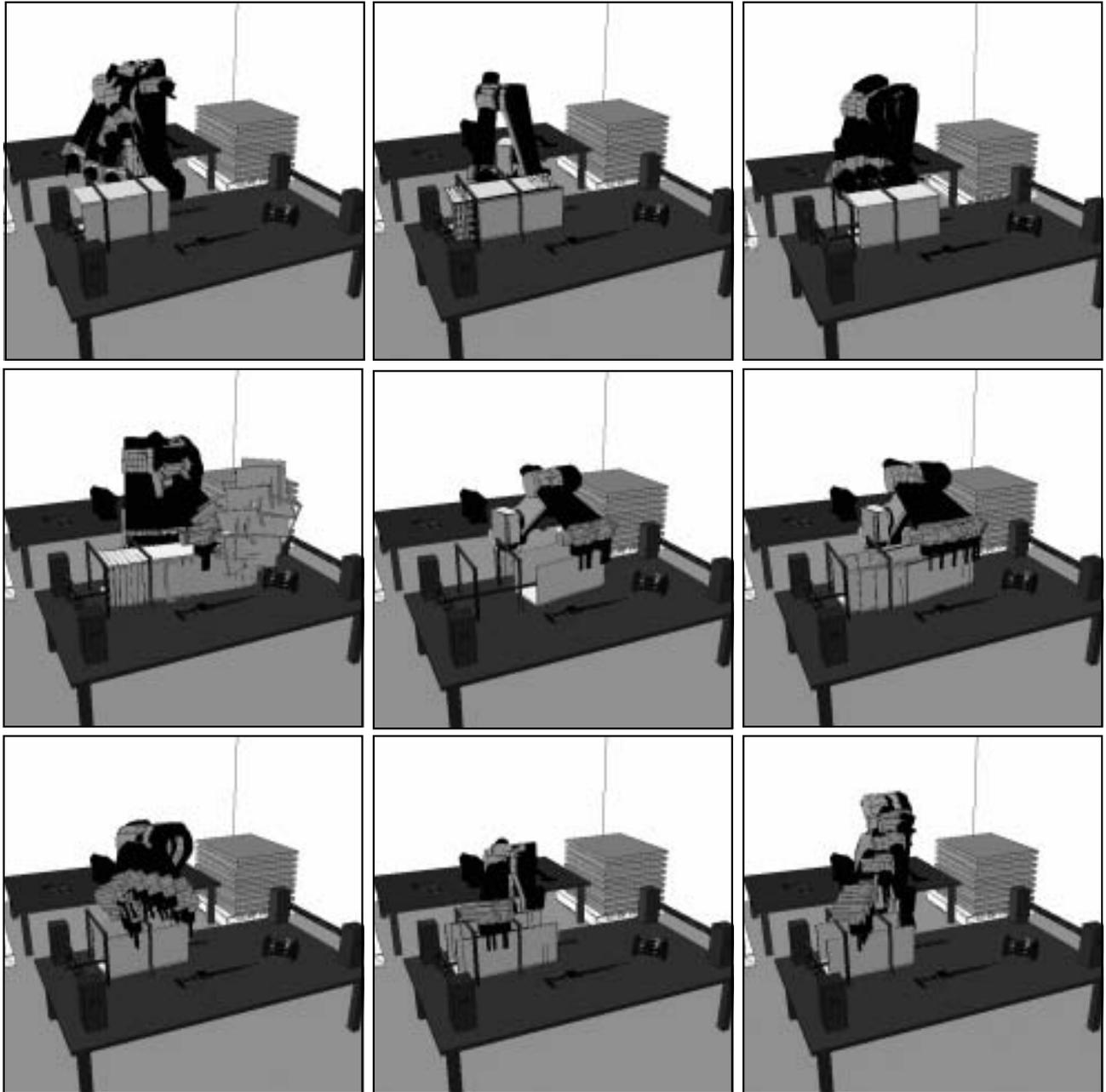


Fig. 13. Manipulation path computed for the MulGP problem.

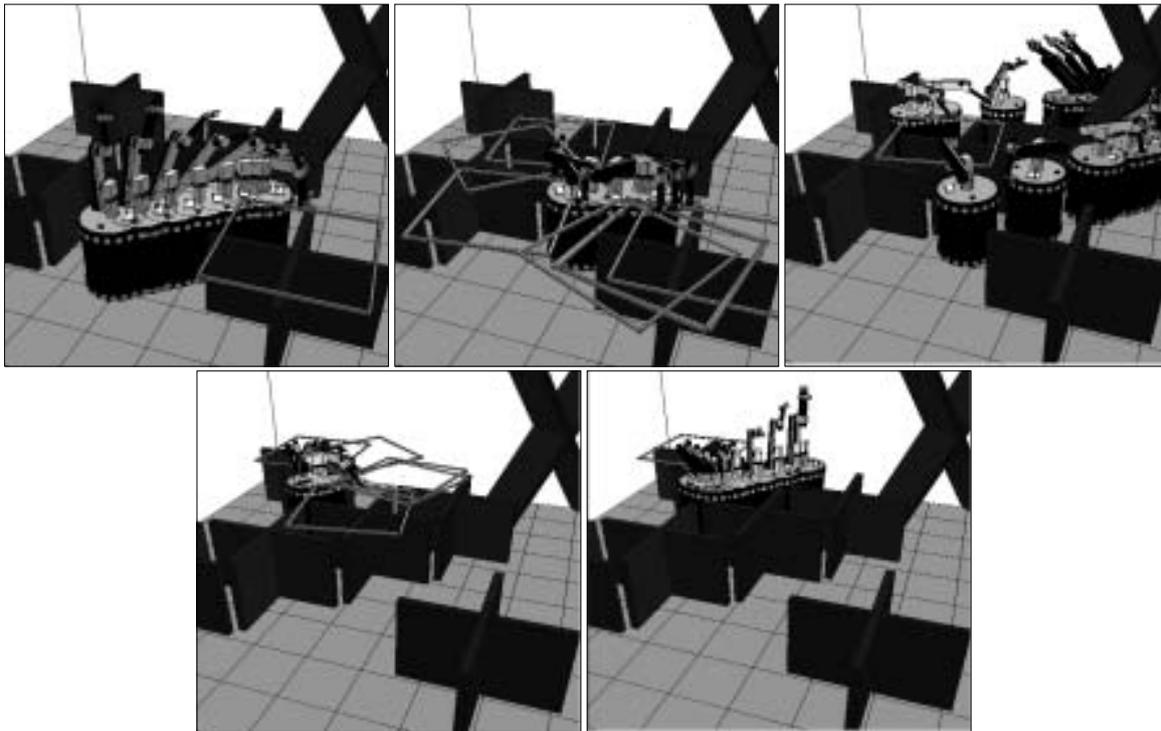


Fig. 14. Manipulation path computed for the MobM problem.

advantage of the proposed approach which limits the number of such connection tests by first computing connected components inside  $CG \cap CP$ . Also, the use of the visibility technique is the reason for the small size of the manipulation roadmaps; such small roadmaps also reduce the number of connections to be tested.

## 8. Conclusion

We have presented a new approach to manipulation planning. Its power lies in the fact that it can deal with a continuous formulation of the manipulation problem. It is based on structuring the search space to directly capture the connectivity of the submanifolds that correspond to the places where transit and transfer paths can be connected. This structure is expressed in a probabilistic roadmap and allows us to design a manipulation planner that automatically generates, inside continuous domains, the grasps and placements that make the problem solvable. Simulation results show the approach's effectiveness for solving complex manipulation problems.

There remain several possible improvements. For example, it is important to improve the performance of the connection planner which remains the most costly operation. This relates to the efficiency of PRM planners when facing dynamic changes of the environment. Although the approach has the potential to handle general models of the grasp and placements

spaces, the planner is currently implemented for the particular case of planning pick-and-place operations for polyhedral objects. One could, however, imagine applications requiring other models. Finally, the manipulation planner is currently restricted to a single movable object manipulated by a single robot. The case of multiple movable objects and robots requires studying the conditions under which the reduction property can be extended to such situations. We also began to investigate a more general approach (Gravot, Alami, and Siméon 2002) combining a symbolic task planning level with the geometric manipulation planner in order to solve problems of higher complexity with multiple objects and robots.

## Acknowledgment

This work has been partly supported by the IST European Project 39250 MOVIE.

## References

- Ahuactzin, J. M., Gupta, K., and Mazer, E. 1998. Manipulation planning for redundant robots: a practical approach. *International Journal of Robotics Research* 17(7):731–747.
- Ahuactzin, J. M., and Gupta, K. 1999. The kinematic roadmap: a motion planning based global approach for inverse kinematics of redundant robots. *IEEE Transactions on Robotics and Automation* 15(4):653–670.

- Alami, R., Siméon, T., and Laumond, J. P. 1989. A geometrical approach to planning manipulation tasks. The case of discrete placements and grasps. *5th International Symposium on Robotics Research*, Tokyo, Japan.
- Alami, R., Laumond, J. P., and Siméon, T. 1994. Two manipulation planning algorithms. *Algorithmic Foundations of Robotics (WAFR94)*, Stanford, CA.
- Barraquand, J., and Ferbach, P. 1994. A penalty function method for constrained motion planning. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, San Diego, CA.
- Barraquand, J., and Latombe, J.-C. 1991. Robot motion planning: a distributed representation approach. *International Journal of Robotics Research* 10(6):628–649.
- Bessiere, P., Ahuactzin, J., El-Ghazali, T., and Mazer, E. 1993. The “Ariadne’s Clew” algorithm: global planning with local methods. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Yokohama, Japan, July 26–30.
- Chen, P. C., and Hwang, Y. K. 1991. Practical path planning among movable obstacles. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Sacramento, CA.
- Cortés, J., Siméon, T., and Laumond, J. P. 2002. A random loop generator for planning the motions of closed kinematic chains with PRM methods. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, EPFL, Lausanne, Switzerland, September 30–October 4.
- Ito, K. (ed). 1987. *Encyclopedic Dictionary of Mathematics*, MIT Press, Cambridge, MA.
- Gravot, F., Alami, R., and Siméon, T. 2002. Playing with several roadmaps to solve manipulation problems. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, EPFL, Lausanne, Switzerland, September 30–October 4.
- Han, L., and Amato, N. 2000. A kinematics-based probabilistic roadmap method for closed kinematic chains. *Algorithmic and Computational Robotics (WAFR00)*, Hanover, NH.
- Kavraki, L., and Latombe, J.-C. 1994. Randomized preprocessing of configuration space for fast path planning. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, San Diego, CA.
- Koga, Y., and Latombe, J.-C. 1992. Experiments in dual-arm manipulation planning. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Nice, France.
- Koga, Y., and Latombe, J.-C. 1994. On multi-arm manipulation planning. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, San Diego, CA.
- Kuffner, J., and Lavelle, S. 2000. RRT-Connect: an efficient approach to single-query path planning. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, San Francisco, CA, April 24–28.
- Latombe, J.-C. 1991. *Robot Motion Planning*, Kluwer, Dordrecht.
- Laumond, J. P., and Alami, R. 1989. A geometrical approach to planning manipulation tasks in robotics. LAAS Technical Report No. 89261.
- Laumond, J. P., and Siméon, T. 2000. Notes on visibility roadmaps for motion planning. *Algorithmic and Computational Robotics (WAFR00)*, Hanover, NH.
- LaValle, S., Yakey, J.H., and Kavraki, L. 1999. A probabilistic roadmap approach for systems with closed kinematic chains. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Detroit, MI.
- Lozano-Pérez, T., Jones, J.L., Mazer, E., and O’Donnell, P.A. 1992. *Handey: A Robot Task Planner*, MIT Press, Cambridge, MA.
- Lynch, K., and Mason, M. T. 1994. Stable pushing: mechanics, controllability and planning. *Algorithmic Foundations of Robotics (WAFR94)*, Stanford, CA.
- Nielsen, Ch., and Kavraki, L. 2000. A two-level fuzzy PRM for manipulation planning. *IEEE Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Takamatsu, Japan, October 30–November 5.
- Overmars, M., and Švestka, P. 1994. A probabilistic learning approach to motion planning. *Algorithmic Foundations of Robotics (WAFR94)*, Stanford, CA.
- Sahbani, A., Cortés, J., and Siméon, T. 2002. A probabilistic algorithm for manipulation planning under continuous grasps and placements. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, EPFL, Lausanne, Switzerland, September 30–October 4.
- Siméon, T., Laumond, J.P., and Nissoux, C. 2000. Visibility-based probabilistic roadmaps for motion planning. *Advanced Robotics* 14(6):477–494. (A short version appeared in Proceedings of the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems).
- Siméon, T., Laumond, J.P., van Geem, C., and Cortés, J. 2001. Computer Aided Motion: Move3D within MOLOG. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Seoul, Korea, May 21–26.
- Siméon, T., Cortés, J., Sahbani, A., and Laumond, J.P. 2002. A manipulation planner for pick and place operations under continuous grasps and placements. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Washington, DC, May 11–15.
- Wilfong, G. 1988. Motion planning in the presence of movable obstacles. *Proceedings of the 4th ACM Symposium on Computational Geometry*, Urbana-Champaign, IL, June 6–8.

# Geometric Algorithms for the Conformational Analysis of Long Protein Loops

J. CORTÉS,<sup>1</sup> T. SIMÉON,<sup>1</sup> M. REMAUD-SIMÉON,<sup>2</sup> V. TRAN<sup>3</sup>

<sup>1</sup>LAAS-CNRS, 7 avenue du Colonel-Roche, 31077 Toulouse, France

<sup>2</sup>Centre de Bioingenierie Gilbert Durand, 135 avenue de Ranguel, 31077 Toulouse, France

<sup>3</sup>Unité de Recherches sur la Biocatalyse, 2 rue de la Houssinière, 44322 Nantes, France

Received 23 July 2003; Accepted 27 November 2003

**Abstract:** The efficient filtering of unfeasible conformations would considerably benefit the exploration of the conformational space when searching for minimum energy structures or during molecular simulation. The most important conditions for filtering are the maintenance of molecular chain integrity and the avoidance of steric clashes. These conditions can be seen as geometric constraints on a molecular model. In this article, we discuss how techniques issued from recent research in robotics can be applied to this filtering. Two complementary techniques are presented: one for conformational sampling and another for computing conformational changes satisfying such geometric constraints. The main interest of the proposed techniques is their application to the structural analysis of long protein loops. First experimental results demonstrate the efficacy of the approach for studying the mobility of loop 7 in amylosucrase from *Neisseria polysaccharea*. The supposed motions of this 17-residue loop would play an important role in the activity of this enzyme.

© 2004 Wiley Periodicals, Inc. J Comput Chem 25: 956–967, 2004

**Key words:** conformational search; robotic motion planning; loop closure; steric clash avoidance; long protein loops

## Introduction

Prime techniques in structural investigations require the exploration of the conformational space  $\mathcal{C}$  of a molecule. Conformational search methods<sup>1</sup> explore  $\mathcal{C}$  to identify the stable structures of molecules, which determine their properties and functions. Molecular simulations<sup>2</sup> explore  $\mathcal{C}$  while computing conformational changes on a molecule under modified environmental conditions. The analysis of such changes of the molecular structure is essential for the understanding of many biologic processes.

Because the goal of the conformational search is to find minimum energy structures, the exploration is much more efficient when it is limited to a subset of  $\mathcal{C}$  excluding energetically unacceptable conformations. Conformational changes explored in simulations can occur only if there is not a high energetic barrier to overcome. Therefore, approaches treating these problems will greatly benefit from efficient techniques able to provide samples and paths in  $\mathcal{C}$  that filter most unfeasible conformations.

The conformational analysis of a whole macromolecule is a difficult problem. From a methodological point of view, two stages are usually necessary: the first corresponds to the identification of rigid segments (i.e., secondary structural elements) capable of participating in the molecular framework; the second is devoted to the remaining segments, so-called loops, assumed to be much

more flexible. However, available techniques to predict low-energy conformations of long loops are limited and much less efficient because of the loop flexibility.

When the global molecular architecture is assumed to be known and only portions (loops) are studied separately, the integrity of molecular chains must be maintained. The first and last atoms of the treated segment of a molecular chain must remain bonded with their neighbor atoms. Breaking these bonds requires a high amount of energy. A strong constraint is thus imposed for the conformational exploration. This same constraint is present in the analysis of cyclic molecules. It is often referred to in the literature as the *loop-closure* constraint. Three main kinds of methods can be applied to solve the loop-closing problem (i.e., computing conformations satisfying loop closure): analytic (e.g., refs. 3–5), optimization-based (e.g., refs. 6–8), and database methods (e.g., refs. 9 and 10). The difficulty of this problem increases with the length of the molecular chain, and available techniques are limited, or at least strongly penalized, by this.

In addition to breaking bonds, another large amount of energy is required to get two nonbonded atoms significantly closer than the sum of their van der Waals (vdW) radii. A violation of this condition is called *steric clash*. Feasible conformations of a mo-

Correspondence to: J. Cortés; e-mail: jcortes@laas.fr

lecular segment cannot contain either internal clashes, which we call *self-clashes*, or clashes with atoms of the rest of the molecule. A possible filter for such unacceptable conformations consists of evaluating the repulsive term of the vdW energy and discarding conformations that exceed a given cutoff value.<sup>11</sup> However, this energetic constraint can also be treated by geometric procedures. The use of “clash grids,” computed from the distances between atoms, to perform this filtering was proposed in ref. 12. An interesting alternative is the use of collision detection algorithms applied on a 3D model of the molecule.<sup>13</sup> Obviously, the higher the number of atoms the more critical the efficiency of the technique.

In robotics, the same kinds of constraint appear when treating the motion planning problem.<sup>14</sup> Paths must be computed in the subset of feasible configurations\* of the robot,  $\mathcal{C}_{\text{feas}}$ . The main feasibility condition is collision avoidance. The robot cannot collide with obstacles in the workspace and self-collisions are also forbidden. Besides, when the robotic mechanism contains kinematic loops closure constraints must be considered in the computed motions. Sampling-based motion planning techniques (e.g., refs. 15–17) have been demonstrated to be efficient and general tools in this field. These techniques capture the topology of  $\mathcal{C}_{\text{feas}}$  within data structures (graphs or trees) by performing a random (or quasirandom) exploration of  $\mathcal{C}$  on a model of the robot and its environment.

In recent publications,<sup>18,19</sup> we described efficient algorithms for planning motions of closed-chain mechanisms. In this article, we investigate the adaptation of these techniques to handle molecular models. Although the method could be applied to any molecular segment or cyclic molecule, we are mainly interested in the application to long protein loops.

## Interest in Protein Loops

Loops play key roles in the function of proteins. They are often involved in active and binding sites. Therefore, when predicting a protein structure an accurate loop modeling is necessary for determining its functional specificity.

Modeling loops in proteins is one of the main open problems in structural biology. Comparative modeling methods (see ref. 20 for a survey) often fail in the prediction of protein loop structures when the percentage of sequence identities between known and predicted protein family members is low. Indeed, it is well established that there is no reliable approach for modeling long loops (more than five residues) available at this time.<sup>21</sup>

The alternatives to comparative modeling are *de novo* (or *ab initio*) methods.<sup>22</sup> Such methods carry out a search of low-energy conformations for a given amino acid sequence. Many different approaches have been proposed for modeling protein loops. One of the most developed techniques is described in ref. 23. This reference article also provides a concise survey of loop modeling methods. The accuracy of *de novo* methods mainly depends on the

energy function they use. Therefore, improvements in the results provided by these approaches require the design of fine-energy models. However, progress in the conformational exploration strategies may also be necessary to increase the efficiency of these techniques, which are today computationally expensive.

Even more important than the prediction of stable loop conformations is the determination of the feasible conformational changes. In many enzymes, for example, surface loops undergo conformational changes to catalyze a reaction.<sup>24</sup> Further, loop motions are in general involved in protein interactions. Therefore, introducing loop flexibility into docking approaches is necessary for a more accurate prediction of these interactions.<sup>25</sup>

## Aim of Our Approach

The techniques proposed in this article aim to be new tools for the structural analysis of long polypeptide segments and, in particular, of protein loops. The efficiency of geometric algorithms developed in the field of robotics can relieve conformational exploration approaches of a part of the heavy energetic treatment.

In Section 5, we propose a conformational sampling technique that generates random conformations satisfying loop-closure and clash avoidance constraints. The backbone conformation is first computed by an algorithm that relies on efficient geometric and kinematic procedures. Side-chain conformations are then generated by combining sampling techniques and an effective collision detection algorithm. Families of approaches requiring conformational sampling, such as Monte Carlo algorithms<sup>26</sup> or stochastic roadmap techniques,<sup>27</sup> would directly benefit from such filtered conformations.

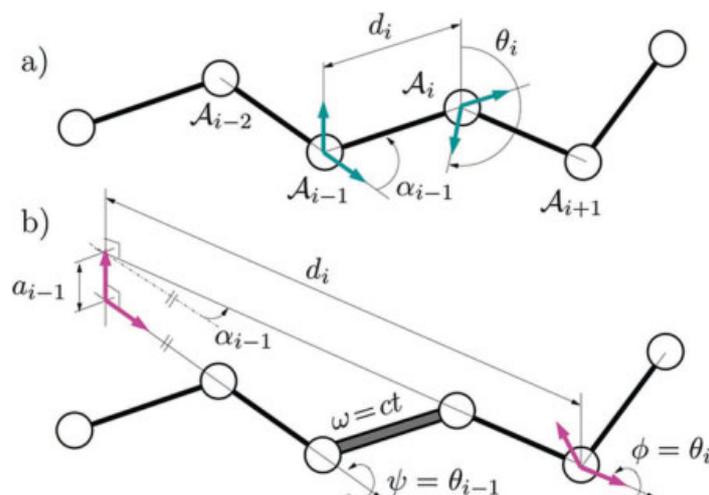
Another interesting feature of our sampling technique is to compute loop conformations avoiding steric clashes with the rest of the protein. Using this technique to compute random samples uniformly distributed in the conformational space will provide useful information about the allowed conformations of the loop in its environment. For instance, this information could be represented in the form of Ramachandran plots,<sup>28</sup> and techniques (e.g., MODELLER<sup>23</sup>) using such statistical distributions could gain in performance.

The geometric analysis can be pushed further. In Section 6, we propose an algorithm to capture the connectivity of the subspace of geometrically feasible conformations. The possible deformations maintaining loop-closure and clash avoidance constraints are explored and encoded in a data structure. Such a data structure would be useful for many existing conformational exploration approaches. Note that a conformational search method sharing similar ideas has been proposed in ref. 29 for small molecules (ligands) under geometric constraints.

## Problem Formulation

The problem is formulated from a robotic point of view. First, the geometric model of the molecule is described. The constraints that must be satisfied during the exploration of the conformational space are then defined.

\*A configuration for a robot is equivalent to a conformation for a molecule. We designate both, the configuration space and the conformational space, by  $\mathcal{C}$ .



**Figure 1.** Molecular chain model. Frames associated (a) with atoms and (b) defining the articulated mechanism. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

### Geometric Model

#### Kinematics-Inspired Model

A molecule is a set of atoms  $\mathcal{A}_i$  partially connected by bonds. A sequence of bonded atoms is called a *molecular chain*. Three parameters, usually called *internal coordinates*, define the relative position of consecutive atoms in a molecular chain: bond lengths, bond angles, and dihedral angles. The widely adopted *rigid geometry assumption* (see ref. 30 as one of the first references) considers that only dihedral angles are variable parameters. Under this assumption, a molecule can be seen as an articulated mechanism with *revolute joints* between bonded atoms. The model of a molecular chain can be built from the internal coordinates using kinematics conventions. We follow the *modified Denavit–Hartenberg* (mDH) convention described in ref. 31. A Cartesian coordinate system  $F_i$  is attached to each atom  $\mathcal{A}_i$  and then the relative location of consecutive frames can be defined by a homogeneous transformation matrix:

$${}^{i-1}T_i = \begin{pmatrix} C\theta_i & -S\theta_i & 0 & 0 \\ S\theta_i C\alpha_{i-1} & C\theta_i C\alpha_{i-1} & -S\alpha_{i-1} & -S\alpha_{i-1}d_i \\ S\theta_i S\alpha_{i-1} & C\theta_i S\alpha_{i-1} & C\alpha_{i-1} & C\alpha_{i-1}d_i \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where  $d_i$  is the bond length between atoms  $\mathcal{A}_{i-1}$  and  $\mathcal{A}_i$ ;  $\alpha_{i-1}$  is the supplement of the bond angle between  $\mathcal{A}_{i-2}$ ,  $\mathcal{A}_{i-1}$ , and  $\mathcal{A}_i$ ;  $\theta_i$  is the dihedral angle formed by atoms  $\mathcal{A}_{i-2}$ ,  $\mathcal{A}_{i-1}$ ,  $\mathcal{A}_i$ , and  $\mathcal{A}_{i+1}$  [see Fig. 1(a)]. C and S represent sines and cosines, respectively.

A molecular chain between atoms  $\mathcal{A}_0$  and  $\mathcal{A}_n$  is then modeled by a kinematic chain,  ${}^1\mathcal{K}_n$ , in which joint variables correspond to dihedral angles. The conformation of the chain is determined by the array  $q$  of the  $\theta_i$ . The kinematic model of a polypeptide segment is composed of a set of chains: the main-chain (the backbone) and the side-chains,

which are built upon it. The conformation of the segment is then specified by an array containing the conformation parameters of the backbone and of all the side-chains.

Often, some portions of molecular models are treated as rigid solids, for instance, peptide units in proteins. The rigid geometry assumption also considers that double-bond torsion angles, such as peptide bonds, are fixed. Hence, the number of frames required in the kinematic modeling is reduced. Figure 1(b) illustrates how the frames corresponding to the mDH parameters are obtained by simple geometric operations when the dihedral angle associated with a peptide bond  $\omega$  is fixed at a given value. Thus, several atoms in each peptide unit have constant coordinates in these frames. As proposed in a recent work,<sup>32</sup> frames only need to be attached to rigid units (called *atomgroups* by the authors). Then, the relative location of atoms in an atomgroup only requires positional coordinates, yielding to a more efficient method for updating conformations.

#### vdW Model

The vdW model consists of a representation of the molecule by the union of solid spheres associated with atoms. A vdW radius is assigned to each atom type. This geometric model of the molecule is the simplest and most ordinary space-filling diagram.<sup>33</sup> In molecular models treated by our approach, such spheres are the mobile *bodies* of the articulated polypeptide segment and the static *obstacles* corresponding to the rest of the atoms in the molecule, which compose what we call the *environment*.

### Geometric Constraints

#### Loop Closure

A loop-closure constraint applied on the kinematic model of a molecular chain  ${}^1\mathcal{K}_n$  fixes the relative location of the frames  $F_0$  and  $F_n$ , which we call *base-frame* and *end-frame*, respectively.

**Algorithm 1:** RANDOMLOOPCONF.

---

```

input : the loop, the rest of the protein
output : the conformation q
begin
   $q_b \leftarrow \text{RANDOMBACKBONECONF}(loop.bkb);$ 
  if not CLASHCHECK( $q_b$ , loop.bkb, protein) then
    if  $q_s \leftarrow \text{GENERATESIDECHAINS}(q_b, loop, protein)$  then
       $q \leftarrow \text{COMPOUNDCONF}(loop, q_b, q_s);$ 
    else return Failure;
  else return Failure;
end

```

---

Therefore, the transform matrix  ${}^0T_n$  is known. This matrix can also be obtained from the sequence of local transformations:

$${}^0T_n = {}^0T_1 T_2 \dots T_{n-1} T_n$$

This equality provides a system of equations, called *closure equations*, where the unknowns are the joint variables  $\theta_i$ . Hence, a relationship must exist between the parameters in  $q$  for satisfying loop closure.

#### Clash Avoidance

Distances between nonbonded atoms that are substantially shorter than the sum of their vdW radii must be avoided. The choice of the limiting contact distance is ambiguous. For our experiments, we model molecules using a percentage (usually 70%) of the vdW radii proposed in ref. 34. Collisions between such reduced vdW spheres must be avoided if they are separated by more than three bonds. This condition must be satisfied between the atoms of the articulated segment and between these atoms and the static atoms of the rest of the molecule.

## Conformational Sampling

Algorithm 1 computes a random conformation of a polypeptide segment (the protein loop) achieving loop-closure and clash avoidance constraints on the 3D model. First, the backbone conformation  $q_b$  is generated. The procedure for obtaining random conformations satisfying closure is explained in the next subsection. These conformations are then tested for clashes of backbone atoms between themselves and with atoms in the environment. Once a feasible conformation for the backbone has been computed, random conformations of the side-chains  $q_s$  are tested. These chains are built iteratively until all of them are free of clashes. The process is explained below.

#### Backbone Conformation with Closure

Obtaining a backbone conformation satisfying loop closure requires the solution of the closure equations mentioned above. Unfortunately, and despite the intensive research in the field, no

efficient general solution is currently available to solve systems of multivariable nonlinear algebraic equations (see ref. 35 for a survey).

It is now well known that, in general, six variables in the closure equations are dependent on the rest (independent variables). Note that six is the minimum number of parameters that allow us to span full-rank subsets of  $SE(3)$  (the position-orientation space in a 3D world).<sup>14</sup> Many articles in computational chemistry and robotics (e.g., refs. 3, 5, and 36–39) propose methods to obtain these six dependent variables as a function of the other parameters. Except for very particular geometries (e.g. regular cyclohexane<sup>40</sup>), only a finite number of solutions exists.

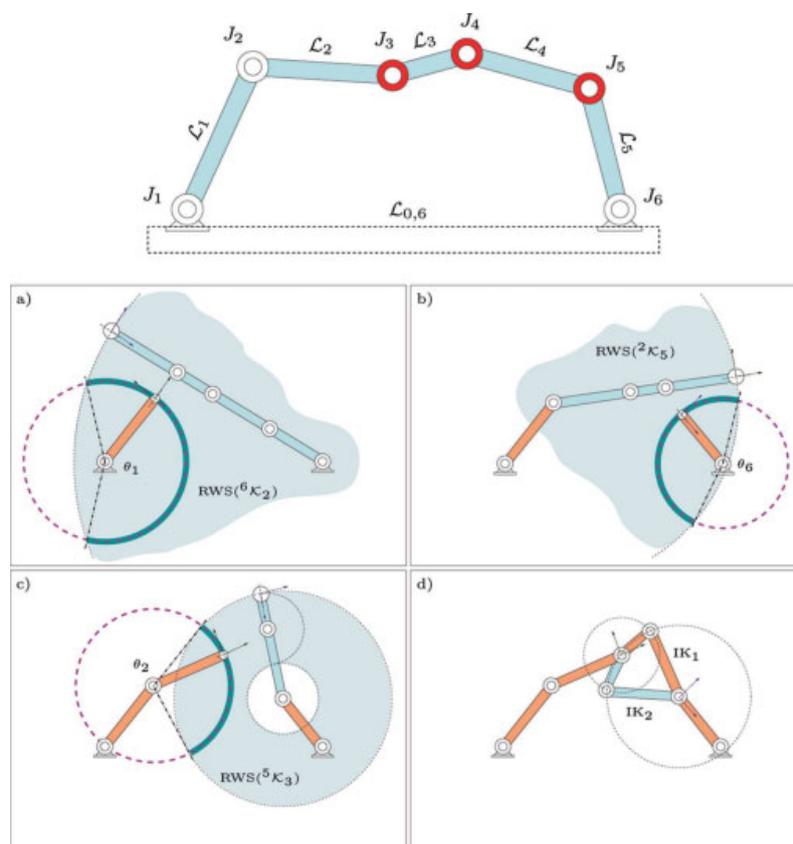
The remaining difficulty is how to obtain values for the independent variables for which a solution of the closure equations exists. In robotics, detailed analytic approaches have been proposed only for planar or spherical closed mechanisms.<sup>41</sup> In computational chemistry, only a few authors have tackled this problem. Decimation approaches and hierarchical decomposition of the closing problem have been proposed for loops with six or more residues.<sup>5</sup> However, for very long loops the efficiency of such methods decreases because closure equations must be solved several times for different fragments of the chain.

We propose an algorithm, called *random loop generator* (RLG), that produces random configurations of articulated mechanisms containing closed chains. This algorithm has demonstrated its efficiency within robotic motion planning techniques.<sup>18,19</sup> The configuration parameters of a closed kinematic chain are separated into two arrays: we call the independent variables of the closure equations the *active variables*  $q^a$  and the dependent variables the *passive variables*  $q^p$ . The RLG algorithm performs a particular random sampling for  $q^a$  that notably increases the probability of obtaining solutions for  $q^p$ .

We next explain the main elements of our approach and how it can be applied to polypeptide backbone segments. Explanations are illustrated on a simple mechanism, the 6R planar linkage in Figure 2. The  $\mathcal{L}_i$  are the rigid bodies and the  $J_i$  the revolute joints connecting them.

#### Loop Decomposition

The choice of the dependent and independent variables in the closure equations is arbitrary. We choose them consecutively in the kinematic chain. Thus, we can refer to a *passive subchain* involving joints whose variables are in  $q^p$  (*passive joints*). Although the passive subchain can be placed anywhere in the closed chain, it is convenient to place it in the middle. In general, the passive subchain is a mechanism with six degrees of freedom. For a polypeptide backbone model under the rigid geometry assumption, only dihedral angles  $\phi$  and  $\psi$  are variable. Therefore, the passive subchain is composed of the backbone of three residues. In the example in Figure 2, three consecutive revolute joints (i.e., a 3R planar mechanism) are sufficient. We have chosen  $J_3$ ,  $J_4$ , and  $J_5$  to be the passive joints of the 6R linkage. Then, the rest of the joints (*active joints* corresponding to  $q^a$ ) can be seen as contained in two *active subchains* rooted on the (fictive) solid on which the base frame and the end frame are fixed ( $\mathcal{L}_{0,6}$  in Fig. 2).



**Figure 2.** Steps of the RLG algorithm performed on a 6R planar linkage. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

### RLG Algorithm

The pseudocode of the algorithm that generates random configurations of a single-loop closed chain (applied to the loop backbone) is synthesized in Algorithm 2. First, the configuration parameters of the active subchains,  $q^a$ , are computed by the function `SAMPLE_` $q^a$  detailed in Algorithm 3. The idea of the algorithm is to progressively decrease the complexity of the closed chain treated at each iteration until only the configuration of the passive subchain,  $q^p$ , remains to be solved. The two active subchains are treated alternately. The ideal solution should be to sample each

#### Algorithm 2: RANDOMBACKBONECONF.

---

**input** : the backbone  
**output** : the conformation  $q_b$   
**begin**  
 $q^a \leftarrow \text{SAMPLE\_}q^a(\text{backbone});$   
**if**  $q^p \leftarrow \text{COMPUTE\_}q^p(\text{backbone}, q^a)$  **then**  
 $q_b \leftarrow \text{COMPOUNDCONF}(\text{backbone}, q^a, q^p);$   
**else return Failure;**  
**end**

---

joint variable from the subset of values, which we call *closure range*, satisfying the closure equations. However, computing this subset is as difficult as solving the general closure equations. Thus, an approximation is used. This approximation must be conservative to guarantee a complete solution (i.e., no region of the subspace satisfying closure constraints is excluded from the sam-

#### Algorithm 3: SAMPLE\_ $q^a$ .

---

**input** : the backbone  
**output** : the active variables  $q^a$   
**begin**  
 $(J_b, J_e) \leftarrow \text{INITSAAMPLER}(\text{backbone});$   
**while not** `ENDACTIVECHAIN`(backbone,  $J_b$ ) **do**  
 $I_c \leftarrow \text{COMPUTECLOSURERANGE}(\text{backbone}, J_b, J_e);$   
**if**  $I_c = \emptyset$  **then go to** line 1;  
 $\text{SETJOINTVALUE}(J_b, \text{RANDOM}(I_c));$   
 $J_b \leftarrow \text{NEXTJOINT}(\text{backbone}, J_b);$   
**if not** `ENDACTIVECHAIN`(backbone,  $J_e$ ) **then**  
 $\text{SWITCH}(J_b, J_e);$   
**end**

---

pling). More details about how to obtain the approximated closure range are given in the next subsection. The closure range of the joint variable treated at one iteration depends on the configuration of the previously treated joints. Hence, this subset must be recomputed for all joints (except the first treated one) in the generation of each new configuration. Because of the conservative nature of the approach, it is possible to obtain an empty set. In this case, the process is restarted.

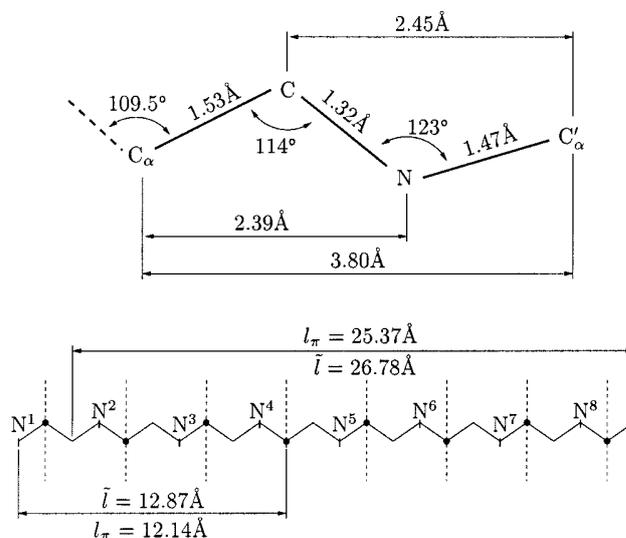
Figures 2(a)–2(c) illustrate how the values of  $\theta_1$ ,  $\theta_6$ , and  $\theta_2$  (the active variables  $q^i$ ) are generated for the 6R linkage. For each joint variable, the estimation of the closure range is computed and a random value is sampled inside this set. Figure 2(d) shows the two solutions of the closure equations for the passive subchain. In this case, these solutions are obtained by simple trigonometric operations. The solution for the passive subchain in the polypeptide backbone model is treated below.

#### Computing Closure Range

The problem can be formulated as follows. Given a closed kinematic chain  ${}^b\mathcal{H}_e$  involving joints from  $J_b$  to  $J_e$  (we consider  $b < e$  in this explanation), two open kinematic chains are obtained by breaking the body  $\mathcal{L}_b$  between  $J_b$  and  $J_{b+1}$ . A suitable break point is the physical placement of  $J_{b+1}$ , but any other point can be chosen. A frame  $F_C$  associated with this break point can be seen as the end frame of both open chains. The closure range of the joint variable corresponding to  $J_b$ ,  $\theta_b$ , is the subset of values for which  $F_C$  is reachable by the open chain  ${}^e\mathcal{H}_{b+1}$ . In general, the exact solution to this problem is extremely complex. Most works in the robot kinematics literature are limited to particular instances (e.g., refs. 42 and 43). For our purpose, a simple and fast method is preferred to a more accurate but slower one. We solve the problem only considering positional reachability.

Because  $J_b$  is a revolute joint, the origin of  $F_C$  describes a circle around its axis. The approximation of the closure range is obtained by the intersection of this circle with a volume (surface for the planar case in Fig. 2) bounding the region mapped by the origin of  $F_C$  attached to the chain  ${}^e\mathcal{H}_{b+1}$ , which is called the *reachable workspace* (RWS) in robotics. This bounding volume is contained between two concentric spheres (circles) centered at the origin of the base frame and whose radii are the maximum and minimum extension of the chain,  $r_{\text{ext}}$  and  $r_{\text{int}}$ . For a general mechanism, obtaining these radii requires the solution of complex optimization problems. If the appropriate (even if computationally slow) method is available, it can be used in a precomputing phase. However, simpler particular solutions can be adopted for particular classes of mechanisms. The solution is straightforward for the planar linkage in our example. The regions designated as RWS in Figures 2(a)–2(c) represent such bounding surfaces at different steps of the algorithm.

In the application to molecular models, frames  $F_C$  are the frames attached to atoms. Particularities in the geometry of polypeptide backbones allow the design of a simple approximated method to compute the spheres bounding RWS. For chains containing more than three residues (which is the size of the passive subchain),  $r_{\text{int}}$  can be simply considered zero without decreasing the performance of the technique.



**Figure 3.** Maximum extension of polypeptide backbone with Pauling–Corey geometry.

The maximum distance between the extreme atoms of a segment of polypeptide backbone\* is often obtained for a conformation with all the dihedral angles at  $\pi$ . We call this length  $l_\pi$ . However, this assumption is not always true, in particular if a slight rotation around peptide bonds is allowed. An upper bound of the maximum is required for guaranteeing completeness. This upper bound  $\hat{l}$  is the sum of the distances between consecutive  $C_\alpha$  atoms (i.e., the length of peptide units). Obviously, when the chain begins or ends with a fragment of a peptide unit (i.e., only one or two of the three concerned atoms in the backbone are contained in the chain) the length of this portion is added. Instead of using a constant value,  $r_{\text{ext}}$  is sampled from a distribution between  $l_\pi$  and  $\hat{l}$  each time this dimension is required in the process. We suggest using a Gaussian distribution with  $\mu = l_\pi$  and  $\sigma^2 = 1$ . This increases the efficiency of the approach while keeping completeness.

This approximated method to obtain  $r_{\text{ext}}$  is not dependent on a particular kind of geometry. It can be applied on standard models or to structures acquired from the Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>). Figure 3 illustrates the application to backbone segments with standard Pauling–Corey geometry.<sup>44</sup>

#### General 6R Inverse Kinematics

The kinematic model of the three-residue backbone corresponding to the passive subchain in our approach can be seen as a 6R manipulator with general geometry.<sup>45</sup> Obtaining the conformation of a serial manipulator, given the location of the base frame and the end frame (i.e., solving the closure equations), is known in robotics as the *inverse kinematics problem*.

The method we use to solve the general 6R inverse kinematics problem is inspired by the work of Lee and Liang.<sup>36</sup> The principle

\*Without proline. This case, studied apart, is not detailed in this article.

of the method is described in ref. 39.\* The algebraic elimination of variables starts in a way similar to that used in related works (e.g., refs. 37 and 38). However, Renaud goes further in the elimination process, arriving at an  $8 \times 8$  quadratic polynomial matrix in one variable instead of the  $12 \times 12$  matrix in the referred methods. The problem can then be treated as a generalized eigenvalue problem (as previously proposed in ref. 38), for which efficient and robust solutions are available.<sup>46</sup> Another important advantage of the method in relation to all previous approaches is that it requires a minimum number of divisions in the elimination process. In particular, divisions by zero are avoided to guarantee robustness.

Note that the general 6R mechanism can have up to 16 inverse kinematic solutions. Thus, several sets of values of the passive variables  $q^p$  satisfy loop-closure equations for a given value of the active variables  $q^a$ . Each backbone conformation obtained by composing  $q^a$  with the different  $q^p$  is treated by the algorithm RANDOMLOOPCONF (Algorithm 1).

### Clashes and Side-Chain Conformation

#### Collision Detection Algorithms

A collision detection algorithm determines if contacts or penetrations exist between 3D bodies. They are important tools in computational geometry and robotics.<sup>47,48</sup> Collision detection is the most computationally expensive process in sampling-based motion planning techniques. Thus, effective algorithms have been developed in this field to try to minimize this cost.

In our current implementation of the approach, clashes in a sampled conformation are checked by a generic collision detection algorithm,<sup>49</sup> which operates well within geometrically complex 3D scenes.

#### Sampling Side-Chain Conformation

The conformation of the side-chains is built upon a feasible backbone conformation. These side-chain conformations are generated by randomly sampling the side-chain dihedral angles and tested until a collision-free solution is found. A progressive construction is carried out. Instead of rebuilding all the side-chains when the collision test is positive, only the conformation of clashing side-chains is resampled. The resampling and collision detection process is performed following an arbitrary order of the side-chains, intending to prevent a privileged conformational sampling. When two side-chains collide together, but self-clashes or clashes with the backbone and the rest of the protein do not exist, only one of them will be resampled. The process is iterated a certain number of times before returning that a clash-free conformation of the side-chains cannot be found.

\*The author is currently working on an extended version with full technical details.

#### Algorithm 4: ExploreByRRT.

---

```

input : the loop, the rest of the protein,  $q_{init}$ 
output : the tree  $\mathcal{T}$ 
begin
   $G \leftarrow \text{INITTREE}(q_{init});$ 
   $n_{fail} \leftarrow 0;$ 
  while not STOPCONDITION( $\mathcal{T}$ ) do
     $q_{rand} \leftarrow \text{GUIDEDRANDOMCONF}(loop);$ 
     $q_{near} \leftarrow \text{NEARESTNEIGHBOR}(q_{rand}, \mathcal{T});$ 
     $q_{feas} \leftarrow q_{near};$ 
     $state \leftarrow OK;$ 
    while  $state = OK$  do
       $q_{step} \leftarrow \text{MAKESTEP}(q_{feas}, q_{rand});$ 
      if FEASIBLECONF( $q_{step}$ ) then  $q_{feas} \leftarrow q_{step};$ 
      else  $state \leftarrow FAIL;$ 
    if not TOO SIMILARCONF( $q_{near}, q_{feas}$ ) then
       $q_{new} \leftarrow \text{INTERMEDIATECONF}(q_{near}, q_{feas});$ 
      GROWTREE( $q_{new}, q_{near}, \mathcal{T}$ );
       $n_{fail} \leftarrow 0;$ 
    else  $n_{fail} \leftarrow n_{fail} + 1;$ 
end

```

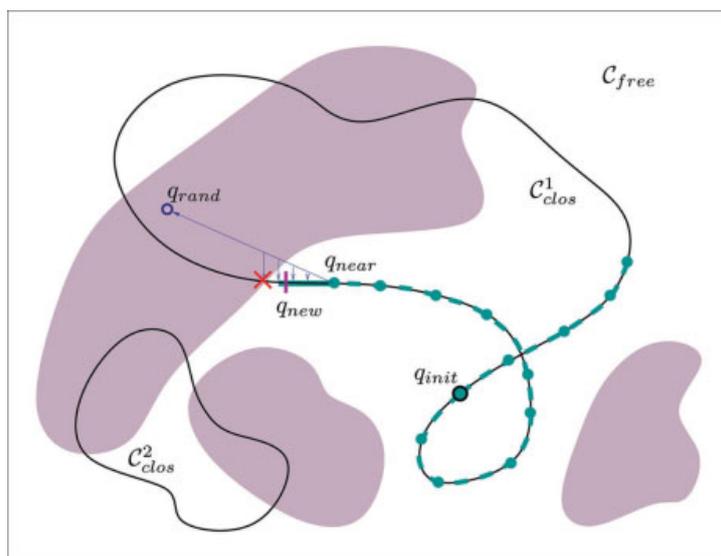
---

## Conformational Space Exploration

### Sampling-Based Motion Planning Techniques

Sampling-based motion planning techniques appeared in robotics as an alternative to exact approaches<sup>14</sup> that cannot be applied to high-dimensional configuration spaces. In particular, algorithms based on the *probabilistic roadmap* (PRM) approach (e.g., refs. 15 and 16) have mostly been developed. The general PRM principle is to construct a graph (roadmap) that captures the topology of the feasible subset of robot configurations,  $\mathcal{C}_{feas}$ . The nodes of this graph are randomly sampled configurations satisfying intrinsic conditions in this subset (e.g., collision avoidance). The edges are short feasible paths (*local paths*) linking “nearby” nodes. Other families of methods aim to efficiently solve *single planning queries* instead of covering the whole search space. The *rapidly-exploring random tree* (RRT)<sup>17</sup> is a data structure and sampling scheme to quickly search high-dimensional constrained spaces.  $\mathcal{C}_{feas}$  is explored by one or two trees rooted at the start and/or goal configurations. The exploration is biased by sampling points in  $\mathcal{C}$  and incrementally pulling the search tree(s) toward them.

This section treats the application of these techniques onto geometric models of molecules. Let us call  $\mathcal{C}_{clos}$  the subset of the conformations satisfying the loop-closure constraints and  $\mathcal{C}_{free}$  the subset of clash-free conformations.  $\mathcal{C}_{feas} = \mathcal{C}_{clos} \cap \mathcal{C}_{free}$  is the subset of the geometrically feasible conformations to be explored. Obviously, not every conformation in  $\mathcal{C}_{feas}$  is energetically acceptable, but a significant number of high-energy structures are excluded from this subset. We assume that  $\mathcal{C}_{feas}$  contains all the energetically feasible conformations:  $\mathcal{C}_{low E} \subset \mathcal{C}_{feas} \subset \mathcal{C}$ .



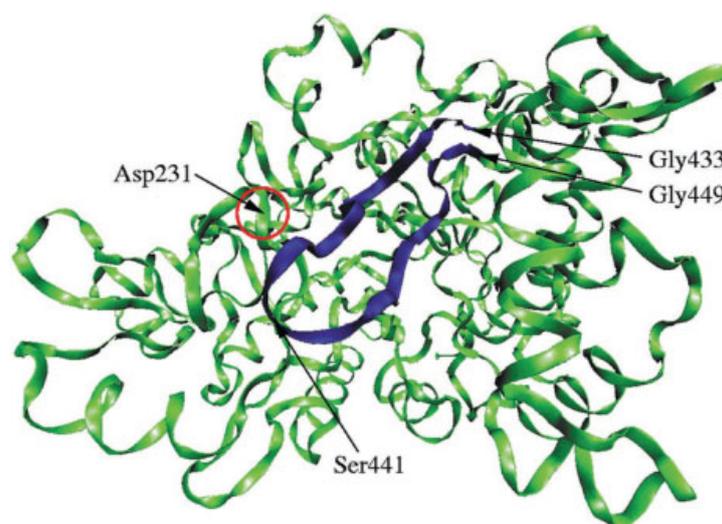
**Figure 4.** Incremental exploration of  $\mathcal{C}_{\text{feas}}$  using an RRT-like technique.

#### Incremental Search Keeping Constraints

We next explain an algorithm to carry out the incremental search of  $\mathcal{C}_{\text{feas}}$  using an RRT-like technique<sup>17</sup> extended to handle the geometric constraints in our problem. Algorithm 4 gives the pseudocode and Figure 4 illustrates the exploration in a simple 2D example. The darker regions in the figure correspond to conformations with steric clashes,  $\mathcal{C}_{\text{free}}$  being the rest of the space. In general, conformations satisfying closure (in  $\mathcal{C}_{\text{clos}}$ ) are grouped into different disjoint continuous manifolds.<sup>50</sup> We considered two manifolds  $\mathcal{C}_{\text{clos}}^1$  and  $\mathcal{C}_{\text{clos}}^2$  for this illustration.

The starting point  $q_{\text{init}}$  can be a randomly sampled feasible conformation (e.g., generated by the technique explained above) or

a known conformation (e.g., acquired from the PDB). For executing an expansion step of the RRT, a random conformation  $q_{\text{rand}}$  is first sampled in  $\mathcal{C}$ .  $q_{\text{rand}}$  need not satisfy either closure or clash avoidance constraints. This conformation is only used as a local goal for the exploration. Nevertheless, we have experimentally shown that a *guided-random sampling* generating  $q_{\text{rand}}$  close to the subset satisfying closure equations improves the process (i.e., a wider portion of the space is explored in less time) in relation to a uniform random sampling.<sup>51</sup> For this, the configuration parameters corresponding to the independent variables of the closure equations,  $q^a$ , are generated by the function `SAMPLE_` $q^a$  (Algorithm 3), explained above. Then, the nearest node in the current tree,  $q_{\text{near}}$ ,



**Figure 5.** Structure of AS from *N. polysaccharea*.

is selected using a distance metric in  $\mathcal{C}$ . A new conformation  $q_{\text{feas}}$  is iteratively pulled from  $q_{\text{near}}$  toward  $q_{\text{rand}}$ . The pulled conformation must remain in the feasible subset. The closure constraint is maintained as follows. A conformation  $q_{\text{step}}$  is obtained by interpolating  $q_{\text{near}}$  and  $q_{\text{rand}}$  following a law (e.g., linear interpolation). The closure equations are then solved for the passive variables of the backbone conformation (called  $q^p$  above). If the solution in the same manifold as  $q_{\text{near}}$  exists, then the conformation satisfying closure,  $q'_{\text{step}}$ , is checked for clashes. The process goes on until one of the feasibility conditions is violated. The new node of the tree,  $q_{\text{new}}$ , is an intermediate conformation between  $q_{\text{near}}$  and the last obtained  $q_{\text{feas}}$ . We use a Gaussian sampling between  $q_{\text{feas}}$  and  $q_{\text{near}}$  to obtain it.

Several criteria can be adopted for stopping the exploration. The simplest one is to build the tree until it contains a given number of nodes. The drawback is that this criterion is not related to a coverage of the explored region. We believe that an estimation of this coverage could be deduced from the number  $n_{\text{fail}}$  of consecutive times the algorithm fails when trying to expand the tree. A similar relationship has been demonstrated in related methods.<sup>16</sup>

While the infinite solutions of the global inverse kinematics problem are grouped into different disjoint continuous manifolds and collision-free portions of each manifold can be also disjoint, the explained algorithm can explore only a region in  $\mathcal{C}_{\text{feas}}$ . Several starting points are required for exploring the different connected components of  $\mathcal{C}_{\text{feas}}$ . An algorithm combining RRT and PRM techniques could be used for the exploration of the whole subset.

#### Exploration with Flexible Geometry

Considering fixed values for bond lengths, bond angles, and double-bond torsion angles is a well-accepted assumption that reduces the complexity of the structural analysis of molecules. However, it implies a severe restriction for conformational space exploration.<sup>52</sup>

The rigid geometry assumption can be relaxed by allowing a slight variation of these parameters within given intervals. Handling these new variables is not a hard problem for our exploration algorithm, proceeding as follows. To generate a conformation  $q_{\text{rand}}$ , parameters  $d$ ,  $\alpha$ , and  $\omega$  (see above) are first randomly sampled within the defined intervals. Then, the approach explained in the Conformational Sampling section can be used. In the incremental variation of the selected conformation  $q_{\text{near}}$  toward  $q_{\text{rand}}$ , the new parameters are treated like the rest of the (nonpassive) variables (i.e., they are interpolated following a given law).

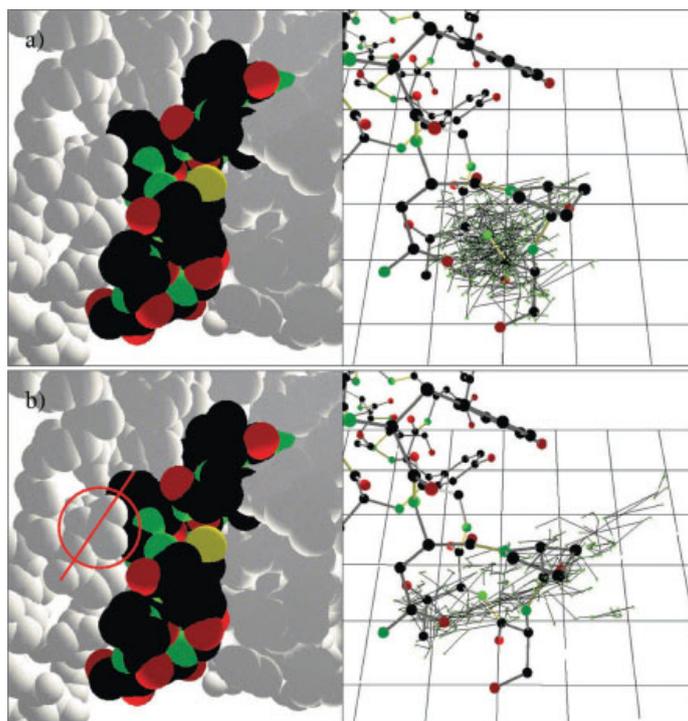
#### First Results: Loop 7 Motions of Amylosucrase from *Neisseria Polysaccharea*

Amylosucrase (AS) is a glucansucrase that catalyzes the synthesis of an amylose-like polymer from sucrose. In the Carbohydrate-Active enZYme database (CAZy) (<http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.htm>), this enzyme is classified in family 13 of glucoside-hydrolases (GH), which mainly contains starch-converting enzymes (hydrolases or transglycosidases). Remarkably, this enzyme is the only polymerase acting on sucrose substrate reported in this family, all the other glucansucrases being gathered in GH

family 70. Which structural features are involved in AS specificity is an important fundamental question. Indeed, the structural similarity of AS to family 13 enzymes is high. The 3D structure reveals an organization in five domains.<sup>53</sup> Three of them are commonly found in family 13: a catalytic  $(\beta/\alpha)_8$  barrel domain, a B domain between  $\beta$ -strand 3 and  $\alpha$ -helix 3 (loop 3), and a C terminal Greek key domain. Two additional domains are found in AS only: a helical *N*-terminal domain and a domain termed B', formed by an extended loop between  $\beta$ -strand 7 and  $\alpha$ -helix 7. Domain B' partially covers the active site located at the bottom of a pocket and is mainly responsible for this typical architecture. Recently, cocrystallization of AS with maltoheptaose revealed the presence of two maltoheptaose binding sites, the first (OB1) in the main access channel to the active site and a second (OB2) at the surface of domain B'. Soaking AS crystals with sucrose also revealed the presence of a second sucrose binding site (SB2) different from the active site initially identified.<sup>54</sup> The comparison of the various structures obtained suggests that motion of the 17-residue fragment of domain B' starting at residue Gly433 and ending at residue Gly449, consecutive to oligosaccharide binding, could facilitate sucrose translocation from SB2 to the active site. In the following part, this fragment will be called loop 7. This loop could play a pivotal role responsible for the structural change and the polymerase activity. In this context, molecular simulation of loop 7 motion appears to be crucial to gain new insight into AS structure–function relationships.

Figure 5 shows the crystallographic structure of AS and the location of the residues we mention in the following paragraphs. The model for our tests was created from the PDB file containing this structure (PDB ID: 1G5A), considering loop 7 as an articulated mechanism and the rest of the atoms as static elements. Atoms were modeled with 70% of their vdW radii. Images on the left in Figure 6 represent the articulated vdW model of the loop and a portion of its environment. Under our modeling assumptions, the results of the geometric exploration showed that only slight conformational variations of the loop are possible if the backbone integrity is maintained and steric clashes are avoided. The image on the right in Figure 6(a) shows the skeleton of the articulated segment and a representation of one of the RRTs computed for this test. Nodes of the RRT are graphically represented by the positions explored by the  $C_\alpha$  atom of Ser441, the middle residue of the loop. This result contradicts presupposed significant loop fluctuations. Of course, our approach is not deterministic and therefore we cannot guarantee that such a motion does not exist. However, after several exhaustive tests we can assert that the probability of its existence is low. The average size of the constructed RRTs is 1000 nodes, for which about 4000 random conformations and 20,000 complete collision tests were necessary. The average computing time was 1 h.\* Note that computing time is mostly spent in collision detection. The generation of random conformations is fast. For this loop, computing a conformation-satisfying closure (including the update of all the frames and atom positions) takes less than 0.1 s with a nonoptimized implementation. The conformational sampling used by the exploration algorithm (i.e., guided-

\*Tests were performed using a Sun Blade 100 workstation with a 500-MHz UltraSPARC-IIe processor.



**Figure 6.** Exploration (a) with and (b) without the side-chain of Asp231. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

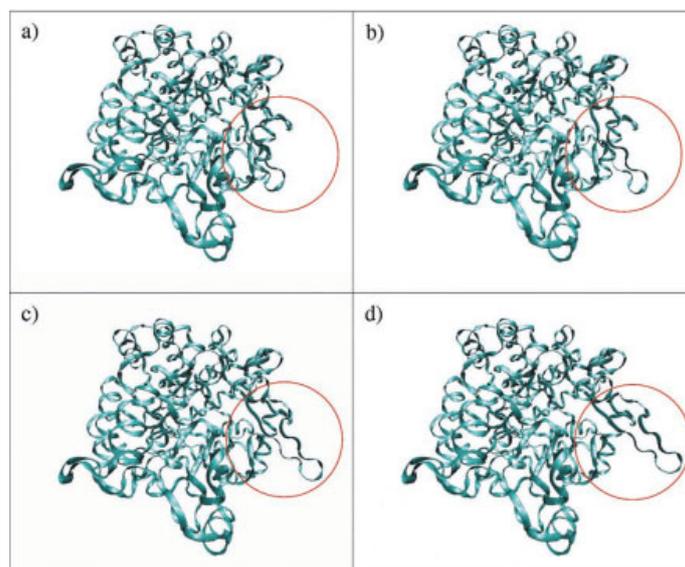
random sampling of  $q^a$  without solving the closure equations for  $q^p$ ) demands only about 0.01 s per conformation.

Several structural elements, and mainly loop 3 (residues 183–262), restrain the mobility of loop 7. Residue Asp231 was identified as the main “geometric lock” responsible for the loop 7 enclosing. The side-chain of this residue was removed from the model to simulate a possible conformational change of this chain or even of the whole loop 3. The conformational exploration in this case showed that the loop is able to effect the expected motions keeping geometric constraints. The  $C_\alpha$  atom of Ser441 can be dislocated more than 9 Å from its crystallographic position. Several tests were performed to see if the random nature of the approach could have an important influence on the nature of the results. Similar motions were obtained for all of them. The loop moves almost as a rigid body with hinges at the extreme residues. Considerable variations of the backbone dihedral angles are concentrated in residues 433–436 and 446–449. Figure 6(b) shows the representation of the RRT constructed in one of these tests. The images in Figure 7 correspond to four frames of the conformational change encoded in the RRT. Therefore, an “opening/closing” mechanism similar to other enzymes (e.g., refs. 24 and 55), termed *conformational gating*, is suspected for this loop. The role that residue Asp231 could play in this mechanism is being investigated. Directed mutagenesis experiments, replacing residue Asp231 by glycine, are currently being developed.

## Discussion and Prospects

We proposed geometric techniques aimed at providing powerful filters for conformational sampling and search methods. Our solution to the loop-closing problem is computationally efficient and its performance is only slightly affected by the length of the molecular chain. To the best of our knowledge, only the CCD algorithm recently proposed in ref. 8 offers a similar performance. While this optimization-based algorithm converges to an approximate closure solution starting from a nearly open conformation of the loop, our RLG sampling method computes exact solutions to the closure problem. Also, one disadvantage pointed out in ref. 8 is that the CCD optimization technique, which considers one degree of freedom at a time, may favor large changes in the first residues of the loop. By comparison, the random strategy of RLG produces more uniformly distributed samples, better suited for exploration of the conformational space.

Our algorithms are currently implemented within the motion planning platform Move3D<sup>56</sup> developed at LAAS for robotics applications. No particular consideration has been given to reducing computation time in the present implementation, which is aimed at demonstrating the efficacy of the proposed techniques. More extensive experimental tests and performance comparisons remain for future work, based on an optimized version taking advantage of the specifics of molecular models. We started the development of such an optimized version and a standalone library that could be made accessible to the scientific community.



**Figure 7.** Simulated conformational gating of loop 7 in AS. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

Concerning the avoidance of steric clashes, collision detection algorithms combined with smart sampling techniques constitute an attractive alternative to methods producing optimization-based rearrangements. We are developing a tailored collision detection algorithm for molecular models that should perform faster than the generic checker currently used. In addition, a different progressive process for building backbone conformations is going to be tried. In contrast to the described sampling approach, clashes between the backbone atoms and the static environment will be checked after each step of the RLG algorithm.

In our current implementation, values for all variable dihedral angles in the side-chains and backbone are randomly sampled in the interval  $(-\pi, \pi]$ . As in other related techniques, our approach could handle information on the statistically preferred values of these angles (e.g., from Ramachandran plots by residue type). Using this information, many local steric clashes should be implicitly avoided.

Concerning the exploration technique, we are working on a method for pruning branches of the RRT to decrease the size of this data structure, thus increasing the speed of the search process. Preliminary results using a visibility-based heuristic<sup>16</sup> seem promising.

The algorithms presented in this article treat conformations of a molecular segment in a static environment. The extension of these algorithms to handle the flexibility of side-chains in this environment could be done without difficulty. Handling several loops that share the same region of the space (e.g., antibody hypervariable loops<sup>57</sup>) is an interesting extension we expect to develop.

The first results of the application of our robotic approach to molecular models show the potential of this technique. A fast geometric analysis can help find the answer to important biochem-

ical questions such as: what are the crucial residues in the biochemical reaction? and what are the possible conformational changes?

Although our next goal is to improve this geometrically constrained exploration, the final aim is to incorporate the energetic analysis into the incremental search technique. An energy function can easily be integrated into this kind of exploration algorithm. Indeed, impressive results have been obtained by conformational search methods inspired by sampling-based motion planning techniques applied to computer-assisted drug design,<sup>29</sup> protein folding,<sup>27,58</sup> and ligand–protein docking.<sup>59,60</sup> Given this energy function, geometrically feasible conformations generated by our approach could be evaluated and labeled, and then only the subset of the conformational space  $\mathcal{C}_{\text{low } E}$  below a certain energetic limit should be explored.

## Acknowledgments

The collaboration and advice of the following people has been essential for carrying out this work: Gwénaëlle André (INRA, Nantes), Paul Bates (CRUK, London), Patrick Danes (LAAS-CNRS, Toulouse), David Guieysse (INSA, Toulouse), Marc Renaud (LAAS-CNRS, Toulouse), Lluís Ros (IRI-CSIC, Barcelona), and Vicente Ruiz (IRI-CSIC, Barcelona). This work has been partially supported by the interdisciplinary CNRS project BioMove3D and the European project IST-37185 MOVIE.

## References

1. Leach, A. R. *Molecular Modeling: Principles and Applications*; Longman: White Plains, NY, 1996; Chapter 8.

2. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: New York, 1996.
3. Gō, N.; Scheraga, H. A. *Macromolecules* 1970, 3, 178–187.
4. Manocha, D.; Zhu, Y.; Wright, W. *CABIOS* 1995, 11, 71–86.
5. Wedemeyer, W. J.; Scheraga, H. A. *J Comput Chem* 1999, 20, 819–844.
6. Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, C.; Levinthal, C. *Biopolymers* 1987, 26, 2053–2085.
7. Zheng, Q.; Rosenfeld, R.; Vajda, S.; DeLisi, C. *J Comput Chem* 1993, 14, 556–565.
8. Canutescu, A. A.; Dunbrack, R. L. Jr. *Protein Sci* 2003, 12, 963–972.
9. Oliva, B.; Bates, P. A.; Querol, E.; Aviles, F. X.; Sternberg, M. J. E. *J Mol Biol* 1997, 266, 814–830.
10. van Vlijmen, H. W. T.; Karplus, M. *J Mol Biol* 1997, 267, 975–1001.
11. Bruccoleri, R. E.; Karplus, M. *Biopolymers* 1987, 26, 137–168.
12. Moulton, J.; James, M. N. G. *Proteins* 1986, 1, 146–163.
13. Lotan, I.; Schwarzer, F.; Halperin, D.; Latombe, J. C. In *Proceedings of the 18th ACM Symposium on Comp Geom*; Barcelona, 2002; p 43–52.
14. Latombe, J. C. *Robot Motion Planning*; Kluwer Academic: Boston, 1991.
15. Kavragi, L. E.; Švestka, P.; Latombe, J. C.; Overmars, M. *IEEE Trans Rob Autom* 1996, 12, 566–580.
16. Siméon, T.; Laumond, J. P.; Nissoux, C. *Adv Rob J* 2000, 14, 477–494.
17. LaValle, S. M. In *Control Problems in Robotics*; Bicchi A.; Christensen H. I.; Prattichizzo D., Eds.; Springer-Verlag: Berlin, 2002; p 19–37.
18. Cortés, J.; Siméon, T.; Laumond, J. P. In *Proceedings of the IEEE International Conference on Robotics & Automation*; IEEE: Washington D.C., 2002; p 2141–2146.
19. Cortés, J.; Siméon, T. In *Proceedings of the IEEE International Conference on Robotics & Automation*; IEEE: Taipei, 2003; p 4354–4359.
20. Contreras-Moreira, B.; Fitzjohn, P. W.; Bates, P. A. *Appl Bioinfo* 2002, 1(4), 177–190.
21. Tramontano, A.; Leplae, R.; Morea, V. *Proteins* 2001, Suppl. 5, 45, 22–38.
22. Baker, D.; Šali, A. *Science* 2001, 295, 93–96.
23. Fiser, A.; Do, R. K.; Šali, A. *Protein Sci* 2000, 9, 1753–1773.
24. Osborne, M. J.; Schnell, J.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. *Biochemistry* 2001, 40, 9846–9859.
25. Janin, J.; Henrick, K.; Moulton, J.; Eyck, L. T.; Sternberg, M. J. E.; Vajda, S.; Vakser, I.; Wodak, S. J. *Proteins* 2003, 52, 2–9.
26. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J Chem Phys* 1953, 21, 1087–1092.
27. Apaydin, M. S.; Brutlag, D. L.; Guestin, C.; Hsu, D.; Latombe, J. C. In *Proceedings of RECOMB*; Washington D.C., 2002; p 12–21.
28. Ramachandran, G. N.; Sasisekharan, V. *Adv Prot Chem* 1968, 23, 283–438.
29. LaValle, S. M.; Finn, P. W.; Kavragi, L. E.; Latombe, J. C. *J Comput Chem* 2000, 21, 731–747.
30. Scott, R. A.; Scheraga, H. A. *J Chem Phys* 1966, 44, 3054–3069.
31. Craig, J. J. *Introduction to Robotics: Mechanics and Control*; Addison-Wesley: Reading, MA, 1989; Chapter 3.
32. Zhang, M.; Kavragi, L. E. *J Chem Info Comp Sci* 2002, 42, 64–70.
33. Edelsbrunner, H. In *Robotics: The Algorithmic Perspective* (WAFR1998); Agarwal P.; Kavragi L. E.; Mason M.; Eds.; A. K. Peters: Boston, 1998; p 265–277.
34. Bondi, A. *J Phys Chem* 1964, 68, 441–451.
35. Nielsen, J.; Roth, B. *Proc NATO Adv Study Inst Comp Meth Mech* 1997, 1, 233–252.
36. Lee, H. Y.; Liang, C. G. *Int J Mech Mach Theory* 1988, 23, 209–217.
37. Raghavan, M.; Roth, B. In *Proceedings of the Int Symp Rob Res*; Tokyo, 1989; 314–320.
38. Manocha, D.; Canny, J. *IEEE Trans Rob Autom* 1994, 10, 648–657.
39. Renaud, M. In *Proceedings of the Int Conf Mech Design & Prod Cairo*, 2000; p 15–25.
40. Crippen, G. M. *J Comput Chem* 1992, 13, 351–361.
41. Celaya, E.; Torras, C. In *Computational Kinematics*; Angeles J.; Hommel G.; Kovács P., Eds., Kluwer Academic: Dordrecht, 1993; p 85–94.
42. Ricard, R.; Gosselin, C. M. *J Mech Design* 1998, 120, 269–278.
43. Merlet, J. P. *J Intell Rob Sys* 1995, 13, 143–160.
44. Pauling, L. *The Nature of the Chemical Bond*; Cornell University Press: Ithaca, NY, 1960.
45. Angeles, J. *Fundamentals of Robotic Mechanical Systems*; Springer-Verlag: New York, 2003.
46. Kwakernaak, H.; Sebek, M. *IEEE Trans Autom Control* 1994, 39, 315–328.
47. Lin, M.; Manocha, D.; Cohen, J.; Gottschalk, S. In *Algorithms for Robotic Motion and Manipulation* (WAFR1996); Laumond J.-P.; Overmars M. H., Eds., A. K. Peters: Boston, 1997; p 129–141.
48. Jiménez, P.; Thomas, F.; Torras, C. In *Robot Motion Planning and Control*; Laumond J.-P., Ed.; Springer-Verlag: Berlin, 1998; p 305–343.
49. van Geem, C.; Siméon, T. *LAAS Report 01073*; Toulouse, 2001.
50. Burdick, J. W. In *Proceedings of the IEEE Conference on Robotics and Automation*, Scottsdale, AZ, 1989; p 264–270.
51. Cortés, J. *PhD Thesis*; Institut National Polytechnique de Toulouse, 2003.
52. Bruccoleri, R. E.; Karplus, M. *Biopolymers* 1985, 18, 2767–2773.
53. Skov, L. K.; Mirza, O.; Henriksen, A.; de Montalk, G. P.; Remaud-Simeon, M.; Sarçal, P.; Willemot, R. M.; Monsan, P.; Gajhede, M. *J Biol Chem* 2001, 276, 25273–25278.
54. Skov, L. K.; Mirza, O.; Sprogøe, D.; Dar, I.; Remaud-Simeon, M.; Albenne, C.; Monsan, P.; Gajhede, M. *J Biol Chem* 2002, 277, 47741–47747.
55. Derreumaux, P.; Schlick, T. *Biophys J* 1998, 74, 72–81.
56. Siméon, T.; Laumond, J. P.; Lamiroux, F. In *Proceedings of the IEEE International Conference on Robotics and Automation*; Seoul, 2001; p 25–30.
57. Bruccoleri, R. E.; Haber, E.; Novotny, J. *Nature* 1988, 335, 564–568.
58. Amato, N. M.; Dill, K. A.; Song, G. *J Comp Biol* 2002, 9, 149–168.
59. Apaydin, M. S.; Singh, A. P.; Brutlag, D. L.; Latombe, J. C. In *Proceedings of the IEEE International Conference on Robotics and Automation*; Seoul, 2001; p 932–939.
60. Bayazit, O. B.; Song, G.; Amato, N. M. In *Proceedings of the IEEE International Conference on Robotics and Automation*; Seoul, 2001; p 954–959.

RESEARCH

Open Access

# Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods

Ibrahim Al-Bluwi<sup>1,2</sup>, Marc Vaisset<sup>1,2</sup>, Thierry Siméon<sup>1,2</sup>, Juan Cortés<sup>1,2\*</sup>

From Computational Structural Bioinformatics Workshop 2012  
Philadelphia, PA, USA. 4 October 2012

## Abstract

**Background:** Obtaining atomic-scale information about large-amplitude conformational transitions in proteins is a challenging problem for both experimental and computational methods. Such information is, however, important for understanding the mechanisms of interaction of many proteins.

**Methods:** This paper presents a computationally efficient approach, combining methods originating from robotics and computational biophysics, to model protein conformational transitions. The ability of normal mode analysis to predict directions of collective, large-amplitude motions is applied to bias the conformational exploration performed by a motion planning algorithm. To reduce the dimension of the problem, normal modes are computed for a coarse-grained elastic network model built on short fragments of three residues. Nevertheless, the validity of intermediate conformations is checked using the all-atom model, which is accurately reconstructed from the coarse-grained one using closed-form inverse kinematics.

**Results:** Tests on a set of ten proteins demonstrate the ability of the method to model conformational transitions of proteins within a few hours of computing time on a single processor. These results also show that the computing time scales linearly with the protein size, independently of the protein topology. Further experiments on adenylate kinase show that main features of the transition between the open and closed conformations of this protein are well captured in the computed path.

**Conclusions:** The proposed method enables the simulation of large-amplitude conformational transitions in proteins using very few computational resources. The resulting paths are a first approximation that can directly provide important information on the molecular mechanisms involved in the conformational transition. This approximation can be subsequently refined and analyzed using state-of-the-art energy models and molecular modeling methods.

## Background

Conformational transitions in proteins are generally related to their capacity to interact with other molecules. Their study is therefore essential for the understanding of protein functions. Unfortunately, it is very difficult to obtain this type of dynamic information at the atomic scale using experimental techniques. Modeling protein conformational transitions with conventional computational methods is

also challenging because, in many cases, these transitions are rare, slow events. Standard molecular dynamics (MD) simulations with current computational resources cannot be applied in practice to model large-amplitude (slow time-scale) conformational transitions. Such simulations require variants of MD methods that enhance sampling of rare events or that bias the exploration in a given direction (e.g. [1-5]), or, alternatively, to have access to outstanding computational power [6].

\* Correspondence: [juan.cortes@laas.fr](mailto:juan.cortes@laas.fr)

<sup>1</sup>CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France

Full list of author information is available at the end of the article

Modeling conformational transitions in proteins has motivated the development of specific methods, computationally more efficient than MD simulations. Many of these methods (e.g. [7-9]) are based on the deformation of a trivial initial path between the two given conformations toward the minimum energy path connecting them. Consequently, the performance of these methods is strongly conditioned by the suitability of the initial path. In recent years, methods to model conformational transitions have also been developed on the basis of robot motion planning algorithms [10-13]. Most of these robotics-inspired methods are aimed at providing qualitative information about the conformational transition using few computational resources. For this, they exploit the efficiency of sampling-based exploration algorithms applied to simplified molecular models.

The high dimensionality of the space to be explored is the main difficulty that all computational methods to model protein conformational transitions have to face. Therefore, several approaches have been developed to reduce the dimensionality of the problem (e.g. [14-16]). Normal mode analysis (NMA) [17] is a particularly interesting tool in this regard, since a small number of low-frequency normal modes provide a good hint of the direction of large-amplitude conformational changes [18-21]. Several recent works apply this property of NMA to improve the performance of conformational exploration methods.

The approach presented in this paper was originally introduced in [22]. The basic principle is to use NMA to bias the conformational exploration performed by a Rapidly-exploring Random Tree (RRT) algorithm [23], aiming to efficiently compute conformational transition paths. The main novelty presented in the present work is the introduction of a multi-scale model for the protein. In this model, an elastic network is defined considering only a single node (called a particle) per tripeptide. Motion directions provided by NMA of such a coarse-grained elastic network are then applied to the all-atom model for a more accurate conformational exploration. The introduction of this multi-scale model has important outcomes. First, the number of normal modes is largely reduced thanks to the use of the coarse-grained model, which significantly reduces the time required to compute them. In addition, generating the all-atom model from the coarse-grained model can be accurately and efficiently achieved using methods from robot kinematics [24], avoiding the need of artifacts such as the RTB approach (rotations-translations of blocks) [25].

Next section presents the overall method, and explains each of its elementary components: elastic network normal mode analysis, tripeptide-based multi-scale protein modeling, and motion-planning-based conformational exploration. Then, several types of results aimed to validate the

approach and to show its good computational performance are presented for a set of proteins with different sizes and topologies. A more detailed analysis of results is presented for adenylate kinase (ADK). Finally, together with the conclusions, we discuss possible directions for future work. Note that a preliminary version of this work was presented in [26]. Compared to this previous version, this paper includes more detailed explanations of the method, a more exhaustive presentation of results, with additional figures and tables, as well as additional results for the ADK protein. In addition, some movies that illustrate results obtained with the proposed method are included as supplementary material.

## Methods

This section presents a new method to model protein conformational transitions. It builds on the combination of several components inside an iterative algorithm. One of these components is NMA performed on a coarse-grained elastic network model of the protein, which enables very fast computation of normal modes. Indeed, a single particle of the elastic network is considered for each group of three consecutive amino-acid residues (i.e. one particle per tripeptide). The all-atom model, which is used to accept or reject sampled states during the conformational exploration, is accurately reconstructed from the coarse-grained one using closed-form inverse kinematics. The RRT algorithm is applied to explore linear combinations of normal modes computed from intermediate conformations along the path. All these elementary components of the method are further explained below.

### Elastic networks and normal mode analysis

Based on a harmonic approximation of the potential energy, normal mode analysis provides information about the directions and frequencies of vibration of a molecule from a minimum-energy conformation. Each mode represents a motion pattern, in which all the atoms move with the same frequency and phase. Low-frequency normal modes correspond to collective motions (e.g. domain motions), whereas high-frequency normal modes correspond to local fluctuations [19,27].

Normal modes are calculated by diagonalizing the Hessian matrix of the potential energy of the molecule. For reducing the computational cost of this operation, several works propose to use simplified potentials and coarse-grained models. An extensively used simplified potential is based on the elastic network model (ENM) [28], which represents the molecule as a set of particles connected by virtual springs. All the protein atoms can be considered as particles in the elastic network. However, a coarse-grained representation that only considers  $C_{\alpha}$  atoms (i.e. a single particle per amino-acid residue) is often applied [19,20]. Moreover, particles are connected by

virtual springs only if they are closer than a user-defined cut-off distance  $d_{cut}$ .

The potential energy function of such an elastic network takes the following form:

$$E = \sum_{d_{ij} < d_{cut}} \frac{C}{2} (d_{ij} - d_{ij}^0)^2$$

where  $d_{ij}$  is the distance between particle  $i$  and particle  $j$ ,  $d_{ij}^0$  is the distance between the two particles at the equilibrium state and  $C$  is the elastic constant. This type of simplified potential has been used in many works and for very different applications [29-32].

In this work, we investigate a further simplification of the ENM. Indeed, the ENM is built using a coarser model based on tripeptides, instead of using  $C_\alpha$  atoms. Figure 1 illustrates the approach. Note that coarse-grained NMA approaches considering more than one residue per particle have already been proposed [25,33,34]. However, these approaches, which are mainly devised to analyze motions of very large systems made of protein assemblies, consider rigid-body motions of groups of residues. In contrast, the approach presented here preserves full flexibility of the protein, which leads to a more accurate simulation of conformational transitions.

Several works show that using a simplified ENM does not necessarily imply a loss of accuracy in the prediction of large-amplitude motion directions [20,25]. However, it certainly leads to a computational performance gain. This issue is further discussed in the results section, where the performance of NMA using tripeptide-based models and  $C_\alpha$ -based models is compared.

The anisotropic network model (ANM) approach, as described in [27,35], is adopted in this work to construct the Hessian matrix from the positions of the particles of the tripeptide-based model. Each  $3 \times 3$  sub-matrix

corresponding to the interaction between two particles is computed as follows:

$$H_{ij} = -\frac{C}{d_{ij}^2} \begin{bmatrix} (x_j - x_i)(x_j - x_i) & (x_j - x_i)(y_j - y_i) & (x_j - x_i)(z_j - z_i) \\ (y_j - y_i)(x_j - x_i) & (y_j - y_i)(y_j - y_i) & (y_j - y_i)(z_j - z_i) \\ (z_j - z_i)(x_j - x_i) & (z_j - z_i)(y_j - y_i) & (z_j - z_i)(z_j - z_i) \end{bmatrix}$$

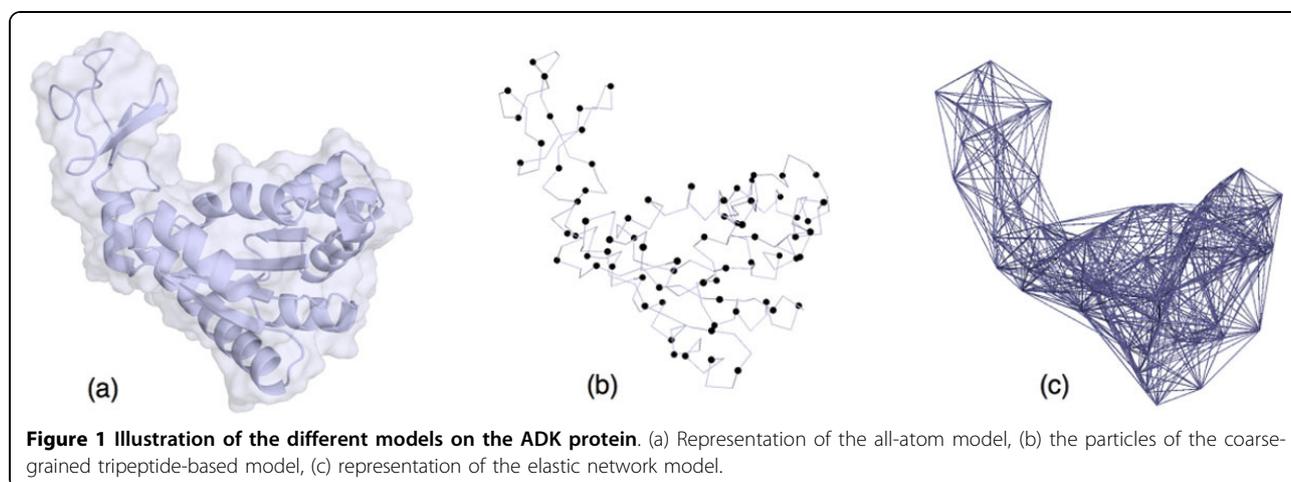
$$H_{ij} = - \sum_{j|j \neq i} H_{ij}$$

If the distance between particles  $i$  and  $j$  is more than the cut-off distance  $d_{cut}$ , then the whole  $3 \times 3$  matrix is replaced by zeros. The Hessian matrix is then diagonalized to compute the eigenvalues and eigenvectors. Each eigenvalue and eigenvector pair corresponds to one normal mode, where the eigenvalue defines the mode frequency and the eigenvector defines the motion direction for each particle in the elastic network.

### Multi-scale model

#### Tripeptide-based model

The multi-scale modeling approach applied in this work is based on a decomposition of the protein chain into fragments of three amino acid residues, which we refer to as *tripeptides*. The reason for choosing such a subdivision is that, assuming fixed bond lengths, bond angles and peptide bond torsions, the backbone of a tripeptide involves 6 degrees of freedom (three pairs of angles  $\phi$ ,  $\psi$ ), and thus, an analogy can be made with a 6R mechanism like a robotic manipulator [24]. Two Cartesian reference frames attached to the N atom in the backbone of the first residue and to the C atom in the last residue define respectively the *base-frame* and the *end-frame* of the 6R mechanism. Since tripeptides are linked through rigid peptide bonds, the location of the end-frame of tripeptide  $i$  can be determined from the base-frame of tripeptide  $i + 1$  by a constant 3D transformation. Given the location of the



base-frame and the end-frame, the conformation of a tripeptide backbone can be obtained by *inverse kinematics*. Consequently, the conformation of the whole protein backbone can be determined from the pose of a single reference frame attached to each tripeptide (this is true for all the protein backbone except two short fragments at the N-terminal and C-terminal ends of the chain, which require a particular treatment). In the following, we will refer to these reference frames as (oriented) *particles*. They are the particles in the coarse-grained ENM.

#### Reconstructing the all-atom model

The interest of the decomposition of the protein into tripeptides explained above is that closed-form inverse kinematics (IK) can be applied to reconstruct the all-atom protein model from the coordinates of the particles. The IK solver applied in this work has been adapted from the method developed by Renaud [36]. This solver is based on algebraic elimination theory, and develops an ad-hoc resultant formulation inspired by the work of Lie and Liang [37]. Starting from a system of equations representing the IK problem, the elimination procedure leads to an 8-by-8 quadratic polynomial matrix in one variable. The problem can then be treated as a generalized eigenvalue problem, as proposed in [38], for which efficient and robust methods such as the Schur factorization can be applied. Note however that our approach is not dependent on this solver, so that other IK methods (e.g. [38,39]) could be applied.

In general, the IK problem for a 6R serial kinematic chain has a finite number of solutions (up to 16 in the most general case). All the solutions correspond to geometrically valid conformations of the tripeptide backbone with fixed ends defined by the pose of the particles. However, when the goal is to simulate continuous motions, the closest conformation to the previous one (i.e. the one before a perturbation applied to the particles) has to be selected in order to avoid jumps in the conformational space. All IK solutions are rejected if none of them remains within a distance threshold that depends on the perturbation step-size.

The explanations above concern only the reconstruction of the all-atom model of the protein backbone from the coarse-grained tripeptide-based model. Side-chains are treated separately, using a simple method based on energy minimization as explained below.

#### Path finding algorithm

The path finding method works by iteratively generating short portions of the transition between two given conformations of a protein, which we will refer to as  $q_{init}$  and  $q_{goal}$ . Algorithm 1 presents the pseudo-code with the main steps of the method. At each iteration, normal

modes are computed for a root conformation  $q_{root}$ . Note that  $q_{root} = q_{init}$  for the first iteration. Then, the RRT algorithm is applied to explore motions corresponding to linear combinations of normal modes. RRT is run until the protein moves a predefined distance toward the target conformation  $q_{goal}$ . The conformational exploration performed by the RRT algorithm is further explained below. Once the RRT exploration is stopped, the closest node  $q_{close}$  in the tree to  $q_{goal}$  is searched. The path between  $q_{root}$  and  $q_{close}$  is then extracted and saved. All the conformations in this path are guaranteed to have a collision-free backbone (including  $C_{\beta}$  atoms) which generally implies getting acceptable energy values after a short minimization to rearrange side-chain conformations. Such an energy minimization procedure is performed on  $q_{close}$ , which will be the root conformation in the next iteration. The algorithm keeps iterating until a predefined distance  $d_{target}$  to  $q_{goal}$  is reached. The resulting path is defined by the sequence of minimized conformations  $q_{close}$  at each iteration. If a finer-grained path is required, other intermediate conformation can be extracted from the sub-paths computed at each iteration. These conformations may require energy minimization to rearrange side-chains, as it is done for  $q_{close}$ .

#### Algorithm 1: COMPUTE\_PATHWAY

**input** : Initial conformation  $q_{init}$ , final conformation  $q_{goal}$ , minimum distance to target  $d_{target}$   
**output** : The transition path  $p$   
**begin**  
 $q_{root} \leftarrow q_{init}$   
**while**  $RMSD(q_{root}, q_{goal}) > d_{target}$  **do**  
 $a \leftarrow COMPUTE\_NORMAL\_MODES(q_{root});$   
 $t \leftarrow BUILD\_RRT(q_{root}, q_{goal}, a);$   
 $q_{close} \leftarrow CLOSEST\_TO\_TARGET(t, q_{goal});$   
 $q_{root} \leftarrow MINIMIZE(q_{close});$   
 $p \leftarrow CONCATENATE(p, q_{root});$   
**end**

#### Algorithm 2: BUILD\_RRT

**input** : Initial conformation  $q_{root}$ , final conformation  $q_{goal}$ , normal modes  $a$   
**output** : The tree  $t$   
**begin**  
 $t \leftarrow INIT\_TREE(q_{root});$   
**while not**  $STOP\_CONDITION(t, q_{goal})$  **do**  
 $q_{rand} \leftarrow SAMPLE(t, a);$   
 $q_{near} \leftarrow BEST\_NEIGHBOR(t, q_{rand});$   
 $q_{new} \leftarrow EXPAND\_TREE(q_{near}, q_{rand});$   
**if**  $ISVALID(q_{new})$  **then**  
 $ADD\_NEW\_NODE(t, q_{new});$   
 $ADD\_NEW\_EDGE(t, q_{near}, q_{new});$   
**end**

### Implementation details

The RRT algorithm, iteratively applied in Algorithm 1, performs the same steps as the basic RRT [23]. The steps are sketched in Algorithm 2. At each iteration, a conformation  $q_{rand}$  is randomly sampled. Note that  $q_{rand}$  is not required to be a feasible conformation. Then, the tree is searched for the closest conformation to  $q_{rand}$ , called  $q_{near}$ . A new conformation,  $q_{new}$ , is generated by moving from  $q_{near}$  towards  $q_{rand}$  with a predefined short step size. The new conformation is added to the tree if it does not violate feasibility constraints, which in the present work are limited to geometric constraints related to no atom overlapping and no bond breaking. The difference with respect to the basic RRT algorithm concerns the implementation of the methods for sampling conformations, searching the nearest neighbor, and expanding the tree. These methods, which are further explained below, are specific to the present framework because of the multi-scale protein model and the application of NMA to bias the exploration.

#### Sampling random conformations

The idea is to randomly sample conformations  $q_{rand}$  using information given by the normal modes. The coarse-grained tripeptide-based model is used at this level. Hence,  $q_{rand}$  is not an all-atom conformation, but an array of particle positions. Random particle positions are generated by moving them from their initial positions, defined by  $q_{root}$  using a linear combination of normal modes with randomly sampled weights. More precisely:

- A sequence of  $3n$  random weights  $w_j$  are sampled in the range  $[-1, 1]$ , where  $n$  is the number of particles, being  $3n$  the number of normal modes (actually, the number of normal modes is  $3n - 6$ , since 6 degrees of freedom correspond to rigid-body motions of the whole set of particles).
- The new positions of the  $n$  particles are computed by a linear combination of all the randomly weighted modes as follows:

$$q_{rand} = q_{root} + \sum_{j=1}^{3n} f * w_j * a_j$$

where  $a_j$  refers to each normal mode, and  $f$  is an amplification factor used to push the sampled conformation away from  $q_{root}$  (this factor is the same for all the normal modes). Note that, since the normal modes are not normalized, low frequency modes have larger norm. Thus, they contribute more significantly in the sum.

#### Finding nearest neighbors

Nearest neighbor search is also performed using the coarse-grained model. Indeed, the computed distance is based on the root mean squared deviation (RMSD) of the particle positions. In the current implementation, the

distance is biased to pull the exploration towards the target conformation as follows:

$$d(q, q_{rand}) = \text{RMSD}(q, q_{rand}) \frac{\text{RMSD}(q, q_{goal})}{\text{RMSD}(q_{init}, q_{goal})}$$

In this work, we have implemented a simple brute-force algorithm to find  $q_{near}$ . More sophisticated nearest neighbor search algorithms could be used to reduce the number of performed distance computations. Note, however, that currently used algorithms based on space partitioning techniques (e.g. kd-trees) do not perform well in high-dimensional spaces [40]. A computationally efficient solution would require the implementation of an approximate nearest neighbor search algorithm.

#### Generating new conformations

For generating  $q_{new}$ , all particle positions in  $q_{near}$  are linearly interpolated towards  $q_{rand}$  with a predefined step size  $k$ . Given these new particle positions, the all-atom model corresponding to  $q_{new}$  is obtained by solving an IK problem for every tripeptide. The implemented method proceeds iteratively. If no IK solution is found for a tripeptide  $t_i$  (the tripeptide between particles  $p_i$  and  $p_{i+1}$ ) or if the solution involves atom collisions, the pose (position and orientation) of particle  $p_{i+1}$  is slightly perturbed and the IK problem is solved again. This process is repeated until a collision-free IK solution is found or a maximum number of trials is reached. If this process fails to find a collision-free IK solution for any tripeptide, failure is reported and the RRT algorithm goes back to the random sampling step.

Once the treatment of all tripeptides has been completed, the conformation of the two terminal fragments is generated. For this, the pose of these fragments is updated with respect to the new poses of the first and last tripeptides. Random perturbations can be applied to these end fragments in order to remove possible collisions with the rest of the protein.

Protein conformations  $q_{new}$  generated using the aforementioned process are guaranteed to satisfy geometric constraints: correct bond geometry and no overlap between backbone atoms. However, in order to speed-up computations, side-chains are not treated at this stage (only  $C_\beta$  atoms are considered for collision avoidance). This is because side-chains are known to be very flexible, and resolving possible collisions along the conformational transition path can be done in a post-processing stage. Indeed, side-chain collisions are resolved during the minimization step at the end of each short RRT execution.

## Results and discussion

This section discusses several experiments aimed to validate the proposed method and to evaluate its performance.

First, the question concerning the accuracy of the tripeptide-based elastic network model is addressed. Then, results are presented on conformational transitions computed for a set of ten proteins with different sizes and topologies. Finally, further results on adenylate kinase are presented and compared to available data on the transition between the open and closed forms of this protein.

### Validating the coarse-grained ENM

Previous works (e.g. [19,20]) have shown that simple ENMs built using  $C_{\alpha}$  atoms perform as well as ENMs built using the all-atom model when studying the dynamic properties of proteins with NMA. Here, we compare the performance of the proposed tripeptide-based model with the  $C_{\alpha}$ -based model for predicting directions of conformational transitions. A set of seven proteins listed in Table 1 was used for this comparison. These proteins were also used in related work [20] for the validation of the  $C_{\alpha}$ -based ENM.

For evaluating the capability of normal modes to predict directions of conformational transitions, we use the notion of *overlap* as proposed in related work [20]. The overlap  $I_j$  between a normal mode  $j$  and an experimentally observed conformational change between two conformations (open and closed)  $q^o$  and  $q^c$  is defined as a measure of similarity between the conformational change and the direction given by the normal mode  $j$ . It can be computed as follows:

$$I_j = \frac{\left| \sum_{i=1}^{3n} a_{ij} \Delta q_i \right|}{\left[ \sum_{i=1}^{3n} a_{ij}^2 \sum_{i=1}^{3n} \Delta q_i^2 \right]^{1/2}}$$

where  $\Delta q_i = q_i^o - q_i^c$  measures the difference between the particle coordinates in conformations  $q^o$  and  $q^c$ ,  $a_{ij}$  corresponds to the  $i^{\text{th}}$  coordinate of the normal mode  $j$ , and  $n$  is the number of particles. A value of 1 for the overlap means that the direction given by the normal mode matches exactly the conformational change, whereas a value around 0.2 or less means that the normal mode is unable to provide any meaningful prediction.

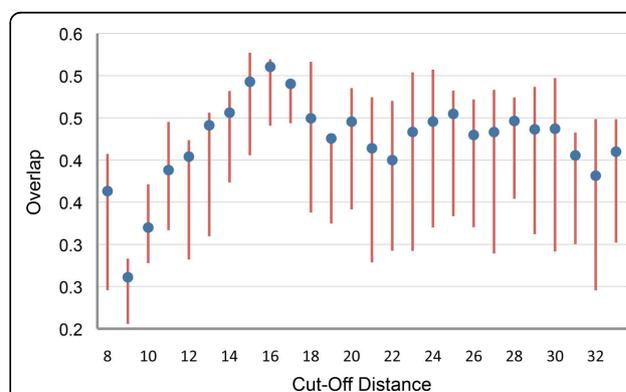
**Table 1 Proteins used in the overlap experiments**

Protein	Residues	PDB <sub>open</sub>	PDB <sub>closed</sub>
Che Y Protein	128	3chy	1chn
LAO binding Protein	238	2lao	1laf
Triglyceride Lipase	256	3tgl	4tgl
Thymidylate Synthase	264	3tms	2tsc
Maltodextrin Binding Protein	370	1omp	1anf
Enolase	436	3enl	7enl
Diphtheria Toxin	523	1ddt	1mdt

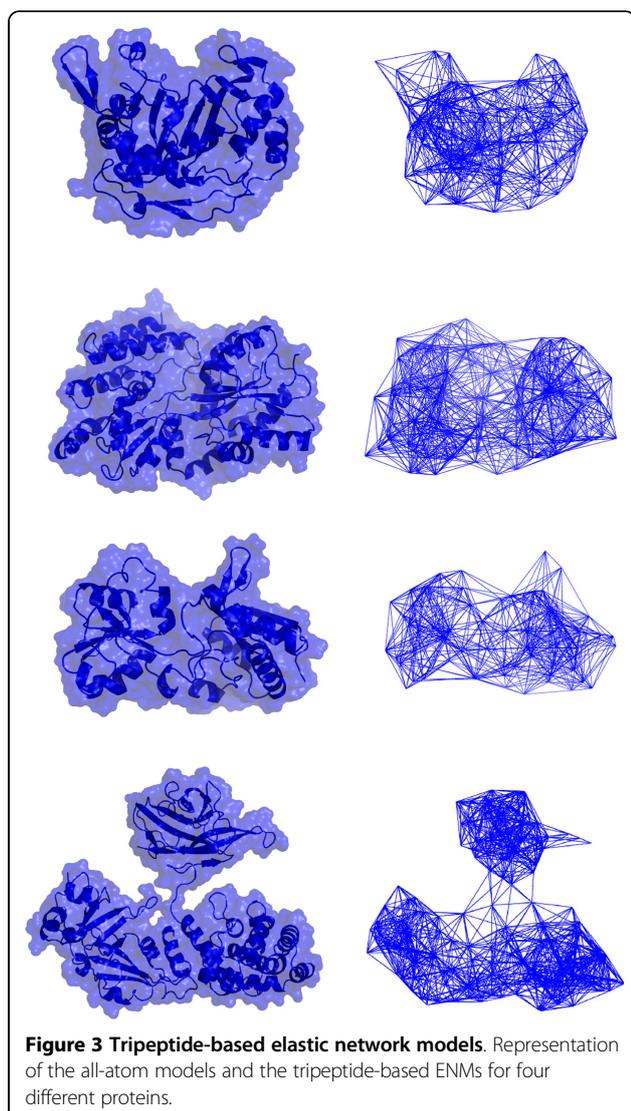
Before conducting the comparative analysis, we need to determine an optimal cutoff distance for the tripeptide-based ENM. A good cutoff distance should create an elastic network that correctly captures the topology of the protein. For  $C_{\alpha}$ -based models, 8 Å is generally used, since this cutoff distance has been empirically shown to provide the best results in most cases. It can be intuitively inferred that the same cutoff distance may not be the optimal choice in our case, because distances between particles of the tripeptide-based model are larger than distances between  $C_{\alpha}$  atoms. Moreover, defining the optimal cutoff value theoretically is not straightforward. Therefore, we have measured and compared the overlap values for the seven proteins with cutoff distances between 8 and 34 Å in order to empirically determine the most suitable range of cutoff values. Figure 2 shows the overlap value for each cutoff distance averaged over the seven proteins. Note that, for each protein, overlap values were computed for all the normal modes, and the best value was considered for the average. As clearly shown in the figure, the best overlap values are for cutoff distances of 15, 16 and 17 Å.

The tripeptide-based ENMs for four of the proteins in Table 1, using a cutoff distance of 16 Å, are represented in Figure 3. The figure shows that the main topological features of the proteins appear in the coarse-grained model.

Table 2 compares overlap values of tripeptide-based ENMs using a cutoff distance of 16 Å with those presented in [20] for  $C_{\alpha}$ -based ENM using a cutoff distance of 8 Å. In the table, columns labeled "Open" correspond to the open-to-closed conformation and columns labeled "Closed" are for the opposite case. The similar overlap values show that the coarse-grained, tripeptide-based ENM is also able to capture the topological information required to compute normal modes that correctly predict directions of large-amplitude motions. Importantly, such a similar performance in terms of overlap is obtained



**Figure 2 Average overlap over the seven proteins of Table 1.** Lines are drawn between the 25th and the 75th percentiles of the overlap values. Average overlap values are indicated with dots.



with much less computational cost. Since the computational complexity of the Hessian matrix diagonalization is  $\mathcal{O}(n^3)$ , the reduction of  $n$  by a factor 3 (a tripeptide involves 3  $C_\alpha$  atoms) provides a theoretical gain of more

**Table 2 Comparison between overlap values for  $C_\alpha$ -based ENMs and tripeptide-based ENMs**

Protein	$C_\alpha$ Overlap		Tripep. Overlap	
	Open	Close	Open	Close
Che Y Protein	0.32	0.34	0.52	0.34
LAO binding Protein	0.84	0.40	0.53	0.52
Triglyceride Lipase	0.30	0.17	0.26	0.35
Thymidylate Synthase	0.56	0.40	0.49	0.29
Maltodextrin Binding Protein	0.86	0.77	0.90	0.84
Enolase	0.33	0.30	0.40	0.30
Diphtheria Toxin	0.58	0.37	0.48	0.30

than one order of magnitude. This theoretical gain has been confirmed with some experiments. In summary, the time required to compute the normal modes with the tripeptide-based model ranges from 0.05 seconds to 0.9 seconds, whereas several minutes may be necessary using the  $C_\alpha$  model.

### Finding conformational transitions

#### Experimental setup

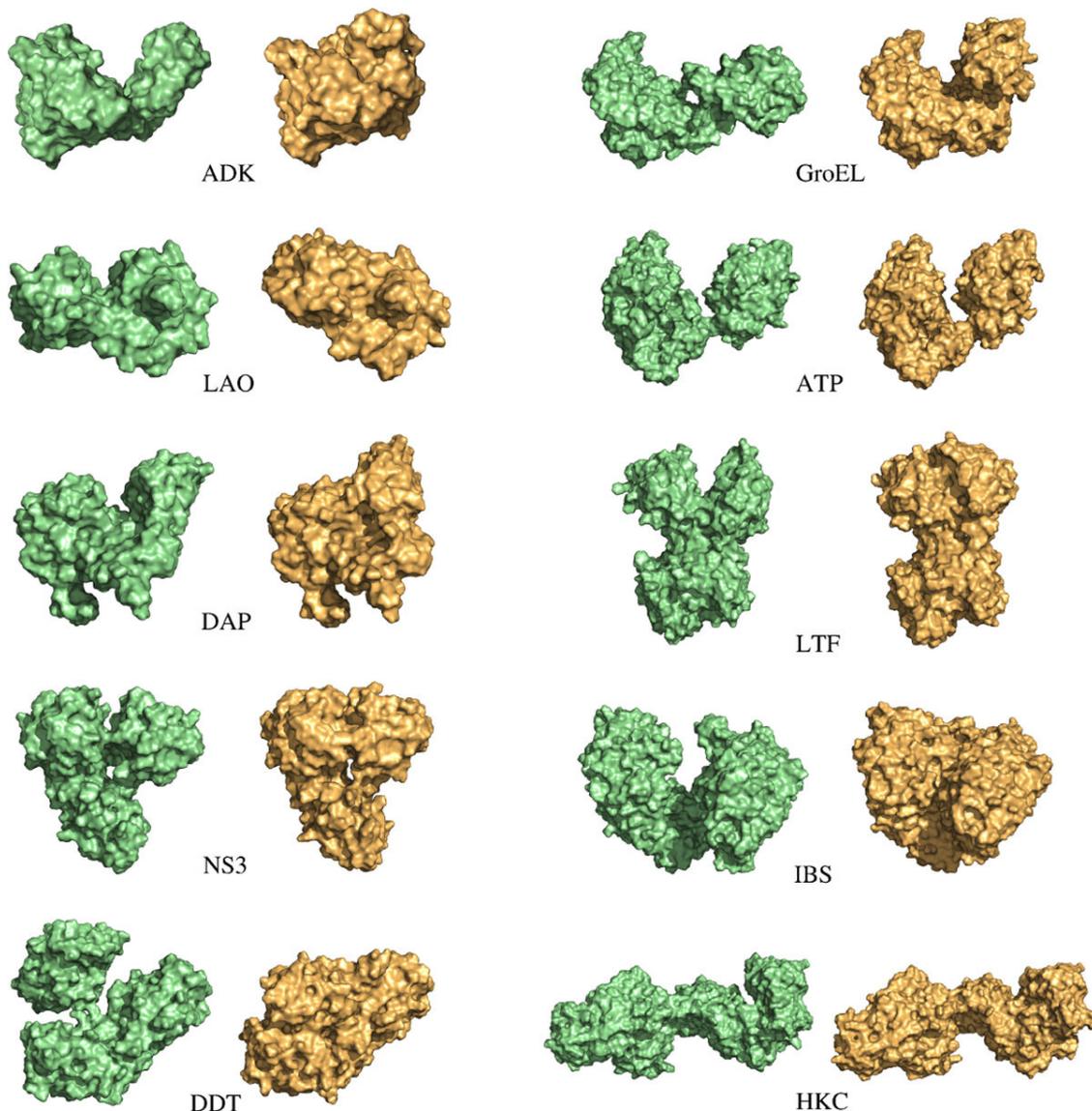
The proposed method was applied to compute conformational transition paths for the ten proteins listed in Table 3, and represented in Figure 4. For each protein, at least two experimental structures corresponding to different conformations are available in the Protein Data Bank (PDB) [41]. The difference between these conformations involves large-amplitude domain motions. The ten proteins are varied in size and topology, as well as in the type of domain motions they undergo. This heterogeneity is important to analyze the reliability and scalability of the method.

Each iteration of the algorithm that computes the transition path performs a short RRT exploration, as mentioned in the previous section. In the current implementation, such a local exploration runs until the protein moves 0.3 Å  $C_\alpha$ -RMSD towards the goal. This distance is gradually reduced to 0.15 Å as the distance to the target conformation decreases. The reason is that the speed of convergence tends to decrease when approaching the target conformation, and recomputing normal modes more frequently provides better results in this situation. If the distance stopping condition is not reached first, the exploration stops after a pre-defined number of iterations (4000 in our case). This additional stopping condition prevents too long runs of RRT in case of blocking situations.

At the end of the RRT exploration, the closest conformation to the goal is identified and submitted to an energy minimization procedure aimed at generating better side-chain conformations. In this work, we have used the AMBER software package [42] for energy minimization.

**Table 3 Proteins used in the experiments**

Protein	Residues	PDB ID <sub>init</sub>	PDB ID <sub>goal</sub>	$C_\alpha$ RMSD
ADK	214	4ake	1ake	6.51
LAO	238	2lao	1laf	3.73
DAP	320	1dap	3dap	3.78
NS3	436	3kqk	3kql	2.75
DDT	535	1ddt	1mdt	10.96
GroEL	547	1aon	1oel	10.49
ATP	573	1m8p	1i2d	3.78
LTF	691	1cb6	1bka	4.75
IBS	876	1ukl	1qgk	6.17
HKC	917	1hkc	1hkb	3.00



**Figure 4** The ten proteins used in the experiment. Representation of open and closed forms of these proteins available in the PDB (IDs are provided in Table 3).

## Results

Table 4 summarizes the results achieved by the proposed method for the set of ten proteins. In this table,  $C\alpha$ -RMSD<sub>end</sub> is the distance between the goal conformation and the conformation obtained at the end of the iterative path finding process. Time<sub>total</sub> is overall computing time, which includes the RRT running time (Time<sub>RRT</sub>) and the time for computing the normal modes and running minimizations at the end of each iteration. The number of iterations of the main algorithm (i.e. the number of NMA calculations) is also indicated in the table. Note that, in all the experiments, the RRT exploration takes more than 90% of the total computing time, which

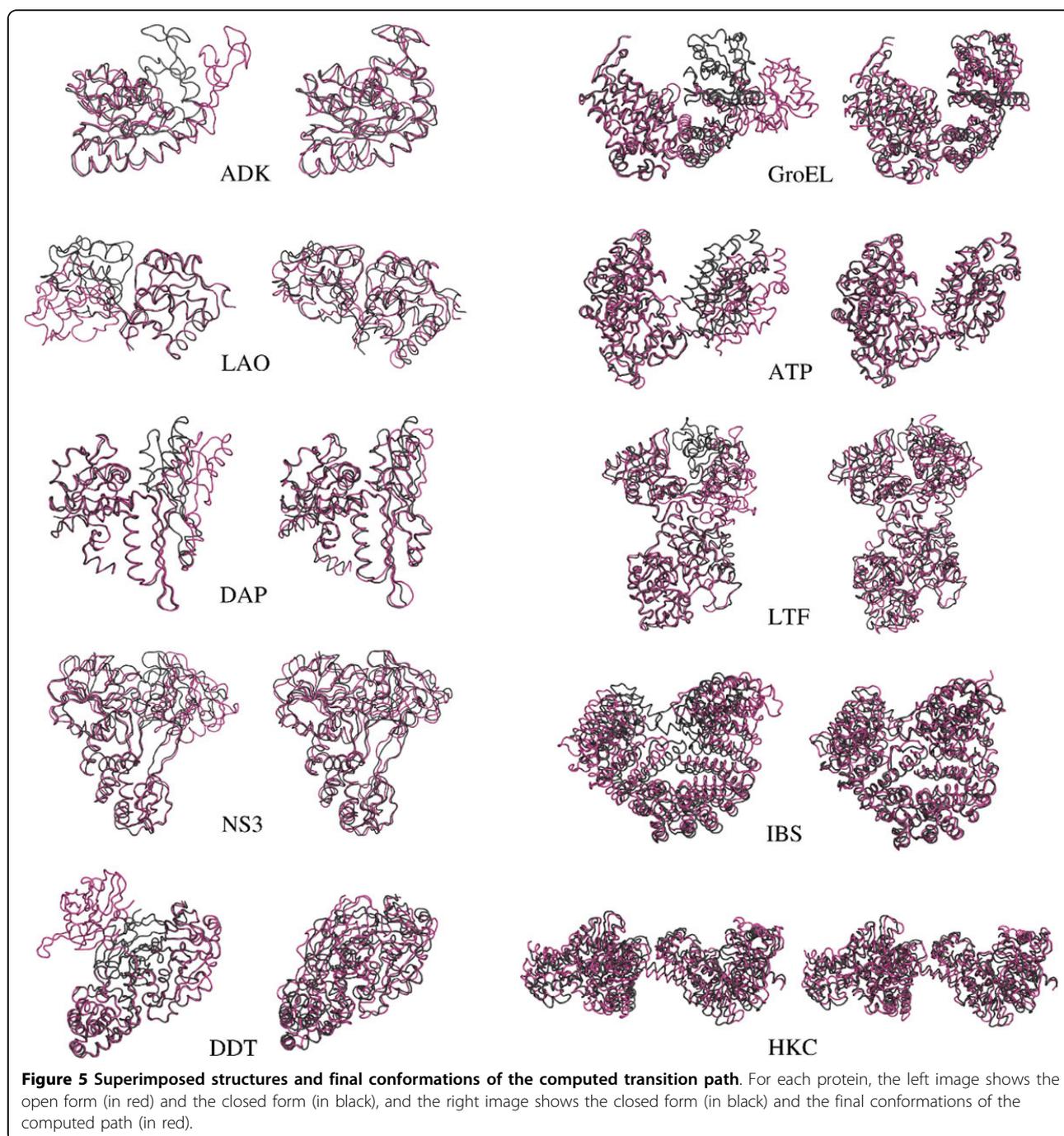
**Table 4** Performance of the method on ten proteins (cf. Table 3)

Protein	$C\alpha$ -RMSD <sub>end</sub>	Iterations	Time <sub>RRT</sub>	Time <sub>total</sub>
ADK	1.56	31	1.82	2.00
LAO	1.32	20	1.52	1.65
DAP	1.31	16	1.78	1.92
NS3	1.29	14	2.82	3.00
DDT	2.88	272	81.54	86.4
GroEL	2.79	142	40.21	42.17
ATP	1.45	30	13.46	14.16
LTF	1.96	74	29.56	31.09
IBS	1.99	80	80.61	82.62
HKC	1.64	38	37.91	39.63

corresponds to runs on a single core of an AMD Opteron 148 processor at 2.6 GHz.

In all cases, the method was able to compute the conformational transition, reaching conformations very close to the given goal conformations. Figure 5 shows superimposed structures (structure superimpositions and images have been done using PyMOL [43]) of open and closed forms of the proteins ( $q_{init}$  and  $q_{goal}$ ), and of the closed form and the last conformation of the computed

transition path ( $q_{goal}$  and  $q_{final}$ ). The distances between the final and goal conformations are below 2 Å (measured using  $C_{\alpha}$ -RMSD) for all the tested proteins with the exception of DDT and GroEL. Note that 2 Å RMSD corresponds to the current accuracy of experimental methods for high-resolution protein structure determination. As can be seen in Figure 5 the superimpositions of the final and goal conformations is very good, even for DDT and GroEL. Note that the method could have



reached closer conformations to the goal with a higher number of iterations. Nevertheless, the strategy applied in these experiments was to stop iterating when the distance to the goal reached a very slow rate of convergence.

We also conducted experiments to analyze the relationship between the computing time and the size of the protein. Since the lengths of the transition paths for the different test systems is variable, we measured the computing time to move 1Å along these paths. The results of these experiments, presented in Table 5 and Figure 6, show a linear relationship between the computing time and the protein size. This scalability is an interesting property of the method. Note that the performance of the method seems not be (or only slightly) affected by the topology of the protein. This is an important advantage over the method presented in [22], which experienced some difficulties in dealing with relative motions of domains connected by several linkers, mainly because of the internal-coordinate representation of proteins used in this previous work.

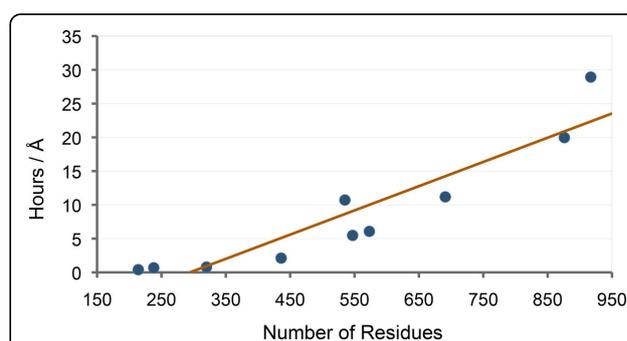
Finally, we did a profiling of the algorithm to identify possible bottlenecks and points to be improved to enhance computational efficiency. Table 6 gives values of the percentage of the time spent in the most time-consuming operations within the RRT exploration: nearest neighbor search (NN), collision checking (CC), inverse kinematics (IK) and random sampling (RS). Surprisingly, nearest neighbor search takes around 60% of the overall computing time. This is due to the brute-force algorithm applied in the current implementation. As mentioned before, a more sophisticated nearest neighbor algorithm should be implemented. The performance of the method could also be enhanced by applying simplified distance metrics (e.g. [16,44]). The use of an appropriate simplified distance metric could reduce computing time while preserving good exploration properties of the algorithm.

#### A closer look at adenylate kinase

Adenylate kinase (ADK) [45] is a widely studied protein involved in signal transduction. The structure of ADK is composed of three domains known as: LID, CORE and

**Table 5 Relationship between the size of the protein and the computing time**

Protein	Residues	Time (hours)
ADK	214	0.4
LAO	238	0.68
DAP	320	0.79
NS3	436	2.11
DDT	535	10.72
GroEL	547	5.84
ATP	573	6.74
LTF	691	11.17
IBS	876	19.96
HKC	917	28.93



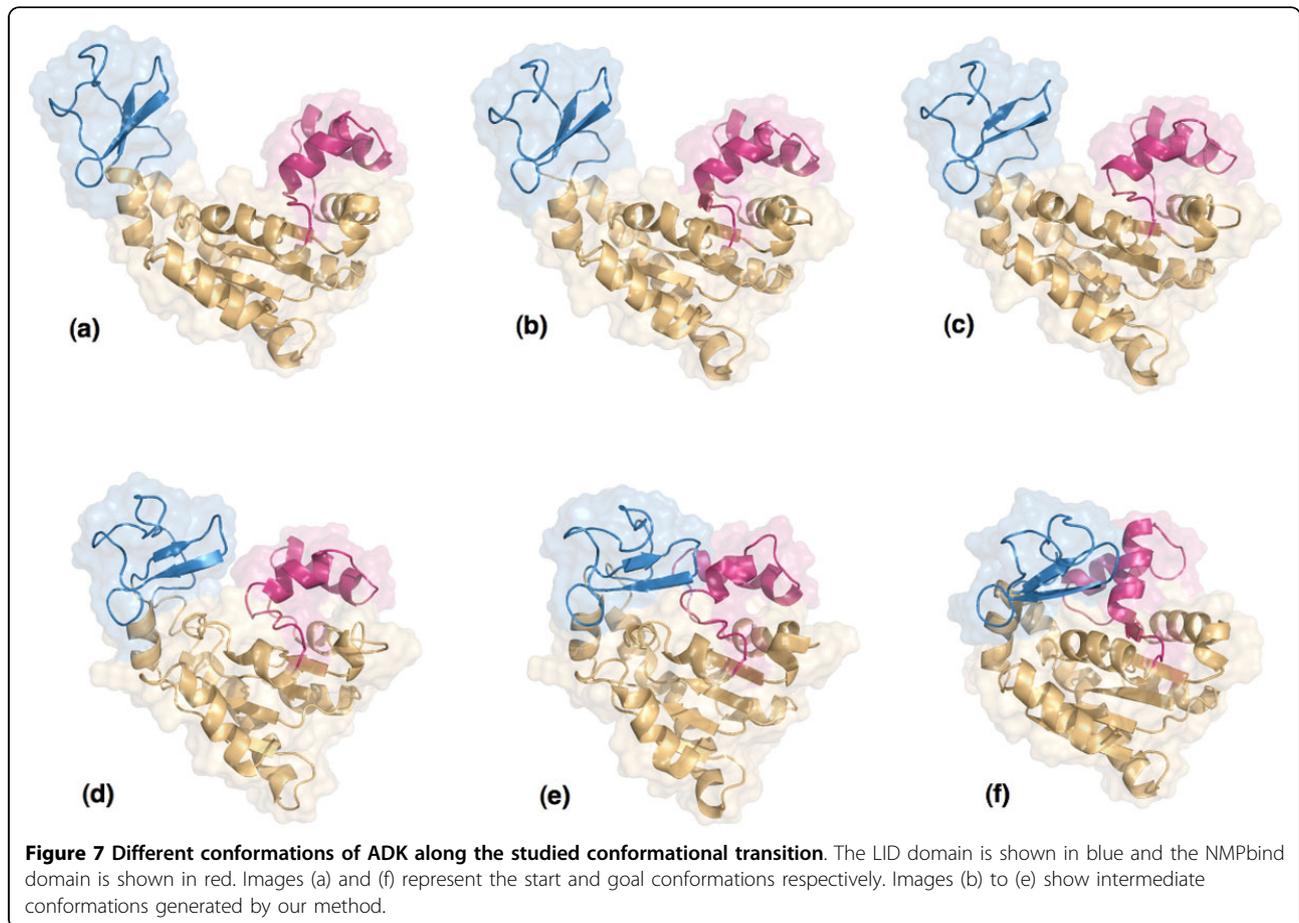
**Figure 6 Plot of the results in Table 5.** The plot shows a linear relationship between the size of the protein and the time required to compute the conformational transition path.

NMPbind. Several works tend to show that the LID and NMPbind domains undergo large-amplitude conformational changes with respect to the CORE domain, which remains stable [46,47]. Some of these works (e.g. [47]) also suggest that the conformational transition between open and closed states of ADK proceeds in two steps: (1) the LID domain moves more clearly than the NMPbind domain at the beginning of the open-to-close transition; (2) then NMPbind domain moves at a faster pace towards the end of the transition path.

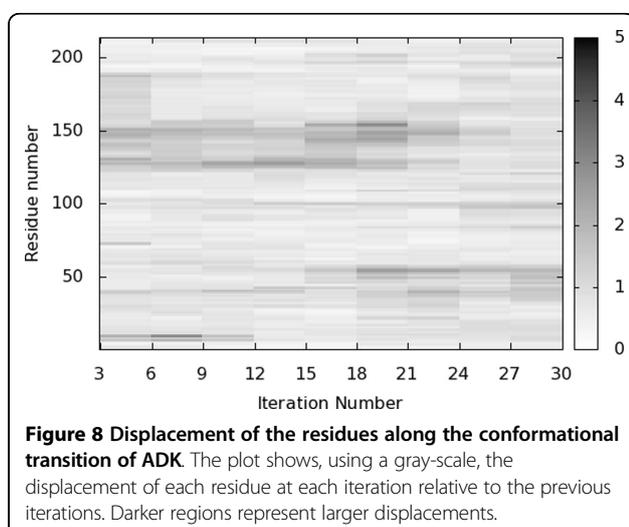
The open conformation of ADK (PDB ID 4AKE), the closed conformations (PDB ID 1AKE) of ADK, and several intermediate conformations obtained with our method are represented in Figure 7. The figure shows significant conformational changes of the LID and NMPbind domains, as expected. The motion of these two regions is also illustrated in Figure 8, which represents the displacement of the residues along the conformational transition. Two darker regions, involving residues 20-60 and 130-160, indicate the parts of the protein that undergo larger displacements. These regions correspond to the NMPbind domain and LID domain, approximately. Figure 8 also shows that residues 20-60, corresponding to the NMPbind

**Table 6 Percentage of the time spent performing the main operations in RRT**

Protein	NN	CC	IK	RS
ADK	57.2%	14.1%	15.0%	6.3%
LAO	51.3%	20.9%	17.0%	5.4%
DAP	50.5%	20.6%	11.0%	12.3%
NS3	67.9%	13.4%	6.6%	8.9%
DDT	64.3%	17.1%	6.9%	9.0%
GroEL	60.4%	17.6%	8.9%	9.8%
ATP	57.3%	20.9%	6.8%	11.9%
LTF	55.1%	16.8%	6.1%	19.3%
IBS	62.9%	15.5%	4.1%	15.5%
HKC	68.9%	5.8%	3.3%	18.2%
Average	59.58%	16.27%	8.57%	11.66%



domain, start moving more significantly near the end of the transition path, whereas residues 130-160, corresponding to the LID domain, start moving at an earlier stage. This reflects the two-step nature of the conformational transition discussed earlier, and shows that our method provides



results that are qualitatively comparable with those presented in previous work on ADK.

The open conformation of ADK (PDB ID 4AKE), the closed conformations (PDB ID 1AKE) of ADK, and several intermediate conformations obtained with our method are represented in Figure 7. The figure shows significant conformational changes of the LID and NMPbind domains, as expected. The motion of these two regions is also illustrated in Figure 8, which represents the displacement of the residues along the conformational transition. Two darker regions, involving residues 20-60 and 130-160, indicate the parts of the protein that undergo larger displacements. These regions correspond to the NMPbind domain and LID domain, approximately. Figure 8 also shows that residues 20-60, corresponding to the NMPbind domain, start moving more significantly near the end of the transition path, whereas residues 130-160, corresponding to the LID domain, start moving at an earlier stage. This reflects the two-step nature of the conformational transition discussed earlier, and shows that our method provides results that are qualitatively comparable with those presented in previous work on ADK.

We have also compared intermediate conformations in the computed transition path of the ADK to a small number of other experimentally solved structures of this protein. These structures correspond to homolog proteins or mutants with very high sequence identity, and some of them are known to be intermediate structures between open and closed forms of the protein. Interestingly, four of these structures are very close to conformations along the transition path. Table 7 shows the distance between each of these structures and the closest conformation in the transition path. The table also shows the position of this conformation in the path. More precisely, the table shows the corresponding iteration number and the percentage of the path length. 2RH5 (A) is very close to the conformation generated by the first iteration, whereas 1E4Y (A) is close to the conformation generated by iteration 27 (near the closed structure). 1DVR (A) is also very close to a conformation toward the beginning of the path (near the open structure), whereas 2RH5 (B) is a slightly less open structure. These results are comparable to those provided by previous studies [12,48], which further validates the proposed method.

## Conclusions

This paper has presented an efficient approach for computing large-amplitude conformational transitions in proteins. It exploits the ability of normal modes to predict directions of collective, large-amplitude motions and the efficiency of the RRT algorithm to explore large spaces. The proposed approach also relies on a multi-scale representation of the protein, based on a decomposition into tripeptides, which significantly contributes to the good performance of the method.

Interestingly, first results presented in the paper show that using an ENM based on the coarse-grained tripeptide-based model instead of a  $C\alpha$ -based model preserves the ability of NMA to predict directions of large-amplitude motions, while significantly reducing computing time.

The proposed method was applied to simulate large-amplitude conformational transitions in proteins of different sizes and topologies. Results show a good performance of the method in all the cases. Computing time scales linearly with the number of residues. It ranges

from a few hours for medium-size proteins to a few days for very large ones. This computational performance could be significantly improved by the implementation of more sophisticated methods to perform the most time-consuming operations within the RRT algorithm, in particular, nearest neighbor search.

A deeper analysis of the conformational transition between open and closed forms of ADK shows that results provided by the proposed method are qualitatively consistent with results obtained with other computational methods and with experimental data. Nevertheless, it is important to note that the resulting paths are a first approximation, which cannot be used directly for an accurate evaluation of energy variations along conformational transitions. This would require a subsequent refinement and analysis using state-of-the-art energy models and molecular modeling methods. It could also be possible to integrate energy evaluations within the RRT exploration with the aim of obtaining better-quality solutions, at the expense of additional computational cost. An interesting extension that could be investigated is to use T-RRT [49,50], instead of RRT, to compute paths that follow more accurately the valleys of the conformational energy landscape.

In this work, we have shown the ability of the proposed method to compute transition paths between two given conformations of a protein. Nevertheless, the approach could also be applied to more challenging problems, such as the prediction of other (meta-)stable states reachable from a given protein conformation, or the discrimination between probable and improbable transitions. This would require some extensions, mainly in the definition of energy/scoring functions to identify interesting intermediate and meta-stable states, as well as high-energy barriers, during the conformational exploration.

## Additional files

### Additional material

- Additional file 1:** Movie of a conformational transition path for AKD
- Additional file 2:** Movie of a conformational transition path for GroEL
- Additional file 3:** Movie of a conformational transition path for LTF
- Additional file 4:** Movie of a conformational transition path for IBS

**Table 7 Known intermediate structures and their distances to the closest conformation in the computed transition path.**

PDB ID	RMSD	Iteration	Path percent
1DVR (A)	1.48	2	9%
2RH5 (A)	1.80	1	4%
2RH5 (B)	1.91	3	15%
1E4Y (A)	2.20	27	94%

### List of abbreviations

MD: molecular dynamics; NMA: normal mode analysis; RRT: rapidly-exploring random tree; RTB: rotations-translations of blocks; ENM: elastic network model; ANM: anisotropic network model; IK: inverse kinematics; RMSD: root mean squared deviation; PDB: Protein Data Bank.

### Competing interests

The authors declare that there are no competing interests.

#### Authors' contributions

TS and JC designed this research and supervised the work. IA implemented the method and carried out experiments. MV participated in the software design and implementation. IA and JC wrote the manuscript. All authors have read and approved the manuscript.

#### Acknowledgements

This work has been partially supported by the French National Research Agency (ANR) under project ProtiCAD (project number ANR-12-MONU-0015-01).

#### Declarations

The publication costs for this article were funded by the French National Research Agency (ANR) under project ProtiCAD (project number ANR-12-MONU-0015-01).

This article has been published as part of *BMC Structural Biology* Volume 13 Supplement 1, 2013: Selected articles from the Computational Structural Bioinformatics Workshop 2012. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcstructbiol/supplements/13/S1>.

#### Authors' details

<sup>1</sup>CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France. <sup>2</sup>Univ de Toulouse, LAAS, F-31400 Toulouse, France.

Published: 8 November 2013

#### References

- Voter AF: A method for accelerating the molecular dynamics simulation of infrequent events. *J Chem Phys* 1997, **106**(11):4665-4677.
- Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, Wrigger W, Schulten K: Steered molecular dynamics. In *Computational Molecular Dynamics: Challenges, Methods, Ideas* Springer-Verlag; 1998, 39-65.
- Sørensen RR, Voter AF: Temperature-accelerated dynamics for simulation of infrequent events. *J Comput Phys* 2000, **112**:9599-9606.
- Laio A, Parrinello M: Escaping free-energy minima. *Proc Natl Acad Sci USA* 2002, **99**(20):12562-12566.
- Hamelberg D, Morgan J, McCammon JA: Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys* 2004, **120**:11919-11929.
- Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wrigger W: Atomic-level characterization of the structural dynamics of proteins. *Science* 2010, **330**(6002):341-346.
- Mills G, Jónsson H: Quantum and thermal effects in H<sub>2</sub> dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. *Phys Rev Lett* 1994, **72**:1124-1127.
- E W, Ren W, Vanden-Eijnden E: String method for the study of rare events. *Phys Rev B* 2002, **66**:052301.
- Bolhuis PG, Chandler D, Dellago C, Geissler PL: Transition path sampling and the calculation of rate constants. *Annu Rev Phys Chem* 2002, **53**:291-318.
- Cortés J, Siméon T, Ruiz de Angulo V, Guieysse D, Remaud-Siméon M, Tran V: A path planning approach for computing large-amplitude motions of flexible molecules. *Bioinformatics* 2005, **21**(suppl 1):i116-i125.
- Raveh B, Enosh A, Schueler-Furman O, Halperin D: Rapid sampling of molecular motions with prior information constraints. *PLoS Comput Biol* 2009, **5**(2):e1000295.
- Haspel N, Moll M, Baker M, Chiu W, Kaviraki LE: Tracing conformational changes in proteins. *BMC Struct Biol* 2010, **10**(Suppl 1):S1.
- Al-Bluwi I, Siméon T, Cortés J: Motion planning algorithms for molecular simulations: a survey. *Comput Sci Rev* 2012, **6**(4):125-143.
- Kim MK, Jernigan RL, Chirikjian GS: Rigid-cluster models of conformational transitions in macro-molecular machines and assemblies. *Biophys J* 2005, **89**:43-55.
- Thomas S, Tang X, Tapia L, Amato NM: Simulating protein motions with rigidity analysis. *J Comput Biol* 2007, **14**(6):839-855.
- Plaku E, Stamati H, Clementi C, Kaviraki L: Fast and reliable analysis of molecular motion using proximity relations and dimensionality reduction. *Proteins* 2007, **67**(4):897-907.
- Cui Q, Bahar I: *Normal mode analysis: theory and applications to biological and chemical systems* Chapman and Hall/CRC mathematical and computational biology series, Chapman & Hall/CRC; 2006.
- Brooks B, Karplus M: Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci USA* 1985, **82**(15):4995-4999.
- Hinsen K: Analysis of domain motions by approximate normal mode calculations. *Proteins* 1998, **33**(3):417-429.
- Tama F, Sanejouand YH: Conformational change of proteins arising from normal mode calculations. *Protein Eng* 2001, **14**:1-6.
- Alexandrov V, Lehnert U, Echols N, Millburn D, Engelman D, Gerstein M: Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. *Protein Sci* 2005, **14**(3):633-643.
- Kirillova S, Cortés J, Stefaniu A, Siméon T: An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins. *Proteins* 2008, **70**:131-143.
- LaValle SM, Kuffner JJ: Rapidly-exploring random trees: progress and prospects. In *Algorithmic and Computational Robotics: New Directions*. A.K Peters; Donald B, Lynch K, Rus D, Boston 2001:293-308.
- Siciliano B, Khatib O: *Springer Handbook of Robotics* Springer; 2008.
- Tama F, Gadea FX, Marques O, Sanejouand YH: Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* 2000, **41**:1-7.
- Al-Bluwi I, Vaissset M, Siméon T, Cortés J: Coarse-grained elastic networks, normal mode analysis and robotics-inspired methods for modeling protein conformational transitions. *Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on: 4-7 October 2012* 2012, 40-47.
- Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I: Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 2001, **80**:505-515.
- Tirion MM: Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 1996, **77**(9):1905-1908.
- Kim MK, Jernigan RL, Chirikjian GS: Efficient generation of feasible pathways for protein conformational transitions. *Biophys J* 2002, **83**(3):1620-1630.
- Tama F, Miyashita O, Brooks CL III: Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol* 2004, **147**(3):315-326.
- Cavasotto CN, Kovacs JA, Abagyan RA: Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc* 2005, **127**(26):9632-9640.
- Mouawad L, Perahia D: Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures. *J Mol Biol* 1996, **258**(2):393-410.
- Schuyler AD, Chirikjian GS: Efficient determination of low-frequency normal modes of large protein structures by cluster-NMA. *J Mol Graph Model* 2005, **24**:46-58.
- Demerdash ONA, Mitchell JC: Density-cluster NMA: a new protein decomposition technique for coarse-grained normal mode analysis. *Proteins* 2012, **80**(7):1766-1779.
- Eyal E, Yang LW, Bahar I: Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* 2006, **22**(21):2619-2627.
- Renaud M: A simplified inverse kinematic model calculation method for all 6R type manipulators. In *Current Advances in Mechanical Design and Production VII*. New York: Pergamon; Hassan MF, Megahed SM 2000:57-66.
- Lee HY, Liang CG: A new vector theory for the analysis of spatial mechanisms. *Mech Mach Theory* 1988, **23**(3):209-217.
- Manocha D, Canny JF: Efficient inverse kinematics for general 6R manipulators. *IEEE Trans Robot Autom* 1994, **10**(5):648-657.
- Coutsias EA, Seok C, Jacobson MP, Dill KA: A kinematic view of loop closure. *J Comput Chem* 2004, **25**(4):510-528.
- Plaku E, Kaviraki LE: Quantitative analysis of nearest-neighbors search in high-dimensional sampling-based motion planning. In *Algorithmic Foundations of Robotics VII*. Berlin: Springer-Verlag; Akella S, Amato NM, Huang WH, Mishra B, 2008:3-18.
- Research Collaboratory for Structural Bioinformatics PDB:[<http://www.rcsb.org/pdb/>].
- Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, et al: *AMBER 9* San Francisco: University of California; 2006.

43. The PyMOL Molecular Graphics System, Version 1.5, Schrödinger, LLC.
44. Shehu A, Olson B: **Guiding the search for native-like protein conformations with an ab-initio tree-based exploration.** *Int J Robot Res* 2010, **29**(8):1106-1127.
45. Müller CW, Schulz GE: **Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 Å resolution. A model for a catalytic transition state.** *J Mol Biol* 1992, **224**:159-177.
46. Müller CW, Schlauderer GJ, Reinstein J, Schulz GE: **Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding.** *Structure* 1996, **4**(2):147-156.
47. Maragakis P, Karplus M: **Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase.** *J Mol Biol* 2005, **352**:807-822.
48. Feng Y, Yang L, Kloczkowski A, Jernigan RL: **The energy profiles of atomic conformational transition intermediates of adenylate kinase.** *Proteins* 2009, **77**(3):551-558.
49. Jaillet L, Cortés J, Siméon T: **Sampling-based path planning on configuration-space costmaps.** *IEEE Trans Robot.* 2010, **26**(4):635-646.
50. Jaillet L, Corcho FJ, Pérez JJ, Cortés J: **Randomized tree construction algorithm to explore energy landscapes.** *J Comput Chem.* 2011, **32**(16):3464-3474.

doi:10.1186/1472-6807-13-S1-S2

**Cite this article as:** Al-Bluwi et al.: Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods. *BMC Structural Biology* 2013 **13**(Suppl 1):S2.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



# Simulating ligand-induced conformational changes in proteins using a mechanical disassembly method†

Juan Cortés,<sup>\*ab</sup> Duc Thanh Le,<sup>ab</sup> Romain Iehl<sup>ab</sup> and Thierry Siméon<sup>\*ab</sup>

Received 9th February 2010, Accepted 25th May 2010

First published as an Advance Article on the web 4th June 2010

DOI: 10.1039/c002811h

Simulating protein conformational changes induced or required by the internal diffusion of a ligand is important for the understanding of their interaction mechanisms. Such simulations are challenging for currently available computational methods. In this paper, the problem is formulated as a mechanical disassembly problem where the protein and the ligand are modeled like articulated mechanisms, and an efficient method for computing molecular disassembly paths is described. The method extends recent techniques developed in the framework of robot motion planning. Results illustrating the capacities of the approach are presented on two biologically interesting systems involving ligand-induced conformational changes: lactose permease (LacY), and the  $\beta_2$ -adrenergic receptor.

## Introduction

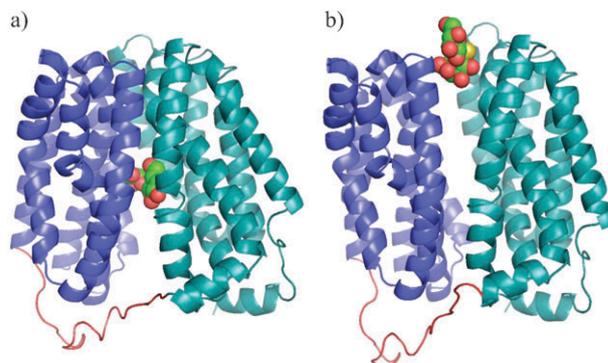
Proteins are flexible macromolecules that fluctuate between nearly isoenergetic folded states.<sup>1</sup> In many cases, conformational changes are associated with their function, and they occur through the interaction with other molecules. For instance, conformational changes are of major importance for protein–ligand and protein–protein recognition.<sup>2,3</sup>

This paper addresses protein conformational changes induced (or required) by the diffusion of a ligand (or substrate/product) molecule inside the protein. An illustrative example is the permeation of lactose through a membrane transport protein (LacY).<sup>4</sup> LacY fluctuates between a conformation where lactose is accessible from the cytoplasm, but the channel toward the periplasmic side is closed (Fig. 1a), and the opposite conformation where the channel is open toward the periplasm and closed in the cytoplasmic side (Fig. 1b). The transition between these two conformational states occurs during lactose diffusion inside the protein.

Despite impressive recent advances on the structural determination of protein motions,<sup>5,6</sup> currently available experimental methods are unable to provide an atomic-resolution structural description of protein conformational changes associated with ligand diffusion. Computational methods are therefore necessary to better understand such processes. However, the time-scale of the ligand diffusion process from a deep active site to the protein surface is out of range for standard molecular dynamics (MD) simulations. Variants of MD methods such as steered molecular dynamics (SMD)<sup>7</sup> and random acceleration molecular dynamics (RAMD)<sup>8</sup> have been proposed for accelerating the simulation of the ligand exit. Both methods introduce an

artificial force in the molecular force field to enhance the ligand motion in a given direction. In SMD simulations, this direction is usually defined by the user through a haptic device. In RAMD simulations, the direction is randomly chosen and iteratively modified after a given number of simulation steps if the ligand gets stuck. Although these methods have been shown to provide biologically relevant information, they remain computationally expensive. Besides, the artificial force introduced for accelerating the simulation may yield biased results about the induced conformational changes, so that the interest of simulating with an accurate molecular force field is partially lost.

This paper presents an alternative method for simulating ligand diffusion motions, together with the possibly induced conformational changes of the protein. Given an initial structure with the ligand docked inside the protein, the proposed method computes a path (*i.e.* continuous sequence of conformations) simulating the ligand exit. Such a path search problem is formulated as a mechanical disassembly problem, where the protein and the ligand are modeled as articulated



**Fig. 1** Lactose permease (LacY) conformational transition. (a) The crystal structure<sup>9</sup> (PDB ID 1PV7), where the substrate is accessible from the cytoplasm. (b) Model of LacY after the conformational change induced by the substrate diffusion toward the periplasm.

<sup>a</sup> CNRS, LAAS, 7 avenue du colonel Roche, F-31077 Toulouse, France

<sup>b</sup> Université de Toulouse, UPS, INSA, INP, ISAE, LAAS, F-31077 Toulouse, France. E-mail: juan.cortes@laas.fr, thierry.simeon@laas.fr

† Electronic supplementary information (ESI) available: Movie of the computed conformational transition of LacY; movies of two different ligand exit paths from  $\beta_2$ -AR. See DOI: 10.1039/c002811h

mechanisms. The main feature of this method is its computational efficiency, enabling to compute large-amplitude conformation transition paths, such as the one illustrated in Fig. 1, in less than one hour of CPU time.

Computing disassembly paths for mechanical parts is an important problem in the fields of robotics and manufacturing engineering. In the last years, randomized search algorithms<sup>10</sup> have been demonstrated to be effective computational tools for disassembly path planning.<sup>11,12</sup> Thanks to their generality, this type of algorithms have also been applied to solve problems in computational structural biology.<sup>13–15</sup> In this framework, the ML-RRT algorithm<sup>16</sup> was introduced as a general method for computing disassembly paths of objects with articulated parts. ML-RRT has been successfully applied in enzyme enantioselectivity studies for computing ligand exit paths considering the flexibility of the protein side-chains.<sup>17,18</sup>

The methodological contribution of this paper is an extension of ML-RRT that enables further introduction of protein flexibility, so that challenging problems involving protein models with flexible backbone segments can be tackled. The improved algorithm is able to consider not only side-chain local flexibility, but also loop or domain motions induced by the ligand along the diffusion pathway. As a proof of concept, the method is applied to two biologically interesting systems involving ligand-induced conformational changes: lactose permease (LacY), and the  $\beta_2$ -adrenergic receptor.

## Methods

### Outline

**Path search problem.** The problem of computing the exit path of a ligand from a protein active site is formulated as a mechanical disassembly problem in which molecules are represented as articulated mechanisms. The degrees of freedom of the molecular models correspond to bond torsions (backbone or side-chains) and to rigid-body motions of atom groups (rigid secondary structure elements). Starting from a given “assembled” (docked) position of the ligand inside the protein, the disassembly problem consists in finding the path leading to a “disassembled” state, where the ligand is located outside the protein. The disassembly path has to be searched in a composite conformational space involving the degrees of freedom of the protein and the ligand. The difficulty for solving such path search problem is due to the very high dimension of this search-space.

**Random diffusion trees.** The conformational exploration algorithm described in this work is derived from the Rapidly-exploring Random Tree (RRT) algorithm,<sup>19</sup> developed in robotics, and which has been demonstrated to perform well for solving complex disassembly problems in constrained spaces. The basic principle of RRT is to iteratively construct a random tree, rooted at a given initial state, and tending to cover the accessible regions of the search-space. The nodes of the tree correspond to states generated by the diffusion process, and the edges correspond to feasible local paths. The RRT construction process is illustrated by Fig. 2 on a simple two-dimensional problem. At each iteration of the

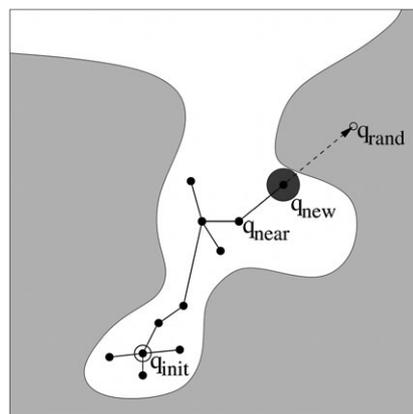


Fig. 2 Illustration of the RRT expansion process.

algorithm, a state  $q_{rand}$  is randomly sampled following a uniform distribution in the search-space. The nearest node in the tree  $q_{near}$  is selected, and an attempt is made to expand it in the direction of  $q_{rand}$ . A new node  $q_{new}$  is generated at the endpoint of the feasible straight-line path (*i.e.* sub-path satisfying motion constraints) from  $q_{near}$  to  $q_{rand}$ . The process is iterated until the final state can be connected to the tree. This tree construction strategy favors an efficient exploration biased toward unexplored regions, while converging to a uniform coverage of the space.<sup>19</sup> This technique performs well for solving moderately high-dimensional problems. However, its performance degrades when applied to very-high-dimensional search-spaces.

**Manhattan-like RRT.** The Manhattan-like RRT (ML-RRT) variant<sup>16</sup> was developed to circumvent this limitation of the basic RRT algorithm for dealing with disassembly problems involving complex articulated objects. The main idea is to facilitate the tree expansion by considering separately two types of conformational parameters, called *active* and *passive*. Active parameters are essential for the disassembly problem, and they are directly treated at each iteration of the algorithm. Passive parameters, however, only need to be treated when they hinder the expansion of active parameters. The advantage of this decoupled treatment, that favors the expansion of the active parameters, is to maintain the exploratory strength of the RRT algorithm while dealing with high-dimensional problems. The ML-RRT algorithm was successfully applied in previous work<sup>17,18</sup> for computing ligand exit paths considering the flexibility of the protein side-chains. For this particular application, the partition of the conformational parameters makes the exploration be focused on the ligand diffusion (active parameters), while the protein side-chain motions (passive parameters) are induced by the ligand motion.

Building on this prior work, we describe below an extension of ML-RRT that enables the simulation of loop/domain motions induced by the ligand diffusion. The proposed generalization of the ML-RRT principle relies on a classification and hierarchization of the different elements in the mechanistic molecular model, receiving each a specific treatment during the exploration.

## Model and parameters

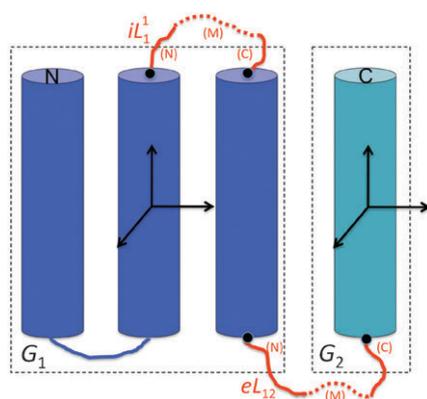
**Mechanistic molecular model.** The proposed method deals with all-atom models of molecules, which are represented as articulated mechanisms. Groups of atoms form the bodies, and the articulations between bodies correspond to bond torsions. The size of the atom groups depends on the level of flexibility allowed to different parts of the molecule. Flexible and rigid regions can be assigned based on structural knowledge. In the present work, flexibility is defined by the user. Note however that the identification of rigid and flexible regions may be automated using computational methods such as FIRST.<sup>20</sup>

Fig. 3 illustrates the mechanistic model of a protein. The following notation is used:

- $G_i$  group: set of rigid secondary structure elements (with flexible side-chains), possibly connected by flexible loops.
- $iL_i^k$  intra-group loop:  $k$ th flexible segment between two secondary structure elements of group  $G_i$ .
- $eL_{i,i+1}$  inter-group loop/linker: flexible segment between secondary structure elements in consecutive groups  $G_i$  and  $G_{i+1}$ .

Each group  $G_i$  holds free rigid body mobility, independently from the other groups. Therefore, loop-closure constraints have to be imposed on flexible segments  $eL_{i,i+1}$  and  $eL_{i-1,i}$  connecting  $G_i$  to its neighboring groups, in order to maintain the molecular chain integrity. As indicated in Fig. 3, several parts are differentiated inside inter- or intra-group loops: the N-terminal and C-terminal segments, and the middle part (M), which is composed by a tripeptide. Such a decomposition is required for the treatment of loop motions that will be explained below. Additionally, geometric (distance and orientation) constraints can be introduced between any pair of elements (rigid groups or loops) in order to model interactions such as hydrogen bonds or disulfide bonds. All these constraints will be satisfied during the conformational exploration.

Side-chains (not represented in the Figure) are generally modeled as flexible elements with freely rotatable bond



**Fig. 3** Schematic representation of a flexible protein model. The three secondary structure elements grouped in  $G_1$  are modeled as a rigid solid. The group  $G_2$  involves only one secondary structure element. The loop/linker  $eL_{1,2}$ , between  $G_1$  and  $G_2$ , is flexible. Loops connecting elements in a group can be flexible or not. Only the intra-domain loop  $iL_1^1$  is flexible in this example.

torsions. By default, the ligand is also fully flexible. Nevertheless, the user can arbitrarily define the flexibility of the ligand and the side-chains.

**Conformational parameters.** The protein conformation is defined by the parameters determining the pose (position and orientation) of all the groups  $G_i$ , the values of the bond torsions in intra- and inter-group loops, and the bond torsions of the side-chains. The conformational parameters of the ligand are the six parameters defining the pose of its reference frame (associated with its center of mass), and the values of the allowed bond torsions.

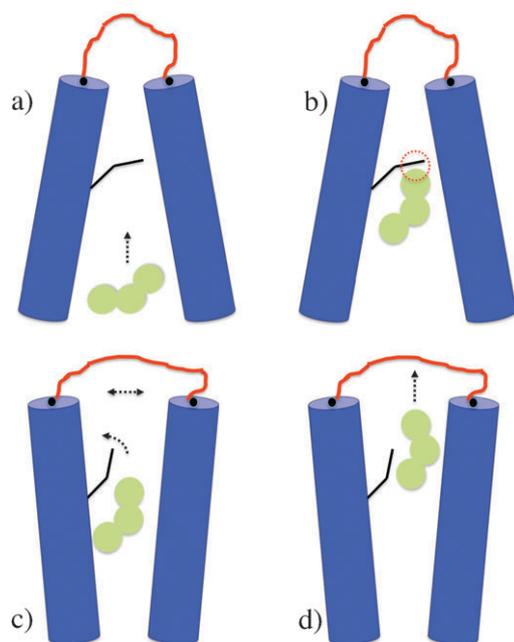
Let  $\mathbf{q}$  denote the array containing the values of all the conformational parameters of the protein and the ligand. The ML-RRT algorithm explores the composite conformational space  $\mathcal{C}$ , which is the set of all conformations  $\mathbf{q}$ . As mentioned above, the conformational parameters are partitioned into *active* and *passive* on the basis of their role in the disassembly problem. Active parameters are essential for carrying out the disassembly task, while passive parameters only need to move if they hinder the progress of the process. Thus, the mobile parts of the molecular model are separated into two lists  $P_{\text{act}}$  and  $P_{\text{pas}}$  containing the active and the passive parts respectively. For a given partition, the conformational parameters are separated into two sets:  $\mathbf{q} = \{\mathbf{q}^{\text{act}}, \mathbf{q}^{\text{pas}}\}$ , where  $\mathbf{q}^{\text{act}}$  is the set of conformational parameters associated with the parts in  $P_{\text{act}}$  and  $\mathbf{q}^{\text{pas}}$  is the set associated with  $P_{\text{pas}}$ . For the protein–ligand disassembly problems addressed in this paper,  $\mathbf{q}^{\text{act}}$  involves the ligand parameters, while  $\mathbf{q}^{\text{pas}}$  concerns the protein flexibility.

Additionally, a mobility coefficient  $\delta \in (0,1]$  is assigned to each passive parameter. This coefficient is used to differentiate passive parts that are allowed to move easily from those that should be moved only if the solution path cannot be found otherwise. By default, the mobility coefficient of all side-chains is set to 1, meaning that they will systematically move if they are identified during the exploration. Lower mobility is allowed to loops and secondary structure groups, with  $\delta = 0.5$  and  $\delta = 0.2$  respectively in the current implementation.

## Conformational exploration algorithm

ML-RRT computes the motion of parts associated with active and passive parameters in a decoupled manner. Fig. 4 provides a simple illustration of the process, which alternates expansion attempts of these parameter subsets.

The ML-RRT algorithm is sketched in Algorithm 1. At each iteration, the motion of active parts is computed first. The function `SampleConf` receives as argument the list of active parts  $P_{\text{act}}$  and samples only the associated parameters  $\mathbf{q}^{\text{act}}$ . Thus, this function generates a conformation  $\mathbf{q}_{\text{rand}}^{\text{act}}$  in a sub-manifold of the conformational space involving the active parameters,  $\mathcal{C}^{\text{act}}$ . The function `NearestNeighbor` selects the node to be expanded  $\mathbf{q}_{\text{near}}$  using a distance metric in  $\mathcal{C}^{\text{act}}$  (*i.e.* involving the ligand pose and its bond torsions). Then, `Expand` performs the expansion of the selected conformation by only changing the active parameters. The returned conformation  $\mathbf{q}_{\text{new}}$  corresponds to the last valid point (*i.e.* satisfying all the geometric constraints) computed along the straight-line path from  $\mathbf{q}_{\text{near}}$  toward  $\{\mathbf{q}_{\text{rand}}^{\text{act}}, \mathbf{q}_{\text{near}}^{\text{pas}}\}$ . If the expansion succeeds



**Fig. 4** Illustration of the decoupled exploration of active and passive parameters within ML-RRT. (a) Expansion of active parameters corresponding to the motion of the ligand. (b) Identification of the passive parts hindering the ligand motion. (c) The expansion of passive parameters yielding the opening motion of the protein. (d) New iteration of the active parameters expansion.

(i.e. the distance from  $\mathbf{q}_{\text{near}}$  to  $\mathbf{q}_{\text{new}}$  is not negligible), a new node and the corresponding edge are added to the tree. The function `Expand` analyzes the collision pairs yielding the stop of the expansion process. If active parts in  $P_{\text{act}}$  collide with potentially mobile passive parts in  $P_{\text{pas}}$ , the list of the involved passive parts  $P_{\text{pas}}^{\text{col}}$  is returned. This information is used in the second stage of the algorithm, which generates the motion of passive parts.

The function `PartsToMove` determines the list  $P_{\text{pas}}^{\text{mov}}$  of passive parts to be moved at one iteration. This function receives as argument the list of colliding passive parts  $P_{\text{pas}}^{\text{col}}$ , and constructs a list with all the parts indirectly involved in the collision based on the kinematic diagram of the molecular model. Fig. 5 illustrates three typical situations. If the ligand motion is hindered by a side-chain in a secondary structure element (Case 1 in Fig. 5), then, the list involves this side-chain and the corresponding group  $G_i$ . When the colliding side-chain is on a flexible loop, then the list involves the side-chain, the loop backbone, and the group  $G_i$  for an intra-group loop  $iL_i$  (Case 2), or the groups  $G_i$  and  $G_{i+1}$  for an inter-group loop  $eL_{i,i+1}$  (Case 3). In all the cases, when a group  $G_i$  is involved in  $P_{\text{pas}}^{\text{mov}}$ , then the backbone of inter-group loops  $eL_{i-1,i}$  and  $eL_{i,i+1}$  (if any) is also considered into the list, since the conformation of these loops needs to be sampled together with the group pose in order to maintain the chain integrity.

The function `PerturbConf` acts on passive parameters. The conformational parameters associated with parts in the list  $P_{\text{pas}}^{\text{mov}}$  are sampled with a probability that depends on their mobility coefficient  $\delta$ , and on the difficulty for expanding  $\mathbf{q}_{\text{near}}$ ,

#### Algorithm 1: Construct\_ML-RRT

```

input      : the conformational space  $C$ ;
              the initial conformation  $\mathbf{q}_{\text{init}}$ ;
              the partition  $\{P_{\text{act}}, P_{\text{pas}}\}$ ;

output    : the tree  $\tau$ ;

begin
   $\tau \leftarrow \text{InitTree}(\mathbf{q}_{\text{init}})$ ;
  while not StopCondition( $\tau$ ) do
     $\mathbf{q}_{\text{rand}}^{\text{act}} \leftarrow \text{SampleConf}(C, P_{\text{act}})$ ;
     $\mathbf{q}_{\text{near}} \leftarrow \text{NearestNeighbor}(\tau, \mathbf{q}_{\text{rand}}^{\text{act}}, P_{\text{act}})$ ;
     $(\mathbf{q}_{\text{new}}, P_{\text{pas}}^{\text{col}}) \leftarrow \text{Expand}(\mathbf{q}_{\text{near}}, \mathbf{q}_{\text{rand}}^{\text{act}})$ ;
    while  $P_{\text{pas}}^{\text{col}} \neq \emptyset$  do
       $P_{\text{pas}}^{\text{mov}} \leftarrow \text{PartsToMove}(P_{\text{pas}}^{\text{col}})$ ;
       $\mathbf{q}_{\text{rand}}^{\text{pas}} \leftarrow \text{PerturbConf}(C, \mathbf{q}_{\text{new}}, P_{\text{pas}}^{\text{mov}}, \mathbf{q}_{\text{near}}, n_{\text{fail}})$ ;
       $(\mathbf{q}'_{\text{new}}, P_{\text{pas}}^{\text{col}}) \leftarrow \text{Expand}(\mathbf{q}_{\text{new}}, \mathbf{q}_{\text{rand}}^{\text{pas}})$ ;
       $P_{\text{pas}}^{\text{col}} \leftarrow P_{\text{pas}}^{\text{col}} \setminus P_{\text{pas}}^{\text{col}}$ ;
       $\mathbf{q}_{\text{new}} \leftarrow \mathbf{q}'_{\text{new}}$ ;
    if not TooSimilar( $\mathbf{q}_{\text{near}}, \mathbf{q}_{\text{new}}$ ) then
      AddNewNode( $\tau, \mathbf{q}_{\text{new}}$ );
      AddNewEdge( $\tau, \mathbf{q}_{\text{near}}, \mathbf{q}_{\text{new}}$ );
       $\mathbf{q}_{\text{near}}.n_{\text{fail}} \leftarrow 0$ ;
    else  $\mathbf{q}_{\text{near}}.n_{\text{fail}} \leftarrow \mathbf{q}_{\text{near}}.n_{\text{fail}} + 1$ ;
  end

```

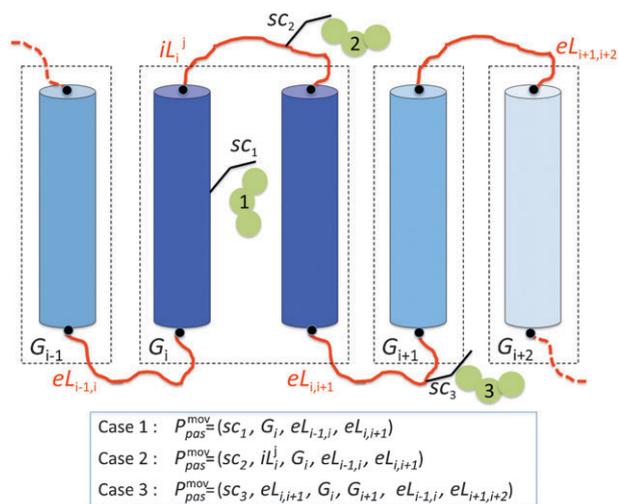
which is estimated by the number of previous expansion failures  $n_{\text{fail}}$ . A parameter is sampled if the following condition is satisfied:

$$\text{Normal Rand}(\mu, \sigma^2) \geq 1 - \delta$$

Where `Normal Rand` returns a random positive real number sampled from a normal distribution with mean  $\mu = 0$  and variance  $\sigma^2 = 0.1 \times n_{\text{fail}}$ . Such a selection strategy maintains a low probability of moving parts with small mobility coefficient (e.g. protein domains) when the diffusion tree grows easily, while the probability is increased when required to unblock the exploration.

The value of the selected passive parameters is perturbed by randomly sampling in a ball centered at  $\mathbf{q}_{\text{near}}$ . Then, an attempt is made to further expand  $\mathbf{q}_{\text{new}}$  toward  $\{\mathbf{q}_{\text{new}}^{\text{act}}, \mathbf{q}_{\text{rand}}^{\text{pas}}\}$ . Note that only parts in  $P_{\text{pas}}^{\text{mov}}$  associated with the perturbed parameters move during this tree expansion. The function `Expand` returns a list  $P_{\text{pas}}^{\text{col}}$  of blocking parts involved in collisions with moving passive parts. If this list contains new passive parts (not contained in  $P_{\text{pas}}^{\text{col}}$ ), the process generating passive part motions is iterated. Such a possible cascade of passive part motions is needed to solve problems where passive parts indirectly hinder the motion of the active ones because they block other passive parts.

The algorithm is iterated until the problem is solved, or when a `StopCondition` determines that the solution cannot be found. The problem is considered to be solved when a conformation with the ligand outside the protein is reached. Failure is returned if a solution is not found after a given maximum number of iterations. Once the random diffusion tree is constructed, the solution path is simply obtained by tracing back the edges from the goal node (“disassembled” state) to the root node (“assembled” state). Finally, a randomized



**Fig. 5** Determination of the list of passive parts to be moved  $P_{\text{pas}}^{\text{mov}}$  based on the contacts with active parts and on the kinematic diagram of the protein model. Three typical cases are illustrated.

path smoothing post-processing<sup>‡</sup> is performed in the composite space of all the parameters, so that simultaneous motions of the ligand and the protein are obtained in the final path, instead of the alternate motions resulting from the Manhattan-like exploration strategy.

### Geometric constraints verification

During the conformational exploration, a set of geometric constraints have to be checked (*e.g.* collision avoidance, hydrogen/disulfide bond integrity) or reinforced (*e.g.* loop closure). These constraints are explained below.

**Collision avoidance.** The main geometric constraint to be verified during the conformational exploration is the avoidance of atom overlaps. The atoms are represented by rigid spheres with a percentage of van der Waals radii. Considering a percentage of the van der Waals equilibrium distance ensures that only energetically infeasible conformations are rejected by the collision checker. The value of 80% is often used in techniques that geometrically check atom overlaps.<sup>22</sup> Collisions are checked between the ligand and the protein, as well as internal collisions between mobile parts of each molecule. The collision test is done inside the function *Expand*, which performs the local expansion motion. Our implementation builds on the efficient *BioCD* algorithm,<sup>23</sup> specially designed for articulated molecular models. *BioCD* uses hierarchical data structures to approximate the shape of the molecules at successive levels of detail, making the number of atom pairs tested for collision to be significantly reduced.

**Loop closure.** The functions *SampleConf* and *PerturbConf* perform a specific sampling procedure of loop conformations, taking into account loop closure constraints. Once the pose parameters of all groups  $G_i$  have been sampled, the Random Loop Generator (RLG) algorithm<sup>24</sup> is applied to sample the

<sup>‡</sup> The probabilistic path shortening method<sup>21</sup> was used for path smoothing.

backbone torsions of the N-terminal and C-terminal segments of each loop. This iterative algorithm, based on simple geometric operations, biases the sampling of these chain segments toward conformations with a high probability of satisfying the loop closure constraint. The constraint is reinforced within the function *Expand*, which applies an inverse kinematics method<sup>25</sup> to compute the bond torsions of the tripeptide in the middle loop part (M) for the conformations along the local expansion motion.

**Hydrogen bonds and disulfide bonds.** These structural constraints can be considered within the mechanistic molecular model. Indeed, they are modeled as distance and angle constraints between the bonded atoms. For hydrogen bonds, the distance  $d$  between the donor and the acceptor atoms, and the bond angle  $\theta$ , must remain within a given range. For instance, for  $\text{O-H}\cdots\text{N}$  bonds:  $d_{\text{O-N}} \in [2.5 \text{ \AA}, 3.8 \text{ \AA}]$  and  $\theta_{\text{O-H-N}} \in [110^\circ, 180^\circ]$ . Disulfide bonds also imply bond length and bond angle constraints between the involved S and C atoms. Additionally, the S-S bond torsion  $\gamma$  is restricted around  $90^\circ$ . The ranges by default are  $d_{\text{S-S}} \in [1.8 \text{ \AA}, 2.2 \text{ \AA}]$ ,  $\theta_{\text{C-S-S}} \in [100^\circ, 130^\circ]$ , and  $\gamma_{\text{S-S}} \in [60^\circ, 120^\circ]$ . All these constraints are checked within the function *Expand*.

## Results and discussion

This section presents results obtained with the proposed method on two biologically interesting systems involving ligand-induced conformational changes. In the first one, the mechanism of sugar permeation through LacY involves a large-amplitude relative motion of transmembrane domains. In the second system, the access/exit of a ligand to the active site of the  $\beta_2$ -adrenergic receptor is related with side-chain motions, loop motions and transmembrane domain rearrangements. The presented results are not aimed to provide new insights into these biological systems, but to serve as a proof of concept and to show the interest of the proposed approach.

The method was implemented within our software prototype *BioMove3D*. *PyMOL*<sup>26</sup> was used for viewing molecular models. The computing times reported below correspond to tests run on a single AMD Opteron 148 processor at 2.6 GHz.

### Lactose permease

Lactose permease (LacY) is a transport protein that transduces electrochemical proton gradients into sugar concentration gradients across the cell inner membrane.<sup>4</sup> LacY is composed of two main domains:<sup>9</sup> the N-domain involving helices I–VI, and the C-domain involving helices VII–XII. The two domains are connected by a long loop containing more than 20 residues. For carrying out its function, LacY is supposed to alternate between two conformational states: the inward-open state, where the substrate is accessible from the cytoplasm, and the outward-open state, where the access is possible from the periplasmic side. However, only the structure of the inward-open conformational state of LacY has been solved by X-ray crystallography.

Different approaches have been used to analyze the conformational transition pathway toward the outward-open state. In particular, experimental studies using double

electron-electron resonance (DEER)<sup>27</sup> suggest that the conformational transition can be mainly described as a rigid-body rotation of the C-domain and the N-domain. Based on such structural knowledge, the mechanistic model of LacY was simplified by considering a rigid backbone for the C- and N- domains. Flexibility was allocated to the loop between helices VI and VII, and to all the protein side-chains. Thus, the mechanical model contains two main groups  $G_1$  and  $G_2$  associated with the C-domain and the N-domain respectively, and an inter-domain loop  $eL_{1,2}$ . The X-ray structure of LacY of *Escherichia coli*<sup>9</sup> (PDB ID 1PV7), corresponding to the inward-open conformation, used as starting point in this work, contains a bound substrate homologue TDG (see Fig. 1a). The substrate molecule was modeled with full flexibility, and it could freely rotate and translate by 50 Å in any direction excepting the direction to the cytoplasm (only 5 Å were permitted in this direction in order to force the exit toward the periplasmic side). Overall, the mechanistic model of LacY-TDG contains 775 degrees of freedom: 12 correspond to the rigid-body motion of the C- and N- domains, 75 to the backbone torsions of the inter-domain loop, 678 to the protein side-chains, and 10 to the substrate mobility and flexibility.

The ML-RRT algorithm was applied to compute the exit pathway of TDG toward the periplasmic side, which involves the conformational transition of LacY. The computing time of a run was about 1 h on a single processor. Such high computational performance is worth to be noted since it represents an important feature of the proposed approach compared to the very long computing times required by other simulation methods such as molecular dynamics. The algorithm was run 10 times in order to analyze a possible variability of results associated with the randomized exploration procedure. All the runs yielded very similar results with regard to the protein conformational change. The obtained “disassembled” conformation, with the ligand outside the protein and LacY in a outward-open state, is represented in Fig. 1b. A movie of the computed conformational transition is provided as ESI.† As it has been pointed out by prior studies,<sup>27</sup> the substrate exit requires the rotation of the two domains. In our results, the observed rotation between the domains is around 20°. Although this is smaller than the 60° suggested by DEER experiments, the overall motion is alike. The comparison of the variation of distances between some residue pairs in the inward- and outward- faces of LacY (see Table 1) shows an approximate overall ratio of 1/3 between the values measured by DEER and our results. The explanation for this quantitative difference is that ML-RRT tends to produce the minimal conformational change required for the molecular disassembly, while larger motions may occur in reality. Interestingly, the distance between residues Ile40 and Asn245 in the outward-open conformation computed by ML-RRT is of approximately 15 Å, which has been shown by cross-linking experiments<sup>28</sup> to be the minimal distance between these residue positions for guaranteeing the activity of LacY.

In other recent studies,<sup>29</sup> steered molecular dynamics (SMD) simulations have been carried out to better understand the physical mechanisms of lactose permeation at the atomic level. SMD results provide detailed information about the interactions between lactose and LacY residues during

permeation. Such information cannot be directly provided by our method, since it does not consider accurate energy functions. However, a straightforward geometric analysis of the paths obtained by ML-RRT can provide the list of residues that the ligand has encountered during its diffusion. The diagram in Fig. 6 represents the residues encountered by the ligand along the path toward the periplasm. A contact between the ligand and a residue side-chain was recorded if the distance between the surface of van der Waals spheres modeling their atoms was below 1 Å. The diagram shows the percentage of times that a contact appeared over the set of 10 paths. Contacts were recorded for three segments of the path: the beginning (0–10 Å), where the ligand is close to its location in the crystal structure, the middle part (10–20 Å), and the final part (above 20 Å), where TDG is near the periplasm. Remarkably, all the residues identified by SMD simulations<sup>29</sup> as interacting residues (through side-chain hydrogen bonds or hydrophobic interactions) appear in the diagram, with the exception of Asp36. Note however that this residue is on the periplasmic surface of the protein. On the other side, only one residue (Thr265) appearing in the contact diagram with a significant percentage is not reported in the referred work. Such an impressive consistency with results of SMD simulations confirms the validity and the potential interest of our approach.

### $\beta_2$ -Adrenergic receptor

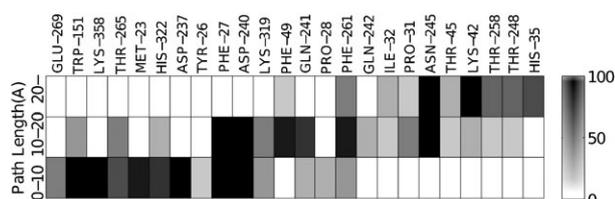
The  $\beta_2$ -adrenergic receptor ( $\beta_2$ -AR) is a membrane protein belonging to the superfamily of the G-protein-coupled receptors (GPCRs),<sup>30</sup> which activate signal transduction inside the cell in response to the binding of hormones and neurotransmitters in the extracellular region. GPCRs are important therapeutic targets for a large class of diseases. Therefore, numerous studies have been devoted to this family of proteins, aiming to better understand their activation/deactivation mechanism. However, many questions remain. In particular, little is known about the functional role of extracellular loops, and about their possible conformational coupling to ligand binding.<sup>31</sup> One major difficulty comes from the lack of structural information inherent to membrane proteins.

A high-resolution crystal structure of  $\beta_2$ -AR has been recently obtained<sup>32</sup> (PDB ID 2RH1). The crystal structure also contains a molecule of carazolol, a partial inverse agonist, in the protein active site. This receptor–ligand structure is the starting point of the conformational analysis presented below. The structure is represented in Fig. 7, using standard notation for the structural elements. Like all GPCRs,  $\beta_2$ -AR contains seven transmembrane helices, which were modeled as rigid groups  $G_i$ . The intracellular and extracellular loops were modeled as flexible elements  $eL_{i,i+1}$ . All the side-chains and the ligand were considered to be fully flexible. The number of degrees of freedom of the whole model is 703: 42 of them correspond to the rigid-body motion of the seven transmembrane helices, 159 to the backbone torsions of the five loops, 490 to the protein side-chains, and 12 to the ligand mobility and flexibility.

The ML-RRT algorithm was applied to compute the exit pathway of carazolol from the active site of  $\beta_2$ -AR. A first set

**Table 1** Distance variation between residue pairs in LacY

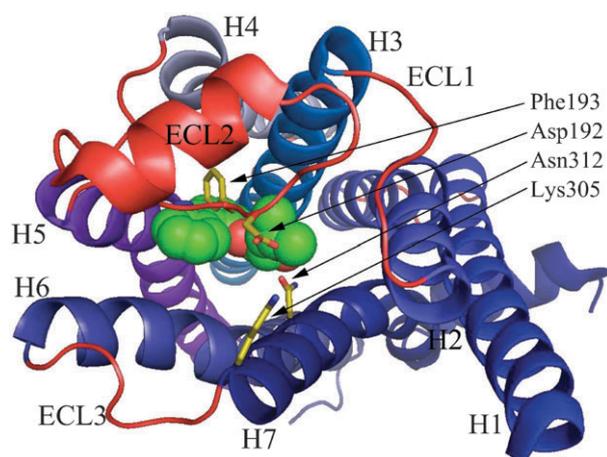
Residue pair	Inward-open	Outward-open (experimental) <sup>27</sup>	Outward-open (simulation)
73–401	41 Å	27 Å	36.9(±1.0) Å
73–340	36 Å	21 Å	31.0(±1.3) Å
136–340	34 Å	17 Å	28.7(±1.4) Å
137–340	32 Å	16 Å	26.7(±1.4) Å
136–401	40 Å	24 Å	35.6(±1.3) Å
137–401	38 Å	22 Å	33.5(±1.4) Å
105–310	34 Å	41 Å	38.0(±1.6) Å
164–310	27 Å	43 Å	32.8(±1.4) Å
164–375	33 Å	49 Å	35.8(±1.2) Å



**Fig. 6** List of residues whose side-chain was encountered by the substrate during its diffusion toward the periplasm. For facilitating interpretation, the pathway is divided into three segments. The greyscale represents the percentage of times that the contact appears over the set of 10 runs.

of 10 runs revealed some variability on the trajectories followed by the ligand. Thus, the algorithm was run 60 times in order to do a more accurate statistical analysis of results. The 60 paths were obtained in 2 h of computing time (each run takes an average of 2 min on a single processor). These exit paths can be divided into two main clusters. In one class of paths, which we refer to as “left-hand” paths, carazolol exits between transmembrane helices H5, H6 and H7. In the other class, called “right-hand” paths, the ligand exits between H2, H3 and H7. The two clusters can be separated by an axis traced between residues Asp192 and Lys305, which form a salt bridge in the crystal structure. Fig. 8 shows snapshots of the ligand exit path for each path class. Movies of these paths are provided as ESI.† Interestingly, these two classes of exit paths have also been observed in prior studies<sup>33</sup> based on random acceleration molecular dynamics (RAMD) simulations. A quantitative comparison can be done between results obtained with ML-RRT and RAMD. The most significant comparable result is that both approaches suggest that left-hand and right-hand exit paths are approximately equiprobable. Indeed, 31/60 of the ML-RRT solutions correspond to left-hand, and 29/60 to right-hand paths. Another result from RAMD simulations concerns the recurrent breakage of the salt bridge Asp192–Lys305 during ligand exit. Paths computed with ML-RRT show a significant motion of the side-chains of these two residues, which lead to the salt bridge breakage for most of the 60 paths. However, in some of the left-hand paths, the ligand exits with only a slight perturbation in the conformation of Asp192 and Lys305. The interpretation is that it is geometrically possible for the ligand to exit between helices H5, H6 and H7 without breaking the salt bridge.

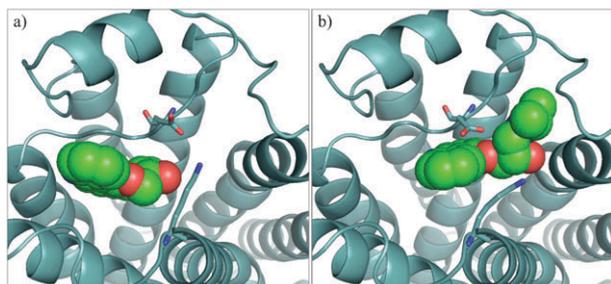
A further comparison between left-hand and right-hand paths obtained with ML-RRT displays other interesting



**Fig. 7** Structure of  $\beta_2$ -AR with carazolol bound in the protein active site viewed from the extracellular side. The secondary structure elements and important residues are displayed on the image.

differences. The first one concerns the orientation of the ligand. In most of the left-hand paths, the ring head of carazolol reaches first the protein surface (see Fig. 8a). Contrarily, the ring and the alkylamine–alcohol tail exit almost simultaneously in most of the right-hand paths (Fig. 8b). A possible interpretation may be that one of the pathways could be preferred for the exit of the ligand, while the other could be more suited to the access. A more accurate analysis of the paths computed by ML-RRT would be required to reinforce such a suggestion. Note however that RAMD simulations from a putative ligand-free model of  $\beta_2$ -AR<sup>33</sup> suggest that carazolol enters the receptor between helices H2, H3 and H7, with its ring head diving first.

Another interesting difference between the two classes of exit paths concerns the conformational changes of the extracellular loop ECL2 induced by the ligand exit. As shown in Fig. 9, right-hand paths imply, on average, a more significant motion of ECL2 than left-hand paths. Note that although the loop ECL2 of  $\beta_2$ -AR is very long, its conformation is constrained by two disulfide bonds, one between residues in the loop (Cys184–Cys190), and one between the loop and H3 (Cys106–Cys191). Thus, in any case, this loop cannot undergo large conformational changes. The observed relationship between right-hand paths and ECL2 flexibility has been confirmed by tests performed on a model of  $\beta_2$ -AR only considering side-chain flexibility. Using this rigid-backbone model, the

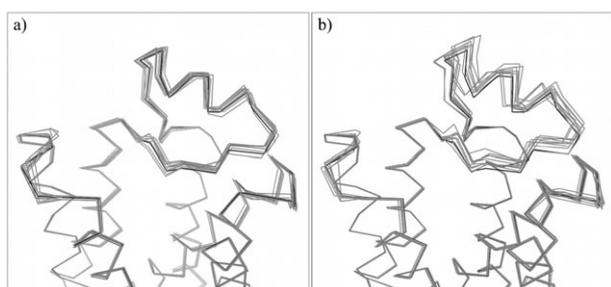


**Fig. 8** Snapshots of the ligand exit from  $\beta_2$ -AR following the left-hand pathway (a), and the right-hand pathway (b).

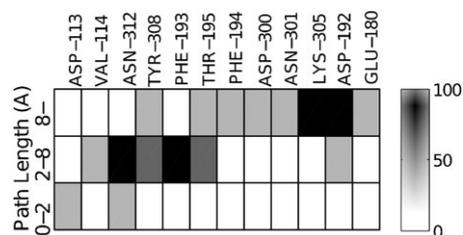
ligand exited through the left-hand pathway in 90% of the ML-RRT runs. These results suggest that right-hand access/exit paths involve a more important interaction between the ligand and ECL2 than left-hand paths. Note that recent studies on GPCRs show important roles of ECL2. Indeed, it can be required for ligand binding,<sup>34</sup> and its motion can be involved in the activation mechanism.<sup>35</sup>

The analysis of contacts between carazolol and  $\beta_2$ -AR residues along the set of 60 exit pathways computed with ML-RRT was performed using the technique described above for the study of LacY. Fig. 10 shows the list of residues whose side-chain was encountered by the ligand. For clarity reasons, the Figure only reports contacts that appeared in more than 30% of the paths. Four residues are clearly highlighted in the diagram: Asp192, Phe193, Lys305, and Asn312. The positions of these residues are indicated in Fig. 7. Two of them, Asp192 and Lys305, form the aforementioned salt bridge, which is broken during the ligand diffusion. Phe193, which is located on ECL2, has also been identified as an important residue in related works. Results of RAMD simulations<sup>33</sup> suggest that this aromatic residue may participate in the ligand entry and stabilization in the active site of  $\beta_2$ -AR. Recent NMR experiments<sup>31</sup> have shown that inverse agonists induce a conformational change of this residue. Finally, Asn312 is an important residue for the stabilization of carazolol in the active site through a polar interaction with its alkylamine-alcohol tail.

Overall, the presented results show that structural information on the access/exit of carazolol to the active site of  $\beta_2$ -AR provided by ML-RRT is in agreement with results of other experimental and computational studies.



**Fig. 9** Superposition of the initial structure of  $\beta_2$ -AR (black) and conformations induced by the ligand exit (grey) following the left-hand pathway (a), and the right-hand pathway (b).



**Fig. 10** List of residues whose side-chain was encountered by the ligand along the exit pathway. For facilitating interpretation, the pathway is divided into three segments. The grey-scale represents the percentage of times that the contact appears over the set of 60 runs. Only contacts that appeared in more than 30% of the paths are displayed.

## Conclusion

The results in this paper show that a mechanistic approach to molecular simulations may lead to the development of efficient computational methods, able to provide relevant information on the interaction of biological molecules. The proposed algorithm, ML-RRT, is a novel and fast method for simulating ligand diffusion inside flexible models of proteins. Indeed, ML-RRT generates long (20–30 Å) diffusion paths within tens of minutes of computing time on a single processor, which is remarkably short compared to the time required by MD-based methods. Such a high computational performance is achieved thanks to the efficiency of the conformational exploration method that operates on geometric models of molecules. Geometrically feasible paths are a reasonably good approximation that provides very useful information. Furthermore, as shown in prior work,<sup>13</sup> the approximate solution path can also be efficiently refined with standard molecular modeling tools (*e.g.* energy minimization) in order to perform a more accurate energetic analysis. However, a current limitation with this approach is that non-bonding interactions such as electrostatic and hydrophobic interactions are neglected when computing the approximate solution path, although they can play some role in the conformational transitions. The accuracy of the approach can be further improved by integrating energy computations during the conformational exploration. For this, we are currently investigating the combination of ML-RRT with a recent algorithm<sup>36</sup> developed for exploring cost-spaces to compute low-energy paths. Other future work involves the improvement of the method to better deal with full molecular flexibility during protein–ligand interactions. We also expect to extend the method for its application to the modeling of protein–protein interactions.

## Acknowledgements

This work has been partially supported by the French National Agency for Research (ANR) under project GlucoDesign, and by the Région Midi-Pyrénées under project Amylo.

## References

- 1 H. Frauenfelder, S. G. Sligar and P. G. Wolynes, *Science*, 1991, **254**, 1598–1603.
- 2 H. A. Carlson, *Curr. Pharm. Des.*, 2002, **8**, 1571–1578.

- 3 M. F. Lensink and R. Méndez, *Curr. Pharm. Biotechnol.*, 2008, **9**, 77–86.
- 4 H. R. Kaback, M. Sahin-Tóth and A. B. Weinglass, *Nat. Rev. Mol. Cell Biol.*, 2001, **2**(8), 610–620.
- 5 G. Katona, P. Carpentier, V. N., P. Amara, V. Adam, J. Ohana, N. Tsanov and D. Bourgeois, *Science*, 2007, **316**, 449–453.
- 6 P. Schanda, V. Forge and B. Brutscher, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 11257–11262.
- 7 S. Izrailev, S. Stepaniants, B. Israelewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers and K. Schulten, in *Computational Molecular Dynamics: Challenges, Methods, Ideas. Vol. 4 of Lecture Notes in Computational Science and Engineering*, Springer-Verlag, Berlin, 1998, pp. 39–65.
- 8 S. K. Ludemann, V. Lounnas and R. C. Wade, *J. Mol. Biol.*, 2000, **303**, 797–811.
- 9 J. Abramson, I. Smirnova, V. Kasho, G. Verner, H. R. Kaback and S. Iwata, *Science*, 2003, **301**, 610–615.
- 10 S. M. LaValle, *Planning Algorithms*, Cambridge University Press, New York, 2006.
- 11 E. Ferré and J.-P. Laumond, *Proc. IEEE Int. Conf. Robot. Automat.*, 2004, 3149–3154.
- 12 D. T. Le, J. Cortés and T. Siméon, *Proc. IEEE Int. Conf. Automat. Sci. Eng.*, 2009.
- 13 J. Cortés, T. Siméon, V. Ruiz, D. Guieysse, M. Remaud and V. Tran, *Bioinformatics*, 2005, **21**, i116–i125.
- 14 A. Enosh, S. J. Fleishman, N. Ben-Tal and D. Halperin, *Bioinformatics*, 2007, **23**, e212–e218.
- 15 S. Kirillova, J. Cortés, A. Stefaniu and T. Siméon, *Proteins: Struct., Funct., Bioinf.*, 2008, **70**, 131–143.
- 16 J. Cortés, L. Jaillet and T. Siméon, *IEEE Trans. Robot.*, 2008, **24**, 475–481.
- 17 D. Guieysse, J. Cortés, S. Puech-Guenot, S. Barbe, V. Lafaquière, P. Monsan, T. Siméon, I. André and M. Remaud-Siméon, *ChemBioChem*, 2008, **9**, 1308–1317.
- 18 V. Lafaquière, S. Barbe, S. Puech-Guenot, D. Guieysse, J. Cortés, P. Monsan, T. Siméon, I. André and M. Remaud-Siméon, *ChemBioChem*, 2009, **10**, 2760–2771.
- 19 S. M. LaValle and J. J. Kuffner, in *Algorithmic and Computational Robotics: New Directions (WAFR2000)*, ed. A. K. Peters, Boston, 2001, pp. 293–308.
- 20 S. Wells, S. Menor, B. Hespeneide and M. F. Thorpe, *Phys. Biol.*, 2005, **2**, 127–136.
- 21 S. Sekhavat, P. Svestka, J.-P. Laumond and M. H. Overmars, *Int. J. Rob. Res.*, 1998, **17**(8), 840–857.
- 22 M. A. DePristo, P. I. W. de Bakker, S. C. Lovell and T. L. Blundell, *Proteins: Struct., Funct., Genet.*, 2003, **51**, 41–55.
- 23 V. Ruiz de Angulo, J. Cortés and T. Siméon, in *Robotics: Science and Systems*, MIT Press, Cambridge, 2005, pp. 6–11.
- 24 J. Cortés, T. Siméon, M. Remaud-Siméon and V. Tran, *J. Comput. Chem.*, 2004, **25**(7), 956–967.
- 25 M. Renaud, in *Current Advances in Mechanical Design and Production VII*, Pergamon, New York, 2000, pp. 57–66.
- 26 W. L. DeLano, <http://www.pymol.org>.
- 27 I. Smirnova, V. Kasho, J.-Y. Choe, C. Altenbach, W. L. Hubbell and H. R. Kaback, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16504–16509.
- 28 Y. Zhou, L. Guan, J. A. Freites and H. R. Kaback, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 3774–3778.
- 29 M. Ø. Jensen, Y. Yin, E. Tajkhorshid and K. Schulten, *Biophys. J.*, 2007, **93**, 92–102.
- 30 G. Vauquelin and B. von Mentzer, *G Protein-Coupled Receptors Molecular Pharmacology from Academic Concept to Pharmaceutical Research*, John Wiley & sons Ltd., Chichester, 2007.
- 31 M. P. Bokoch, Y. Zou, S. G. F. Rasmussen, C. W. Liu, R. Nygaard, D. M. Rosenbaum, J. J. Fung, H.-J. Choi, F. S. Thian, T. S. Kobilka, J. D. Puglisi, W. I. Weis, L. Pardo, R. S. Prosser, L. Mueller and B. K. Kobilka, *Nature*, 2010, **463**, 108–112.
- 32 V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka and R. C. Stevens, *Science*, 2007, **318**, 1258–1265.
- 33 T. Wang and Y. Duan, *J. Mol. Biol.*, 2009, **392**, 1102–1115.
- 34 V. A. Avlani, K. J. Gregory, C. J. Morton, M. W. Parker, P. M. Sexton and A. Christopoulos, *J. Biol. Chem.*, 2007, **282**, 25677–25686.
- 35 S. Ahuja, V. Hornak, E. C. Yan, N. Syrett, J. A. Goncalves, A. Hirshfeld, M. Ziliox, T. P. Sakmar, M. Sheves, P. J. Reeves, S. O. Smith and M. Eilers, *Nat. Struct. Mol. Biol.*, 2009, **16**, 168–175.
- 36 L. Jaillet, J. Cortés and T. Siméon, *IEEE Trans. Robot.*, 2010.

# Control of Lipase Enantioselectivity by Engineering the Substrate Binding Site and Access Channel

Vincent Lafaquière,<sup>[a]</sup> Sophie Barbe,<sup>[a]</sup> Sophie Puech-Guenot,<sup>[a]</sup> David Guieysse,<sup>[a]</sup> Juan Cortés,<sup>[c]</sup> Pierre Monsan,<sup>[a, b]</sup> Thierry Siméon,<sup>[c]</sup> Isabelle André,<sup>[a]</sup> and Magali Remaud-Siméon<sup>\*[a]</sup>

Lipase from *Burkholderia cepacia* (BCL) has proven to be a very useful biocatalyst for the resolution of 2-substituted racemic acid derivatives, which are important chiral building blocks. Our previous work showed that enantioselectivity of the wild-type BCL could be improved by chemical engineering of the substrate's molecular structure. From this earlier study, three amino acids (L17, V266 and L287) were proposed as targets for mutagenesis aimed at tailoring enzyme enantioselectivity. In the present work, a small library of 57 BCL single mutants targeted on these three residues was constructed and screened for enantioselectivity towards (*R,S*)-2-chloro ethyl 2-bromophe-

nylacetate. This led to the fast isolation of three single mutants with a remarkable tenfold enhanced or reversed enantioselectivity. Analysis of substrate docking and access trajectories in the active site was then performed. From this analysis, the construction of 13 double mutants was proposed. Among them, an outstanding improved mutant of BCL was isolated that showed an *E* value of 178 and a 15-fold enhanced specific activity compared to the parental enzyme; thus, this study demonstrates the efficiency of the semirational engineering strategy.

## Introduction

A great challenge for pharmaceutical and industrial chemistry of the next decades will be the discovery and development of fast and economic resolution processes. To obtain pure enantiomers, various methods such as diastereomeric salts, racemate resolution, or asymmetric synthesis with chiral auxiliaries and chiral catalysts can be used.<sup>[1–4]</sup> Among the different approaches, biocatalysis based on the enzymatic resolution of racemic mixtures remains a method of choice. In particular, lipases (triacylglycerol hydrolases, EC 3.1.1.3) are among the most employed catalysts in organic synthesis to catalyse the kinetic resolution of a wide range of substrates and yield optically pure compounds.

Recent advances in tailoring enzymes for high activity and selectivity and their combined use with chemocatalytic reactions has expanded the role of biocatalysis to produce enantiopure compounds from racemic mixtures. Directed evolution or rational design techniques have proven to be successful for the development of lipases with either enhanced or inverted enantioselectivity compared to their parental enzymes.<sup>[5–7]</sup> Amongst lipases used for racemic resolution, lipase from *Burkholderia cepacia* (BCL) has proven to be very useful for the resolution of racemic mixtures of primary and secondary alcohols<sup>[8]</sup> and carboxylic acids.<sup>[9,10]</sup> In a previous study<sup>[11,12]</sup> focused on the understanding of the factors involved in racemic resolution of 2-substituted carboxylic acids, it was suggested that given the topology of BCL active site, which is located at the bottom of a narrow 17 Å pocket, the enzyme ligand affinity and, by extension enantioselectivity, could be affected by the accessibility of the substrate to the catalytic site and the diffi-

culty encountered by the substrate in adopting a productive conformation at the reaction site.

Path-planning algorithms, originating from robotics and adapted for the investigation of various molecular motion problems,<sup>[13–22]</sup> were previously applied for investigating the access/exit of ligands to the BCL active site.<sup>[11]</sup> The amino acids hindering the displacement of (*R,S*) enantiomers along the active site pocket were highlighted by using the atom distance information collected during the path searches by the collision detector BioCD<sup>[23]</sup> integrated in the software prototype Bio-Move3D.<sup>[16,24]</sup> On the basis of the analysis of contacts found along the computed access, two amino acids, L17 and V266, were identified as playing a major role in the discrimination of the pairs of enantiomers and were proposed as targets for mutagenesis aiming at tailoring enzyme enantioselectivity.

[a] V. Lafaquière, Dr. S. Barbe, Dr. S. Puech-Guenot, Dr. D. Guieysse, Prof. P. Monsan, Dr. I. André, Prof. M. Remaud-Siméon  
Université de Toulouse; INSA, UPS, INP, LISBP  
135 Avenue de Rangueil, 31077 Toulouse (France)  
CNRS, UMR5504  
31400 Toulouse (France)  
INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés  
31400 Toulouse (France)  
E-mail: magali.remaud@insa-toulouse.fr

[b] Prof. P. Monsan  
Institut Universitaire de France  
103 Boulevard Saint-Michel, 75005 Paris (France)

[c] Dr. J. Cortés, Dr. T. Siméon  
LAAS-CNRS, Université de Toulouse  
7, avenue du Colonel-Roche, 31077 Toulouse (France)

In the present work, we have constructed libraries of variants in which amino acids identified by path analysis and located both in the bottleneck and at the entrance of BCL active site were targeted. The libraries were first screened on the basis of enzyme activity towards the substrate *para*-nitrophenyl butyrate (*p*NPB). Then, the enantioselectivity of the mutants towards the substrate (*R,S*)-2-chloro ethyl 2-bromophenylacetate (**1**) was experimentally determined. Here, we also discuss the interpretation of molecular modelling results with regards to understanding BCL enantioselectivity.

## Results and Discussion

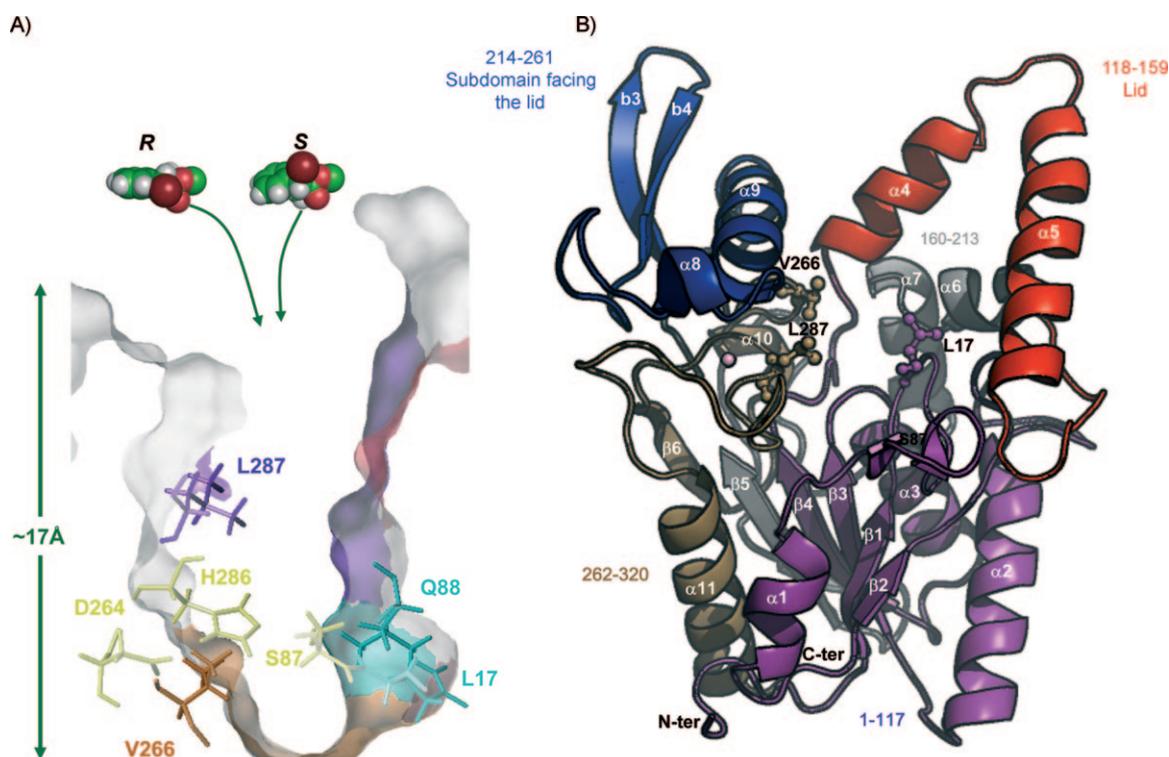
### Construction of the single mutant library

Three amino acids were chosen as mutagenesis targets: L17, V266 and L287 (Figure 1). The selection of L17 and V266 was based on a previous study<sup>[11]</sup> in which we showed that the bottleneck formed by these residues in BCL active site influences both the trajectory and the positioning of various two substituted-acids. The third amino acid, L287, was selected to examine the possible role of a residue located along the access channel but farther from the catalytic centre. Notably, mutation of L287 residue was found in the V266L-L287I-F221L variant of *B. cepacia* lipase (Ps-FVL mutant from Amano CO., Japan), which displayed an enantioselectivity enhanced by

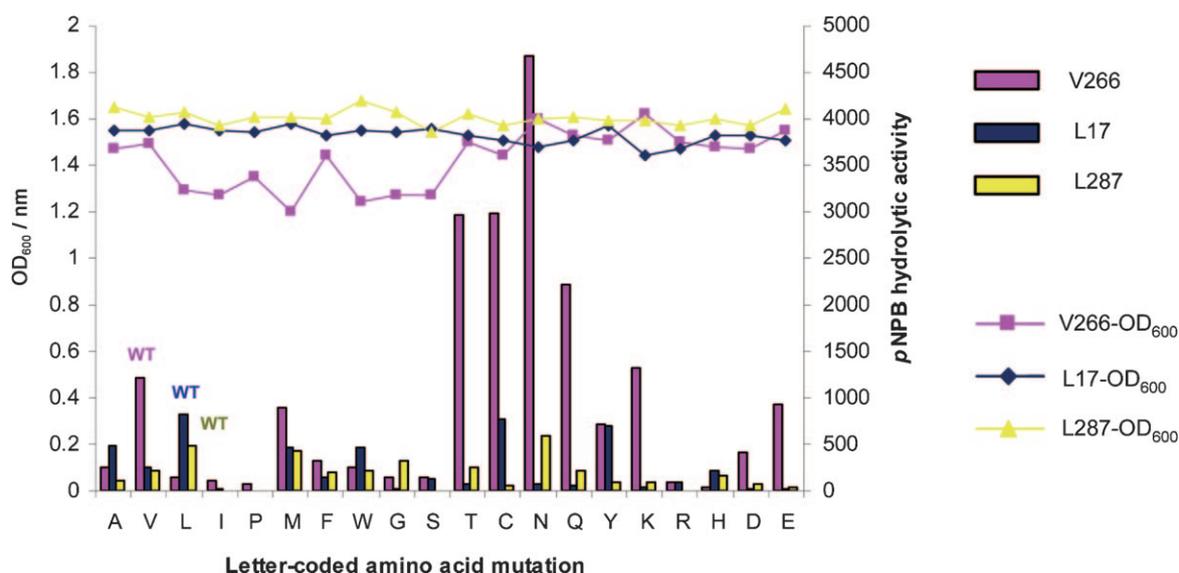
200-fold as compared to the wild-type BCL for the resolution of (*R,S*)-2-bromophenyl acetic acid ethyl ester.<sup>[12]</sup> To further investigate the role of these residues on enzyme enantioselectivity, the three amino acids (L17, V266 and L287) were systematically replaced by the 19 other possible amino acid residues to obtain a library of 57 single mutants.

### Preliminary screening on *p*NPB substrate

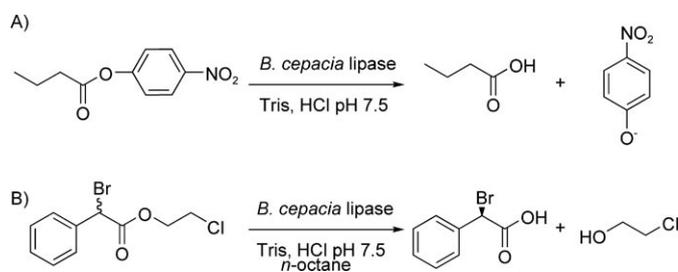
The variants were first grown in deep-well plates. As shown in Figure 2, bacterial growth was rather homogenous for all of them. Following cell lysis, the mutants were screened for their ability to hydrolyse *p*NPB (Scheme 1 A), a substrate easily monitored by spectrophotometry at 405 nm and often used to determine lipase activity.<sup>[25]</sup> The variability of *p*NPB hydrolytic activity found amongst the variants is shown in Figure 2. Overall, ten variants out of 57 appeared more active on *p*NPB than the wild-type BCL. In particular, replacement of V266 by polar and uncharged Thr, Cys, Asn or Gln amino acids led to variants displaying a two- to four-fold enhanced hydrolytic activity, with the highest enhancement being reached for the V266N variant (Figure 2). In contrast, almost all the mutations at positions 17 and 287 had either a neutral or a negative effect and reduced the hydrolytic activity.



**Figure 1.** Representation of *B. cepacia* lipase structure. A) Cross-section view of the enzyme active site. Catalytic triad residues are D264, H286, S87. Amino acid residues forming the oxyanion hole are L17 and Q88. The three amino acid residues selected as mutagenesis targets are: L17, V266 and L287. B) The overall fold of BCL is shown in a cartoon representation in which  $\beta$ -strands are represented by arrows and helices as coils. The regions 1–117, lid (118–159) in red, region 160–213, sub-domain facing the lid (214–261) and region 262–320 are labeled. The position of a  $\text{Ca}^{2+}$  is indicated by a sphere. The catalytic triad (S87, D264 and H286) and amino acid residues involved in oxyanion hole stabilization (L17 and Q88) are shown in stick.



**Figure 2.** Profile of cell density and activity obtained for wild-type and variants of BCL expressed in *E. coli* JM109. Activity and  $OD_{600}/nm$  were an average of values obtained in eight wells.



**Scheme 1.** Hydrolysis of A) *para*-nitrophenyl butyrate (pNPB) and B) (*R,S*)-2-chloro ethyl 2-bromophenylacetate (**1**).

### Enantioselectivity of BCL single mutants towards (*R,S*)-2-chloro ethyl 2-bromophenylacetate (**1**)

The library of single mutants produced in 96-deep well plates was assayed in the presence of racemate **1** (Scheme 1B). Reactions to determine enantioselectivity were performed in deep wells by using crude extracts. The enantiomeric excess of the substrate ( $ee_s$ ) was plotted as a function of the conversion for all BCL variants (Figure 3). As shown in the figure, the single mutations at positions 266 and 287 had limited effect on enzyme enantioselectivity. In one single case, a variant was found to reverse its enantioselectivity from the *R* to the *S* sub-

strate. Indeed, when changing V266 to glycine, which is the most compact amino acid and is also known to contribute greatly to conformational flexibility of polypeptide chains, *S* selectivity was observed. Conversely, three L17 variants (namely L17S, L17G and L17M) showed an enhanced enantioselectivity for the *R* enantiomer (Figure 3B). Variants L17S, L17G and L17M were produced from cultures in Erlenmeyer flasks to validate the microtiter-plate assay. They exhibited *E* values of 87, 24 and 78, respectively; this shows that microtiter-plate assay was a reliable screen for enantioselectivity.

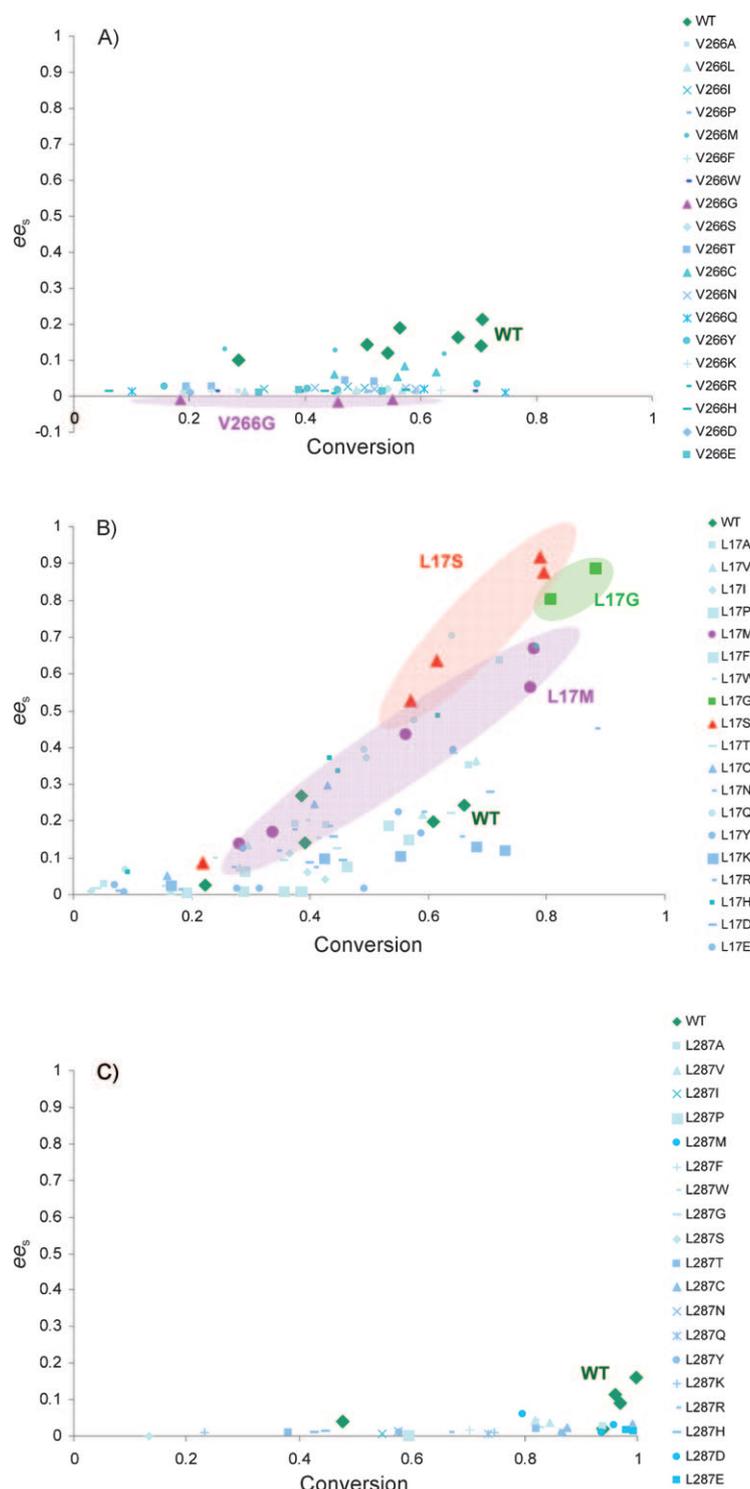
### Characterization of V266G, L17S and L17M variants

The most enantioselective variants (L17S and L17M) and the variant with reversed enantioselectivity (V266G) were then characterized in more detail. Following production and immuno-affinity purification, their ability to hydrolyse substrate **1** was determined. For the V266G mutant, the specific activity remained unchanged compared to the wild-type enzyme ( $v_i = 0.38 \text{ mU mg}^{-1}$ ), but the reversal of enantioselectivity was confirmed by assay of the purified enzyme. The V266G variant was indeed found to hydrolyse the *S* enantiomer significantly faster than the *R* one, with an *E* value of 20 (Table 1). Compared to the parental wild-type enzyme, variants L17S and L17M

**Table 1.** Specific hydrolytic activity and Enantioselectivity values determined for the wild-type BCL and its three single mutants towards each enantiomer of the racemic substrate (*R,S*)-2-chloroethyl-2-bromophenylacetate (**1**).

Variants	$v_{iR}^{[a]}$ [ $\text{mU mg}^{-1}$ ]	$v_{iS}^{[a]}$ [ $\text{mU mg}^{-1}$ ]	$v_i^{[a]}$ [ $\text{mU mg}^{-1}$ ]	Enantioselectivity	<i>E</i> value	Conversion [%]
Wild-type	0.354	0.028	0.38	<i>R</i>	13 ( $\pm 1.8$ ) <sup>[b]</sup>	6.5 (48 h)
V266G	0.02	0.36	0.38	<i>S</i>	20 ( $\pm 4$ ) <sup>[c]</sup>	6.6 (51 h)
L17S	1.57	0.01	1.58	<i>R</i>	128 ( $\pm 35$ ) <sup>[b]</sup>	15.6 (49 h)
L17M	2.07	0.014	2.09	<i>R</i>	133 ( $\pm 31$ ) <sup>[b]</sup>	15.5 (48 h)

[a]  $v_{iR}$ ,  $v_{iS}$ : initial rates;  $v_i = v_{iR} + v_{iS}$ . [b] *E* value =  $v_{iR}/v_{iS}$ . [c] *E* value =  $v_{iS}/v_{iR}$



**Figure 3.** Evolution of substrate enantiomeric excess ( $ee_s$ ) as a function of the conversion for single variants on positions V266 (A), L17 (B) and L287 (C). Enantiomeric excess was calculated as defined below:  $ee_s = \frac{[S] - [R]}{[S] + [R]}$ , ( $s$  = substrate) and the conversion:  $C = 1 - \frac{[(R+S)_t / (R+S)_{t=0}]^d}{d}$ . The results for the best variants are highlighted.

showed an increase of up to sixfold in the initial rate of  $R$  enantiomer consumption ( $v_{ir}$ ). In parallel, the initial rate of  $S$  enantiomer consumption ( $v_{is}$ ) decreased by up to threefold (Table 1). This resulted in a four- and fivefold improvement of the specific activity of the L17S and L17M variants, respectively,

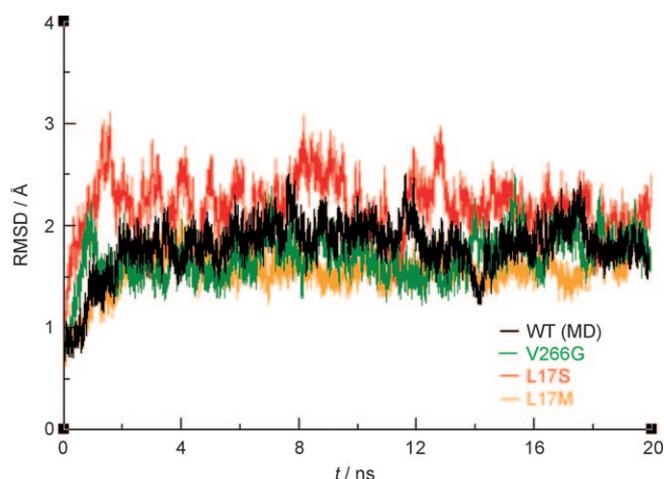
compared to the wild-type enzyme. In similar way, the enantioselectivity of these mutants was increased by a remarkable tenfold respect to wild-type BCL.

### Insights on structural features that control activity and selectivity of BCL single mutants through molecular modelling

The dynamics of the improved single mutants were first investigated to ensure that the mutations did not induce major modifications in the flexibility of the enzyme structure nor in the lid conformation of BCL variants.<sup>[26]</sup> Dynamics studies were then completed by i) the analysis of the binding interactions of each enantiomer in the active site of the variants and ii) the analysis of substrate trajectories during the transport out of the active site.

### Dynamics of wild-type BCL and its variants at water/octane interface

The starting model of wild-type BCL was built from the X-ray structure of BCL in open conformation<sup>[27]</sup> (PDB ID: 3LIP), which adopts the characteristic  $\alpha/\beta$ -hydrolase fold<sup>[28]</sup> (Figure 1B). The model was then subjected to a 20 ns molecular dynamics (MD) simulation under constant temperature and pressure conditions in an explicit water/octane interface, in which the enzyme is known to be active, following a procedure previously described.<sup>[26]</sup> In a similar way, we performed MD simulations on L17S, L17M and V266G mutants by using 3D-models generated from the wild-type BCL model. To verify the stability of the MD trajectories, we inspected the potential energy of the system and the RMSD throughout the simulations. The RMSD trajectories calculated for the backbone atoms of the proteins were compared to the initial structures over the 20 ns of simulations. All profiles show a plateau with an average RMSD around 2 Å, reached after 2 ns; this indicates that the systems have reasonably converged to stable states (Figure 4). To detect the regional motions in the protein structures, B-factors were calculated from MD simulations (Figure 5A). Main differences between the wild-type BCL and its variants reside in two regions, mostly constituted by hydrophobic amino acid residues, surrounding the catalytic pocket. The first region is defined by the lid (residues 118–159) while the second (residues 221–246) corresponds to a sub-domain facing the lid (Figure 1B). Simulated B-factors revealed a particularly high mobility of the hydrophobic  $\beta$ -hairpin constituted by the b3 and b4 strands for all variants as well as for the wild-type BCL. Of note, the mobility of the loop between the  $\alpha 5$  and  $\alpha 6$  helices was found to be more pronounced for the V266G variant than for the wild-type BCL (Figure 5A).



**Figure 4.** Time variation of the RMSD of backbone atoms of the proteins (wild-type (black), V266G (green), L17S (red) and L17M (orange) variants) during the course of MD simulations carried out in explicit water/octane solvent.

Comparison of initial structures with those obtained after 20 ns MD simulation shows that wild-type BCL and its mutants stay in the wide open conformation observed in the crystal (Figure 5B–E). Overall, no major effect of the mutation on the flexibility of the protein backbone was observed. The two variants exhibited similar conformational rearrangements along the simulation and residues involved in the catalytic machinery located at the bottom of a narrow and deep pocket (catalytic triad: S87, H286, D264 and oxanion hole stabilisation: Q88 and L17) did not undergo any significant variation. These results suggest that the amino acid substitutions introduced in BCL did not perturb the protein folding.

#### Covalent docking of substrate in active sites of wild-type BCL, L17S, L17M and V266G variants

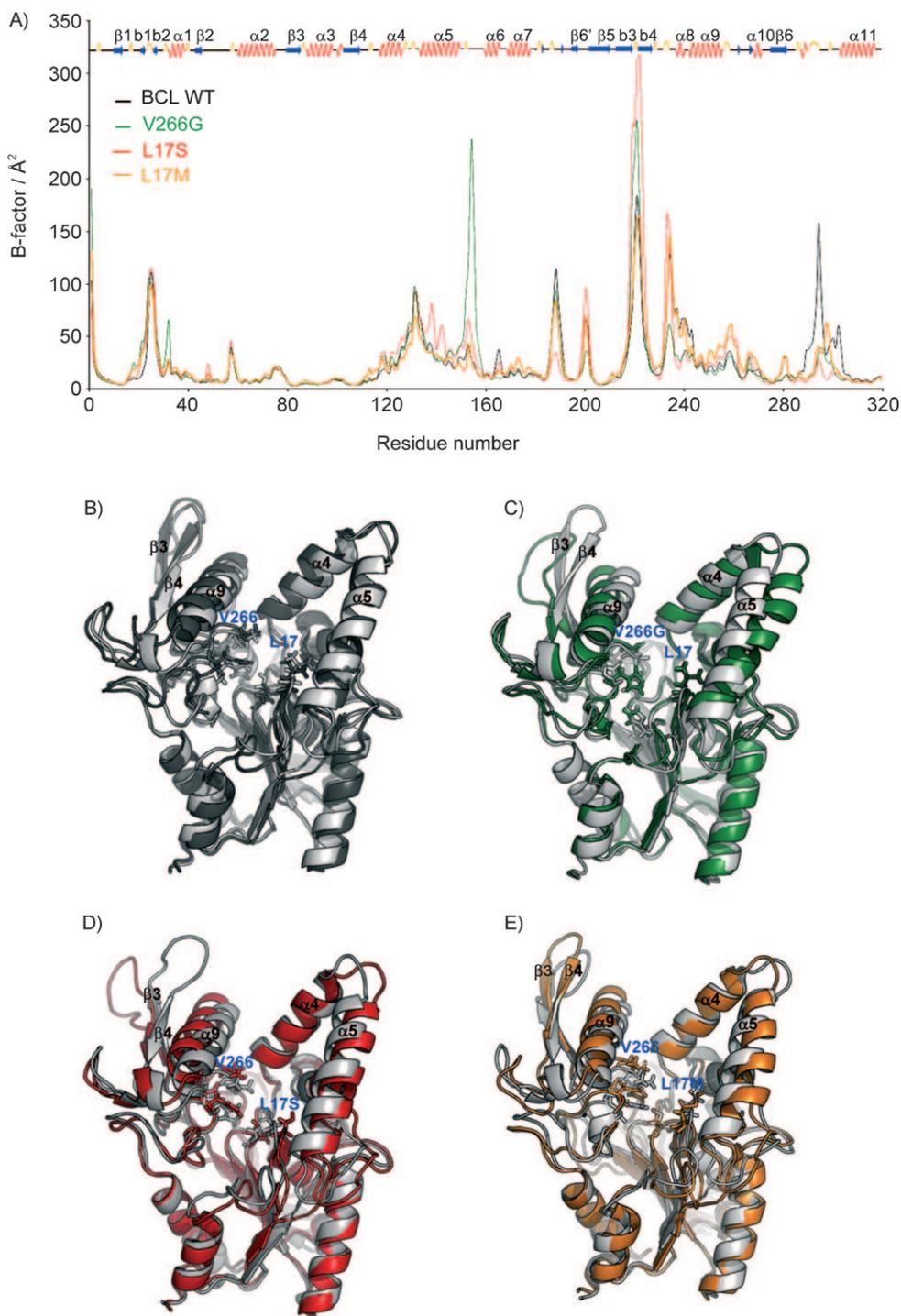
Using a covalent docking procedure, we generated models of the substrate covalently bound to each enzyme; these models corresponded to the tetrahedral intermediate known to be a good mimic of the transition state. In wild-type BCL, L17S and L17M variants, the bulky bromine atom of the rapidly transformed *R* enantiomer is positioned in the inner hydrophobic pocket, the so-called acyl pocket HA,<sup>[12]</sup> pointing towards the inner L167 with the aromatic ring pointing towards the outside of the catalytic pocket (Figure 6A). In contrast, the least rapidly hydrolysed *S* enantiomer was also docked with its bromine atom in the HA pocket, but oriented towards the bottleneck-forming V266. Interestingly, mutation L17S also introduced an additional hydrogen bonding interaction with the covalent intermediate; this led to greater stabilization of the transition states for both *R* and *S* enantiomers compared to the wild-type BCL (Figure 6C). These results are in agreement with a greater hydrolytic activity of L17S variant compared to the wild-type BCL towards substrate **1** (Table 1). However, the difference between scores of the *S* and *R* enantiomers in L17S is lower than that observed for the wild-type enzyme; thus, this

fails to explain the tenfold increase of enantioselectivity observed for the variant in favour of the *R* enantiomer. The docking scores obtained respectively for the *R* and the *S* enantiomers in the L17M variant are in agreement with the enantio-preference of the enzyme. The V266G mutation uncluttered the active site and thus favoured the docking of the *S* enantiomer of substrate **1** in the catalytic site, as reflected by the more favourable docking score (Figure 6B). Figure 6 shows that the orientation of the aromatic ring in the docked poses within all enzymes varies between the *R* and the *S* enantiomers, except for the V266G variant. In the latter, the aromatic ring fits nicely in the space left empty by the V266G mutation; this allows stacking of the aromatic ring onto the catalytic H286. Close values of docking scores were obtained for both enantiomers, that is,  $-1.03$  and  $-1.20$  for the *R* and *S* enantiomers, respectively. The favourable docking score for the *S* enantiomer is in agreement with the experimentally observed reverse of enantioselectivity exhibited by the V266G variant.

Overall, the docking scores (Figure 6) are in agreement with the enantioselectivity, but they failed to predict quantitatively the enantioselectivity, in particular the large improvement of L17S and L17M enantioselectivity. Molecular docking was performed on rigid enzyme 3D-models and solely, the substrate was allowed to adapt its conformation to the active site. Given that molecular adaptation and flexibility are important components for an accurate description of specific ligand binding interactions, this is obviously a limitation.

#### Computing substrate pathways to BCL single mutants catalytic site

Dynamic simulations indicate that the amino acid substitutions introduced in *B. cepacia* lipase did not perturb the protein folding. However, they contribute to significant changes in the enantiomer accessible volume and may have an impact on substrate accessibility. The geometrically feasible motions of the substrates within the enzyme active site were explored using the path-planner integrated in the Biomove3D software.<sup>[16,24]</sup> The algorithm used in this work is derived from the previously described disassembly path-planning technique of articulated objects.<sup>[11,24,29]</sup> In the new version of the Biomove3D software, the basic RRT algorithm was replaced by ML-RRT (Manhattan-Like Rapidly-exploring Random Tree) algorithm.<sup>[24]</sup> ML-RRT outperforms the basic RRT algorithm by considerably decreasing the computing time and its variance. Furthermore, it is able to efficiently treat protein models with all side chains being potentially flexible.<sup>[24]</sup> Within a few minutes on a standard monoprocessor PC, the algorithm is able to compute ligand pathways while taking into account the entire flexibility of both the ligand and the protein side chains that are required to move during the computation of the pathways. In the latest version of Biomove3D, a new computational tool has been included that facilitates the analysis of the ML-RRT search trees by encoding molecular motions in voxel maps.<sup>[30]</sup> Such representation permits arrangement of the information obtained from the exploration of a high-dimensional space (the conformational space of the molecular model) into a three-dimensional

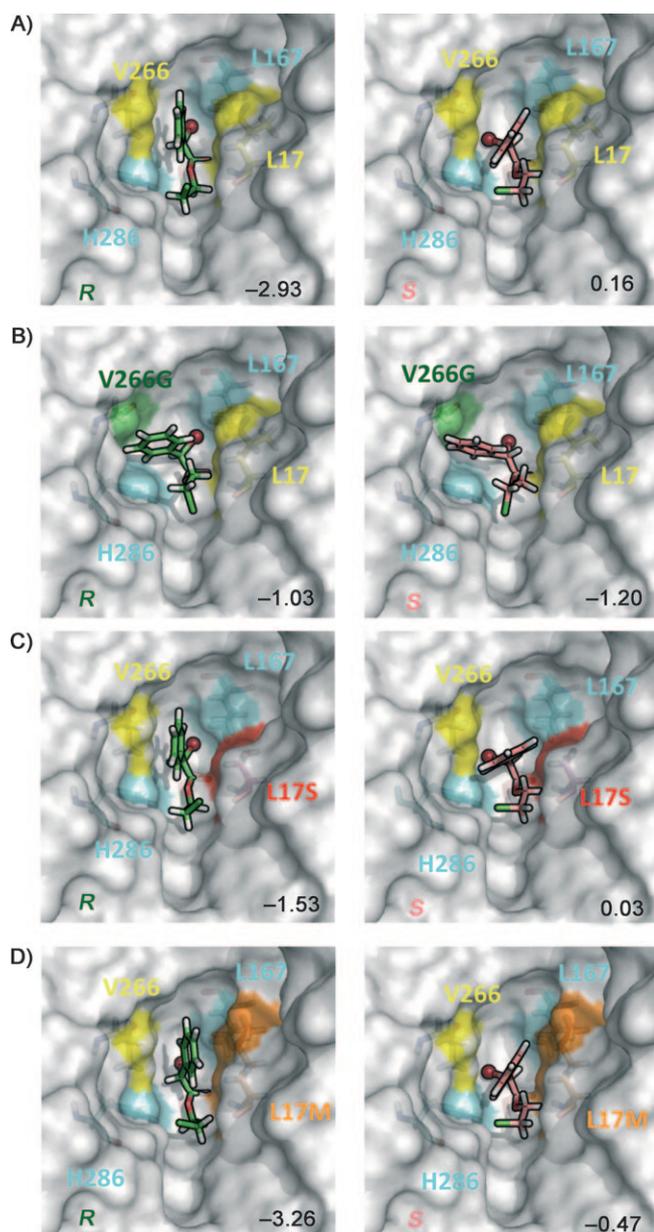


**Figure 5.** A) Calculated B-factors of BCL residues from MD simulations carried out in an explicit water/octane interface for wild type BCL (black), for variant V266G (green), for L17S (red) and for L17M (orange). The secondary structure of BCL is shown on the graph for reference. Snapshots of the enzyme backbone structures taken at the beginning of the production phase (grey coloured) and after 20 ns of MD simulation in water/octane environment: B) wild-type BCL (dark grey), C) V266G (green) D) L17S (red) and E) L17M (orange).

data structure. By combining a geometric conformational exploration and an arrangement of the resultant information into a voxel map, the method permits visualisation of the differences between the space explored by the different enantiomers when accessing the catalytic site of the different BCL enzymes.

duction of the V266G mutation notably facilitates the movement of the *S* enantiomer in the mutant active site compared to the parental enzyme. On the other hand, the V266G mutation exposed a secondary pocket (encircled in Figure 7B) that is unsuccessfully explored by the *R* enantiomer as it searches

Starting from the docked tetrahedral intermediate of the *R* and *S* enantiomer of substrate **1**, we sampled the space that the enantiomers can explore to go from the bottom (catalytic site) of selected BCL variants (L17S, L17M and V266G) to the protein surface. The voxel maps reflect the volume of the catalytic pocket explored by the substrate during its access/exit pathways to the active site. In Figure 7 the computed voxel maps, which represent the positions that can be geometrically reached by the substrate center of mass during the conformational explorations, are shown. As voxels have been coloured following the chronological order of generation, one can easily identify the regions that the substrate reaches first during the geometrical exploration. Comparison of the voxel maps reveals significant differences in the behaviour of racemate **1** into the different enzyme catalytic pockets. Voxel maps obtained for the wild-type BCL indicate a narrower distribution for the *S* enantiomer than the *R* form; this reflects a more constrained motion of the *S* enantiomer as it goes through the bottleneck formed by V266 and L17. Light-blue voxels are found from the bottom to the entrance of the catalytic pocket for the *R* enantiomer; this indicates a faster displacement that requires less iterations than for the *S* enantiomer for which orange voxels are seen around the bottleneck revealing a geometrically difficult passage. These results are in agreement with the faster conversion of the *R* enantiomer by the wild-type BCL (Table 1). As can be seen by the larger voxel map, the intro-



**Figure 6.** Highest score docking modes of (*R,S*)-2-chloro ethyl 2-bromophenylacetate (**1**) covalently bound in the enzyme active site. **1** is docked in the catalytic pocket of wild type enzyme is represented on A) in V266G variant on B), in L17S on C) and in L17M on D). Bromine atom is shown as a sphere. Docking scores are indicated on the figures.

for a way out of the pocket. The exploration of this dead-end region resulted in an increase of the CPU time taken by the path planner to find a solution. Interestingly, upon L17S mutation, the topology of the active site considerably favoured the movement of the *R* enantiomer along the access/exit pathway. Indeed, a network of dark-blue voxels connects the bottom to the top of the catalytic site and reveals the existence of pathways that require less iterations than for the wild-type BCL to exit the catalytic pocket. Of note, the voxel maps computed for the *S* enantiomer in the wild-type and the L17S variant are nearly the same, and this is in agreement with the comparable

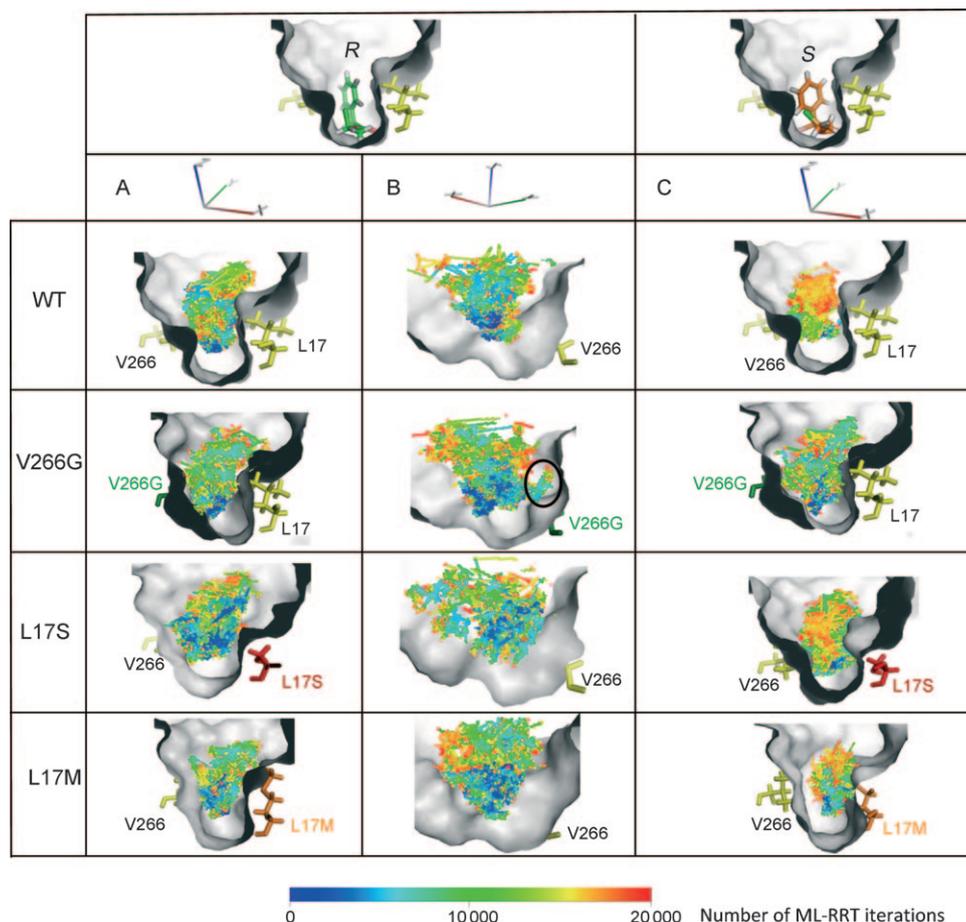
initial rates determined experimentally. Overall, the computation of the exit pathway is greatly facilitated for the *R* enantiomer upon the L17S mutation, and this is in qualitative agreement with the tenfold enhanced enantioselectivity of the L17S variant compared to the wild-type enzyme. The differences observed between the *R* and the *S* voxel maps in the L17M variant reflect the variations in the behaviour of each enantiomer during its displacement along the active site. Indeed, the explored space of the *S* enantiomer is significantly reduced compared to the *R* form; thus, this indicates a greater difficulty of the active site exploration to find an exit pathway. These differences are in agreement with the *E* value of 133 in favour of the *R* substrate.

In the examples presented here, it clearly appears that modelling of the tetrahedral intermediates for both enantiomers was not enough to explain the differences observed in the enantioselectivity of BCL and its variants. Often neglected, the effect of entropy variation has been shown in many cases to be as important as the enthalpic component in enzyme enantioselectivity.<sup>[31,32]</sup> Several studies have attempted to understand the role of entropy on a molecular level and they have estimated the substrate accessible volume within the active site and showed that it could be correlated to transition state entropy.<sup>[31,32]</sup> The modelled volumes could in some cases predict the correct enantiopreference, and the topology of the active site is also proposed to be more suited for the near-attack conformation of the favoured enantiomer.<sup>[33]</sup> In our approach, the full flexibility of both the protein side chains and the substrate is considered for the calculation of the enantiomer pathways. In view of the importance of taking into account the entire molecular motions involved in catalysis when modelling the enzyme reaction, we believe that the computational methods applied in this work might help to expand our knowledge on how enzymes distinguish between enantiomers and to understand the role of entropy on a molecular level.

### Designing improved BCL double mutants

The enantiomers docking, access trajectories and voxel map analyses confirmed the predominant role of the V266/L17 bottleneck on racemate **1** accommodation. To generate diversity at these positions, we first selected the best single mutants L17S and L17M and attempted to introduce an additional mutation at position V266 (Table 2). In the same way, we also attempted the modification of the substrate access to the active site by introducing an additional mutation at position 287 in mutant L17S.

Docking scores, analysis of collision and voxel maps were used to target the mutations. A total of 13 double mutants were constructed and their enantioselectivity was roughly determined without purification of the variants. As shown in Table 2, except for the L17M/V266M mutant, none of the double mutants targeted at positions 17 and 266 led to a notable improvement of the enzyme enantioselectivity. This underlines that engineering of the active site bottleneck is a tricky task, likely due to the proximity of the catalytic triad and oxyanion hole. The behaviour of the L17S/L287 double mutants



**Figure 7.** Voxel maps representing locations of the center of mass of the (*R,S*) enantiomers of (*R,S*)-2-chloro ethyl 2-bromophenylacetate (**1**) reachable from the catalytic position within the wild-type BCL, V266G, L17S and L17M mutants. Voxels resolution is 0.1 Å and colours indicate the chronological order of generation.

turned out to be quite interesting as all of them were found to be more enantioselective than the wild type BCL. Altogether, four mutants were finally purified to homogeneity and characterized. As shown in Table 3, L17M/V266M enantioselectivity was confirmed. The mutant was found slightly more enantioselective than the single L17M mutant and also slightly more active. For the double mutants L17S/L287A; L17S/L287I and L17S/L287W an interesting modulation of the enantioselectivity towards racemate **1** was observed that correlates pretty well with the steric hindrance of the substituent at position 287. In addition, these results show that distant mutations allow a more subtle tune-up of the enzyme structure to enhance its enantioselectivity than mutations located near the catalytic center. With a remarkable *E* value of 178, the L17S/L287I variant was the most enantioselective BCL mutant isolated. Furthermore, this mutant was also found to be 15 times more active than the wild-type enzyme.

**Table 2.** Preliminary Enantioselectivity screening of BCL single and double variants towards the racemic substrate (*R,S*)-2-chloro ethyl 2-bromophenylacetate (**1**). *E* values were determined on nonpurified enzyme extracts. All variants showed an enantiopreference for the *R* enantiomer.

Variants	<i>E</i> value <sup>[a]</sup>
Wild-type	12
<i>Bottleneck: L17S–V266 combinations</i>	
Single mutant: L17S	87
L17S/V266I	19
L17S/V266M	15
L17S/V266Q	1
L17S/V266F	18
L17S/V266D	2
L17S/V266W	1
L17S/V266L	4
L17S/V266T	10
<i>Bottleneck: L17M–V266 combinations</i>	
Single mutant: L17M	78
L17M/V266F	9
L17M/V266M	122
<i>L17S–L287 combinations</i>	
L17S/L287A	39
L17S/L287I	45
L17S/L287W	24

[a]  $E$  value =  $\frac{\ln[(1-C)(1-ee_r)]}{\ln[(1-C)(1+ee_r)]}$

## Conclusions

Overall, our study shows that the catalytic properties of *B. cepacia* lipase can be exploited “à la carte” for the kinetic resolution of 2-substituted racemic acids by structural perturbations via amino acid substitution. The screening of a small library of only 57 monomutants on two structurally different substrates clearly revealed the importance of designing screening assays specifically adapted to the substrate of interest. Another achievement is that the screening of this library enabled the fast isolation of several lipases with a remarkable ten times enhanced or reversed enantioselectivity for the resolution of (*R,S*)-2-chloro ethyl 2-bromophenylacetate.

The voxel map analysis revealed differences that could be linked to accessible volumes and by extension to entropy variations. However, to investigate further the role of both enthalpy and entropy contributions on the enantioselectivity of the BCL variants, we will need to determine experimentally these thermodynamic parameters.

Finally, the semirational engineering strategy developed, which combines the use of classical and novel molecular modelling tools, enabled us to select a restricted number of double mutants (only 13) to construct. Out of this library, the best iso-

**Table 3.** Specific hydrolytic activity and enantioselectivity values determined for the wild-type BCL and its six double mutants towards each enantiomer of the racemic substrate (*R,S*)-2-chloro ethyl 2-bromophenylacetate (1). (Error range for the double mutants is within 10 to 20%).

Variants	$v_{iR}^{[a]}$ [mU mg <sup>-1</sup> ]	$v_{iS}^{[a]}$ [mU mg <sup>-1</sup> ]	$v_i^{[a]}$ [mU mg <sup>-1</sup> ]	Enantioselectivity	<i>E</i> value <sup>[b]</sup>	Conversion [%]
Wild-type	0.354	0.028	0.38	<i>R</i>	13	6.5 (48 h)
L17M/V266M	2.81	0.017	2.83	<i>R</i>	166	9 (19 h)
L17S/L287A	1.83	0.081	1.91	<i>R</i>	22.5	15.6 (20 h)
L17S/L287I	5.95	0.033	5.98	<i>R</i>	178	15.5 (20 h)
L17S/L287W	0.55	0.01	0.56	<i>R</i>	55	6 (20 h)

[a]  $v_{iR}$ ,  $v_{iS}$ : initial rates;  $v_i = v_{iR} + v_{iS}$ . [b] *E* value =  $v_{iR}/v_{iS}$ .

lated variant displayed a 15-fold increased activity and a ten-fold enhanced enantioselectivity. Although additional analyses are now necessary to give a comprehensive interpretation of the molecular factors involved in those enhancements, we can still conclude on the basis of the positive hit ratio (15% of enhanced variants out of the screened library) that the semi-rational engineering strategy that we followed was fairly efficient.

## Experimental Section

**Site-directed mutagenesis:** By using the plasmid pFLAG-ATS-Lip-Hp as a template, the QuickChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) was used according to the manufacturer's instruction to introduce 19 other amino acids at positions 266,<sup>[25]</sup> 17 and 287. The primers employed are L17 forw (5'-CTC GTG CAC GGG XXX ACG GGC ACC GAC-3') and L17 rev (5'-GTC GGT GCC CGT XXX CCC GTG CAC GAG-3'), where X corresponds to the 19 other amino acids (L17A forw=GCC; L17A rev=GGC; L17C forw=TGC; L17C rev=GCA; L17D forw=GAC; L17D rev=GCT; L17F forw=TTC; L17F rev=GAA; L17G forw=GGC; L17G rev=GCC; L17H forw=CAC; L17H rev=GTG; L17I forw=ATC; L17I rev=GAT; L17M forw=ATG; L17M rev=CAT; L17N forw=AAC; L17N rev=GTT; L17P forw=CCC; L17P rev=GGG; L17Q forw=CAG; L17Q rev=CTG; L17R forw=CGC; L17R rev=GCG; L17S forw=AGC; L17S rev=GCT; L17T forw=ACC; L17T rev=GGT; L17V forw=GTC; L17V rev=GAC; L17W forw=TGG; L17W rev=CCA; L17Y forw=TAC; L17Y rev=GTA; L17E forw=GAG; L17E rev=CTC; L17K forw=AAG; L17K rev=CTT).

The primers employed for position L287 are L287 forw (5'-CG AGC TAC AAG TGG AAC CAT XXX GAC GAG ATC AAC CAG-3') and L287 rev (5'-CTG GTT GAT CTC GTC XXX ATG GTT CCA CTT GTA GCT CG-3') where XXX correspond to the codon of the other 19 amino acids (L287A forw=GCC, L287A rev=GGC; L287C forw=TGC, L287C rev=GCA; L287D forw=GAC, L287D rev=GTC; L287E forw=GAG, L287E rev=CTC; L287F forw=TTC, L287F rev=GAA; L287G forw=GGC, L287G rev=GCC; L287H forw=CAC, L287H rev=GTG; L287I forw=ATC, L287I rev=GAT; L287K forw=AAG, L287K rev=CTT; L287M forw=ATG, L287M rev=CAT; L287N forw=AAC, L287N rev=GTT; L287P forw=CCC, L287P rev=GGG; L287Q forw=CAG, L287Q rev=CTG; L287R forw=CGC, L287R rev=GCG; L287S forw=AGC, L287S rev=GCT; L287T forw=ACC, L287T rev=GGT; L287V forw=GTC, L287V rev=GAC; L287W forw=TGG, L287W rev=CCA; L287Y forw=TAC, L287Y rev=GTA).

Double mutants were constructed by following the same procedure used for single mutants construction. Mutant plasmids were used as template for site directed mutagenesis (QuickChange Site-

Directed Mutagenesis Kit). Primers employed are the same as described above.

Mutant plasmids were transformed into *E. coli* JM109 (Promega, Madison, WI). Clones were stored in glycerol (15% v/v) in cryotube vials (Nunc™Brand Products, Denmark) at -20 °C and they were used to inoculate culture in 96-deep-well microplate (Nunc™Brand Products).

**Production of BCL variant in microplate scale:** Starter cultures in sterile 96-well microplates filled with 2×YT medium (150 μL) were inoculated using clones of wild-type or BCL variants (10 μL) stored in cryotube. After 24 h of growth at 30 °C under horizontal shaking (250 rpm), starter cultures (100 μL) were used to inoculate sterile 96-deep-wells culture for protein production (ABgene, Cambridge, UK; 1.1 mL 2×YT; isopropyl-β-D-thiogalactopyranoside (IPTG) at 0.01 mM supplied by Euromedex (Mundolsheim, France)).

These cultures were grown at 30 °C during 24 h under agitation. Growth was measured by OD<sub>600 nm</sub> after shaking to avoid settlement and the cells were then centrifuged at 3700 rpm for 30 min at 4 °C. The pellets were resuspended in lysis buffer (200 μL; 100 mM Tris-HCl pH 7.5, protease inhibitor cocktail tablet EDTA free (Roche Diagnostic), 1 mg mL<sup>-1</sup> lysozyme (Euromedex), 5 μg mL<sup>-1</sup> DNase I (Euromedex) and 1× BugBuster (Novagen), sonicated in a sonicating bath (four cycles; one cycle: 2 min of sonication and 2 min on ice), and frozen at -80 °C over night. After thawing, crude extract were used to measure lipase activity and for screening procedures.

**Determination of hydrolytic activity on pNPB of BCL variants produced in microplate:** Clones were screened for their hydrolytic activity by using an integrated robotic TECAN Genesis RSP-200 platform on pNPB supplied by Sigma-Aldrich as substrate. The lipase extract (20 μL) diluted (v/v) in Tris-HCl buffer (100 mM, pH 7.5) were mixed with Tris-HCl buffer (175 μL; 100 mM, pH 7.5) and pNPB (5 μL; 40 mM in solution in 2-methylbutan-2-ol). pNPB consumption was measured at 405 nm on a VersaMax tunable microplate reader (Molecular Devices, Sunnyvale, CA, USA) at 30 °C. One unit was defined as the amount of enzyme that released 1 μmol of *p*-nitrophenol per minute under these conditions. Activity measurements were performed once for each well. BCL is almost exclusively produced in the form of inclusion bodies and only a small part of the enzyme is in the soluble fraction. The quantity of enzyme in soluble fraction varies between productions and the hydrolytic activity measured can thus be different. To overcome this problem, wild-type enzyme was produced in each deep-well plate and the hydrolytic activity of single variants was always compared to the wild-type enzyme of the same plate.

The activity values obtained from deep-well cultures and Erlenmeyer flask cultures were compared. A positive correlation was found

between the hydrolytic activities produced either by Erlenmeyer or deep-well cultures (data not shown), in agreement with that previously reported for V266 variants.<sup>[25]</sup>

#### Determination of BCL variants' enantioselectivity towards hydrolysis of (*R,S*)-2-chloro ethyl 2-bromophenylacetate

**Classical test:** When BCL variant were produced in Erlenmeyer flasks using 2x YT medium (tryptone 16 g L<sup>-1</sup>, yeast extract 10 g L<sup>-1</sup>, NaCl 5 g L<sup>-1</sup>) supplemented with ampicillin (100 µg mL<sup>-1</sup>) by using conditions previously reported,<sup>[25]</sup> a classical enantioselectivity test was used. Crude extract (500 µL) containing recombinant BCL was added to racemic substrate (250 µL; 50 mM in *n*-octane). The hydrolysis reaction was carried out at 30 °C under magnetic stirring. The progress of the reaction was followed by sampling the reaction at regular intervals.

**Miniaturized test:** For this test, hydrolysis reaction was carried out in 96-deep-well plates used for the enzyme production. After lipase hydrolytic activity measurement, (*R,S*)-2-chloro ethyl 2-bromophenylacetate (100 µL; 50 mM in *n*-octane) were added into each well containing enzyme extract (180 µL). The deep-well plate was sealed with a thermosealing system (Thermofischer Scientific, ALPS 50 V) and the reactions were shaken at 800 rpm and 30 °C in Infors Microtron. The reaction was stopped by addition of *n*-hexane/IPA (500 µL; 70:30 v/v) and the reaction medium was mixed, centrifuged at 13000 rpm during 5 min. The organic extract was analysed by chiral liquid chromatography.

**HPLC analysis:** The chiral HPLC instrument was equipped with a chiral column (Chiralcel OJ-H (25 cm × 4.6 mm), Daicel Chemical Industries Ltd, Osaka, Japan) connected to a UV detector (at 254 nm). The following conditions were used: *n*-hexane/isopropanol 70:30, v/v; flow rate of 1.0 mL min<sup>-1</sup> at 40 °C.

**Determination of conversions, substrate enantiomeric excess (ees) and enantioselectivities (E values):** The conversion was calculated from HPLC results by using the following relationship:  $C = 1 - [(R+S)_t / (R+S)_{t_0}] \times 100$ . The substrate enantiomeric excess is calculated as:  $ee_s = \{[R] - [S]\}_t / \{[R] + [S]\}_t$  (*s* = substrate). Enantioselectivity value was determined as the ratio of the initial rate of *R* enantiomer production versus the initial rate of *S* enantiomer production:  $E \text{ value} = (v_{iR} / v_{iS})$ . Initial rates were determined, before 10% of substrate conversion, by linear regression over at least five points. For reaction on nonpurified enzyme, *E* value was calculated using the conversion, the substrate enantiomeric excess and the mathematical expression [Eq. (1)]:

$$E \text{ value} = \frac{\ln[(1-C)(1-ee_s)]}{\ln[(1-C)(1+ee_s)]} \quad (1)$$

**Purification and protein assay:** Enzyme was expressed in *E. coli* JM109 in fusion with a FLAG peptide. This peptide was used then to purify the enzyme by immuno-affinity purification by using ANTI-FLAG M2 affinity gel (Sigma). Purification was made in batch. One volume of enzyme extract interacts with one volume of affinity gel during 1 h at 4 °C. One volume of gel was then washed by 36 volumes of TBS (50 mM TRIS, 150 mM NaCl, pH 7.4). Elution was made by interaction between gel and FLAG-peptide (300 µg mL<sup>-1</sup>, Sigma) dissolved in TBS with Triton X-100 (1%, Euromedex) and Tween 20 (1%, Sigma-Aldrich). One volume of gel interacts with one volume of elution solution over 5 min. This step was performed three times. The assay of purified protein was made with a NanoDrop 1000 with 2 µL of solution.

**General procedure for the preparation of (*R,S*)-2-chloroethyl 2-bromophenylacetate (1):** Compound 1 was synthesised according

to the procedure previously described,<sup>[34]</sup> but with 2-chloroethanol as the alcohol.

**Spectroscopic data:** Infrared spectra were recorded on a Perkin-Elmer, 1310 infrared spectrophotometer. <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on a Bruker AC-200.1 (<sup>1</sup>H 200.1 MHz and <sup>13</sup>C 50.3 MHz) spectrometer.

**(*R,S*)-2-chloro ethyl 2-bromophenylacetate (1):** Yield: 53%; <sup>1</sup>H NMR (CDCl<sub>3</sub>),  $\delta$  = 3.65–3.71 (t, *J* = 5.7 Hz, 3H; -CH<sub>2</sub>-Cl), 4.35–4.45 (td, *J* = 2.7, 5.7 Hz, 2H; -OCH<sub>2</sub>CH<sub>2</sub>Cl), 5.40 (s, 1H; -CHBr), 7.33–7.59 (m, 5H; ArH); <sup>13</sup>C NMR (CDCl<sub>3</sub>):  $\delta$  = 41.12 (-CH<sub>2</sub>Cl), 46.31 (-CHBr), 65.75 (-OCH<sub>2</sub>-), 128.79(×2), 128.97(×2), 129.52, 135.43, 168.09 (COO); IR: 1750 and 1730 (ν<sub>C=O</sub>), 1600 and 1475 (ν<sub>C-C</sub>), 1280–1140 cm<sup>-1</sup> (ν<sub>C-O</sub>); elemental analysis calcd (%) for C<sub>10</sub>H<sub>10</sub>O<sub>2</sub>BrCl: C 43.49, H 3.65; found: C 43.13, H 3.46.

## Computational methods

**MD calculations:** All MD simulations were carried out by using the AMBER 9 suite of programs<sup>[35]</sup> and the all-atom ff03 forcefield.<sup>[36,37]</sup> The starting model of BCL was derived from the high-resolution crystal structure of BCL in an open conformation, available from the Protein Data Bank (PDB ID: 3LIP).<sup>[27]</sup> The calcium ion that plays a structural role in BCL was conserved in the model and appropriately parameterized according to prior work.<sup>[38]</sup> Four Na<sup>+</sup> cations were added to neutralize the protein. Models of BCL mutants were constructed from wild-type BCL X-ray structure using the Biopolymer module of Sybyl7.3 (Tripos). MD simulations were performed in explicit solvent at water/octane interface using a procedure previously described.<sup>[26]</sup> The simulations were carried out for a total of 20 ns at constant temperature and pressure conditions, using the Berendsen algorithm with a coupling constant of 2 ps for both parameters. Electrostatic interactions were calculated by using the Particle-Mesh Ewald method<sup>[39]</sup> with a nonbonded cutoff of 10 Å. All bonds involving hydrogen atoms were constrained with the SHAKE algorithm,<sup>[40]</sup> permitting the use of 2 fs time steps to integrate the equations of motion. The centre of mass of translational and rotational motion was removed every 6000 MD steps to avoid methodological problems described by Harvey et al.<sup>[41]</sup> The trajectories were extended, as noted above, to 20 ns, and conformations of the system were saved every 0.4 ps for further analysis. Trajectories analysis was carried out using the Ptraj module of the AMBER 9 package.<sup>[35]</sup> The root mean square deviation (RMSD) was calculated for the protein backbone atoms using least squares fit. Atomic positional fluctuations (Δ*r*<sub>*i*</sub><sup>2</sup>) of backbone were calculated. A mass-weighted average value was then calculated for each residue. These parameters are related to the B-factors through the following relationship [Eq. (2)]:

$$B_i = \frac{8\pi^2}{3} \langle \Delta r_i^2 \rangle \quad (2)$$

The simulated *B* factors were calculated by using the coordinates of the 20 ns trajectories.

**Covalent docking of substrates:** Starting coordinates were extracted from the crystal structure of *B. cepacia* lipase (PDB ID: 3LIP) to generate models of wild-type BCL and its variants. Models of BCL mutants were constructed by replacing in silico the target amino acid (L17 or V266) with either L17S, L17M or V266G mutation using the Biopolymer module of Insight II software package (Accelrys, San Diego, USA). The conformation of the mutated residue side chain was optimized by manually selecting a low-energy conformation from a side-chain rotamer library. Steric clashes (van

der Waals overlap) and nonbonded interaction energies (Coulombic and Lennard-Jones) were evaluated for the different side-chain conformations.

The tetrahedral intermediates of both the *R* and the *S* form of the (*R,S*)-2-chloroethyl 2-bromophenylacetate (**1**) were generated using Insight II. Energy minimization of each substrate was performed to generate low-energy starting conformation with suitable bond distances and angles. Starting from these initial structures, the conformational space accessible by the substrate covalently bound to the catalytic serine (Ser 87) in a tetrahedral form was investigated using the automated flexible docking program FlexX<sup>[42,43]</sup> (Biosolve-it). All parameters were set to the standard values as implemented in Version 3.1.1. FlexX uses a fast algorithm for the flexible docking of small ligands into fixed protein binding sites using an incremental construction process of the ligand.<sup>[43]</sup> The docking region was defined to encompass all protein amino acids for which at least one heavy atom was located within a 6.5 Å radius sphere, whose origin was located at the centre of mass of catalytic Ser87. All crystallographic bound waters were removed prior to docking. Hydrogen atoms were added to the protein using standard Sybyl (Tripos) geometries. Residues were kept fixed in their crystallographic positions in all docking experiments. For each docking, the top 30 solutions corresponding to the best FlexX scores were retained.

**Calculation of enantiomer trajectories:** To allow the displacement of the substrates along the active site path, the covalent bond between the catalytic serine and the carbon of the substrate carbonyl function was broken to create two separate molecular entities. The hybridization of the carbonyl function was corrected from  $sp^3$  to  $sp^2$  to generate the molecular models used to search for trajectories. Starting from the docked position, enantiomer trajectories were computed using Biomove3D integrated path planner,<sup>[16,24]</sup> going from the bottom towards the entrance of the active site.

Within BioMove3D, both the lipase and the substrate are modelled as polyarticulated mechanisms, and their motion is restricted by geometric constraints such as steric clash avoidance between spherical atoms with (a percentage of) van der Waals radii. The substrate is considered entirely flexible and all protein side chains that are required to move during the computation of the pathway can do it freely. The ML-RRT (Manhattan-Like Rapidly-exploring Random Tree) algorithm<sup>[24]</sup> is used to explore the conformational space. This algorithm considers two sets of parameters—*active* and *passive*—that are treated in a decoupled manner. *Active* parameters consider the location and the internal torsions of the ligand and they are treated at each iteration of the algorithm. *Passive* parameters take into account the torsion angles of the protein side chains and they only need to be treated when such side chains hinder the motion of *active* parts or other *passive* parts. Therefore, the algorithm favours the ligand motion and determines the protein side chains that must move to allow the ligand progression in the catalytic pocket of the protein. Of note, although the *active* and *passive* parts move alternately in the path computed by the ML-RRT algorithm, a randomized path smoothing post-processing is performed in the composite configuration space of all parameters, so that simultaneous motions are obtained in the final path. For each enzyme–enantiomer pair, the search tree generated by the ML-RRT algorithm was embedded in a three-dimensional data structure called voxel map. The procedure used to construct the voxel maps is identical to that described in our earlier report.<sup>[30]</sup> Visualization and graphics were done using VMD<sup>[44]</sup> and PyMol<sup>[45]</sup> softwares.

**Equipment:** Biomove3D calculations, molecular constructions and graphic displays were performed on an Intel Pentium4 PC with a 3.2 GHz processor. MD calculations were carried out at the Computing Centre of the Région Midi-Pyrénées (CALMIP, Toulouse, France) and on the Linux cluster available at the Centre for Computing Resources (CRI) of INSA–Toulouse, France.

## Acknowledgements

The authors wish to thank the Computing Centre of the Région Midi-Pyrénées (CALMIP, Toulouse, France) and the Centre for Computing Resources (CRI) of INSA–Toulouse for providing calculation resources and support. This work was supported by the Institut des Technologies Avancées du Vivant (ITAV), Toulouse Canceropole campus.

**Keywords:** chiral resolution · enantioselectivity · enzymes · lipases · molecular modeling

- [1] M. J. Kim, Y. Ahn, J. Park, *Curr. Opin. Biotechnol.* **2002**, *13*, 578–587.
- [2] M. J. Kim, Y. K. Choi, S. Kim, D. Kim, K. Han, S. B. Ko, J. Park, *Org. Lett.* **2008**, *10*, 1295–1298.
- [3] J. Nyhlén, B. Martin-Matute, A. G. Sandström, M. Bocola, J. E. Bäckvall, *ChemBioChem* **2008**, *9*, 1968–1974.
- [4] O. McConnell, A. Bach II, C. Balibar, N. Byrne, Y. Cai, G. Carter, M. Chlenov, K. Fan, I. Goljer, H. Yanan, D. Herold, M. Kagan, M. Kerns, F. Koehn, C. Kraml, V. Marathias, B. Marquez, L. Mc Donald, L. Nogle, C. Petucci, G. Schlingmann, G. Tawa, M. Tischler, T. Williamson, A. Sutherland, W. Watts, M. Young, M.-Y. Zhang, Y. Zhang, D. Zhou, D. Ho, *Chirality* **2007**, *19*, 658–682.
- [5] F. Bordes, E. Cambon, V. Dossat-Letisse, I. André, C. Croux, J. M. Nicaud, A. Marty, *ChemBioChem* **2009**, *10*, 1705–1713.
- [6] R. Kato, H. Nakano, H. Konishi, K. Kato, Y. Koga, T. Yamane, T. Kobayashi, H. Honda, *J. Mol. Biol.* **2005**, *351*, 683–692.
- [7] Y. Koga, K. Kato, H. Nakano, T. Yamane, *J. Mol. Biol.* **2003**, *331*, 585–592.
- [8] T. Schulz, J. Pleiss, R. D. Schmid, *Protein Sci.* **2000**, *9*, 1053–1062.
- [9] X.-G. Li, M. Lähitie, M. Päiviö, L. T. Kanerva, *Tetrahedron: Asymmetry* **2007**, *18*, 1567–1573.
- [10] K. Dąbkowska, K. W. Szewczyk, *Biochem. Eng. J.* **2009**, *46*, 147–153
- [11] D. Guieysse, J. Cortés, S. Puech-Guenot, S. Barbe, V. Lafaquière, P. Monsan, T. Siméon, I. André, M. Remaud-Siméon, *ChemBioChem* **2008**, *9*, 1308–1317.
- [12] D. Guieysse, C. Salagnad, P. Monsan, M. Remaud-Siméon, V. Tran, *Tetrahedron: Asymmetry* **2003**, *14*, 1807–1817.
- [13] M. S. Apaydin, D. L. Brutlag, C. Guestrin, D. Hsu, J. C. Latombe, C. Varma, *J. Comput. Biol.* **2003**, *10*, 257–281.
- [14] A. P. Singh, J. C. Latombe, D. L. Brutlag, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1999**, 252–261.
- [15] J. Cortés, T. Siméon, M. Remaud-Siméon, V. Tran, *J. Comput. Chem.* **2004**, *25*, 956–967.
- [16] J. Cortés, T. Siméon, V. Ruiz de Angulo, D. Guieysse, M. Remaud-Siméon, V. Tran, *Bioinformatics* **2005**, *21 Suppl 1*, i116–125.
- [17] S. Kirillova, J. Cortés, A. Stefaniu, T. Siméon, *Proteins Struct. Funct. Bioinf.* **2008**, *70*, 131–143.
- [18] X. Tang, B. Kirkpatrick, S. Thomas, G. Song, N. M. Amato, *J. Comput. Biol.* **2005**, *12*, 862–881.
- [19] A. Enosh, B. Raveh, O. Furman-Schueler, D. Halperin, N. Ben-Tal, *Biophys. J.* **2008**, *95*, 3850–3860.
- [20] B. Raveh, A. Enosh, O. Schueler-Furman, D. Halperin, *PLoS Comput. Biol.* **2009**, *5*, e1000295.
- [21] N. M. Amato, K. A. Dill, G. Song, *J. Comput. Biol.* **2003**, *10*, 239–255.
- [22] G. Song, S. Thomas, K. A. Dill, J. M. Scholtz, N. M. Amato, *Pac. Symp. Bio-comput.* **2003**, 240–251.
- [23] V. Ruiz de Angul, J. Cortés, T. Siméon, *Proc. Robotics: Science and Systems* **2005**.

- [24] J. Cortés, L. Jaillet, T. Siméon, *IEEE Int. Conf. Robotics and Automation* **2007**.
- [25] S. Puech-Guenot, V. Lafaquière, D. Guieysse, L. Landric-Burtin, P. Monsan, M. Remaud-Siméon, *J. Biomol. Screening* **2008**, *13*, 72–79.
- [26] S. Barbe, V. Lafaquière, D. Guieysse, P. Monsan, M. Remaud-Siméon, I. André, *Proteins* **2009**, *77*, 509–523.
- [27] J. D. Schrag, Y. Li, M. Cygler, D. Lang, T. Burgdorf, H. J. Hecht, R. Schmid, D. Schomburg, T. J. Rydel, J. D. Oliver, L. C. Strickland, C. M. Dunaway, S. B. Larson, J. Day, A. McPherson, *Structure* **1997**, *5*, 187–202.
- [28] K. K. Kim, H. K. Song, D. H. Shin, K. Y. Hwang, S. W. Suh, *Structure* **1997**, *5*, 173–185.
- [29] S. M. La Valle, J. J. Kuffner, *Algorithmic and Computational Robotics: New Directions* **2001**, 293–308.
- [30] J. Cortés, S. Barbe, M. Erard, T. Siméon, *Proc. IEEE Int. Conf. Robotics and Automation (ICRA09)*, Kobe, Japan, May **2009**.
- [31] J. Ottosson, L. Fransson, K. Hult, *Protein Sci.* **2002**, *11*, 1462–1471.
- [32] J. Ottosson, J. C. Rotticci-Mulder, D. Rotticci, K. Hult, *Protein Sci.* **2001**, *10*, 1769–1774.
- [33] E. Y. Lau, K. Kahn, P. A. Bash, T. C. Bruice, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9937–9942.
- [34] D. Guieysse, C. Salagnad, P. Monsan, M. Remaud-Siméon, *Tetrahedron: Asymmetry* **2003**, *14*, 317–323.
- [35] D. A. Case, T. E. Darden, I. T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, P. A. Kollman, *AMBER 9*, University of California, San Francisco **2006**.
- [36] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- [37] M. C. Lee, Y. Duan, *Proteins Struct. Funct. Bioinf.* **2004**, *55*, 620–634.
- [38] J. Aqvist, *J. Phys. Chem.* **1990**, *94*, 8021–8024.
- [39] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- [40] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–336.
- [41] S. C. Harvey, R. K. Z. Tan, I. T. E. Cheatham, *J. Comput. Chem.* **1998**, *19*, 726–740.
- [42] B. Kramer, M. Rarey, T. Lengauer, *Proteins Struct. Funct. Genetics* **1999**, *37*, 228–241.
- [43] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *J. Mol. Biol.* **1996**, *261/3*, 470–489.
- [44] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.
- [45] W. L. DeLano, The PyMOL Molecular Graphics System on World Wide Web <http://www.pymol.org>, **2002**.

---

Received: July 15, 2009

Published online on October 8, 2009