Encoding molecular motions in voxels maps

Juan Cortés, Sophie Barbe, Monique Erard, Thierry Siméon

Encoding Molecular Motions in Voxel Maps

Juan Cortés, Sophie Barbe, Monique Erard and Thierry Siméon

Abstract— **Understanding life at the atomic level requires the development of new methodologies, able to overcome the limitations of available experimental and computational techniques for the analysis of processes involving molecular motions. With this goal in mind, we develop new methods, combining robotic path planning algorithms and molecular modeling techniques, for computing large-amplitude motions. This paper builds on these new methods, and introduces voxel maps as a computational tool to encode and to represent such motions. Voxel maps can be used to represent relative motions of two molecules, as well as conformational changes in macromolecules. We investigate several applications and show results that illustrate the interest of such representation. In particular, voxel maps are used to display channels into proteins, to analyze protein-ligand specificity, and to represents protein loop and domain motions.**

I. Introduction

Nowadays, the dynamic nature of biological macromolecules, opposed to the static picture provided by X-ray crystallography, is generally accepted. Furthermore, it has been shown that flexibility plays key roles in molecular interactions such as protein-ligand [1], [2] and protein-protein docking [3], [4]. Unfortunately, and despite great advances achieved in the last years [5], [6], an atomic-resolution structural description of slow-timescale (large-amplitude) molecular motions is out of reach for currently available experimental methods [7]. Computational methods are therefore necessary to complement experimentation.

Molecular dynamics (MD) [8], [9] is the most widely used computational method to simulate molecular motions. MD is an appropriate method to analyze motions taking place in a short timescale (up to some nanoseconds). However, it is too computationally expensive for routine simulations of large-amplitude motions of macromolecules. In some cases, MD simulations can be accelerated by the introduction of artificial forces [10]. Nevertheless, devising such forces may require prior knowledge of the particular problem, and they can excessively bias the resulting trajectories. Simulation methods based on Monte Carlo (MC) algorithms [8], [9] have been developed to overcome the limitations of MD. Such methods present however a major drawback for computing large-amplitude motions since the conformational exploration tends to get trapped into the many local minima of the complex molecular energy landscape. Alternatives to MD and MC simulations have been proposed using very different methods such as iterative NMA calculations [11], [12], or structurally constrained conformational exploration using models from rigidity theory [13].

Our method builds on path planning algorithms [14], [15], originally developed in the field of robotics. Such algorithms are efficient tools for exploring constrained high-dimensional spaces. In the recent years, path-planning-based methods have been successfully applied for investigating different problems in structural biology such as: protein-ligand access and docking [16], [17], [18], protein and RNA folding [19], [20], [21], protein loop motions [22], domain motions [23], and motions of pairs of α-helices in transmembrane proteins [24].

This paper recalls our approach for computing molecular motions (Section II-B) and introduces voxel maps as a new and general computational tool to encode them (Section II-C). Section III illustrates the potential interest of the method on several structural biology problems. The presented results show how voxel maps can effectively represent relative motions of two molecules, as well as conformational changes in proteins. In Section III-A, voxel maps are applied to identify channels in proteins, by exploring and encoding possible motions of a single atom between the active site and the surface. The second application (Section III-B) addresses protein-ligand interactions. Voxel maps permit to reflect differences between the access/exit pathways of different ligands to the active site of a protein. Finally, Section III-C deals with the representation of conformational changes due to loop and domain motions.

II. Methodology

A. Overview

The method presented in this paper builds on the two-stage approach proposed in [18] for computing large-amplitude molecular motions. The first and main stage consists in a geometric processing of the strongest molecular constraints (no atom overlaps, no bond breaking). Fast geometric computation [25] combined with efficient path planning algorithms [26] permits our method to generate large-amplitude motions of flexible molecules with very low computational cost. In the second stage, results of the geometric exploration can be refined and analyzed using classic molecular modeling tools (e.g. energy evaluation/minimization).

The voxel-map representation described below can be seen as an intermediate layer between the two stages. It permits to arrange the information obtained from the exploration of a high-dimensional space (the molecular conformational space) into a simple three-dimensional data structure. The

J. Cortés and T. Siméon are with the LAAS-CNRS, F-31077 Toulouse, France {jcortes,nic}@laas.fr

S. Barbe is with the LISBP, F-31077 Toulouse, France sbarbe@insa-toulouse.fr

M. Erard is with the IPBS, F-31077 Toulouse, France monique.erard@ipbs.fr

All the authors are with the Université de Toulouse; UPS, INSA, INP, ISAE; F-31077 Toulouse, France

choice of the three dimensions and the size of voxels depends on the application (see Section III). In addition to the information structuring, voxel maps permit a visual analysis of the results of the conformational exploration.

B. Exploring geometrically feasible motions

The conformational search method applied in this work (described in more detail in [18]) is based on a mechanistic modeling of molecules [27]. Groups of bonded atoms form the bodies of the mechanism, which are linked by articulations corresponding to bond torsions. These torsions are the molecular degrees of freedom. The atoms are represented by rigid spheres with (a percentage of) van der Waals radii. These spheres cannot overlap. Additional distance and orientation constraints can be imposed between elements of this mechanistic model in order to simulate interactions such as hydrogen bonds.

The technique applied to explore feasible motions of the mechanistic molecular model is derived from the *Rapidly-exploring Random Trees* (RRT) algorithm [28]. The basic principle of RRT is to incrementally grow a random tree, rooted at a given initial conformation $\mathbf{q}_{\text{init}}$, to explore the search space for finding feasible paths. The exploration process is illustrated in Figure 1-b. At each iteration, the tree is expanded toward a randomly sampled conformation $\mathbf{q}_{\text{rand}}$. This random sample is used to simultaneously determine the tree node to be expanded and the direction in which it is expanded. Given a distance metric in the search space, the nearest node $\mathbf{q}_{\text{near}}$ in the tree is selected. The expansion of $\mathbf{q}_{\text{near}}$ in the direction of $\mathbf{q}_{\text{rand}}$ leads to the generation of a new node $\mathbf{q}_{\text{new}}$ and a feasible local path. The key feature of the RRT expansion strategy is to bias the exploration toward unexplored regions before uniformly covering the space.

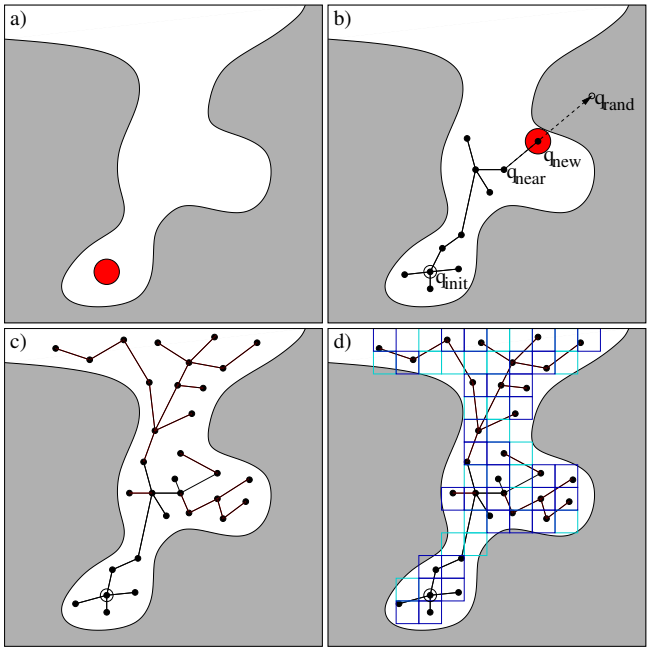

Fig. 1. Two-dimensional illustration of a voxel map construction. Given a geometric model of the molecules (a), a path planning algorithm is used to explore the subset of the conformational space that is reachable from the initial conformation satisfying motion constraints. b) Illustration of an expansion step of an RRT-like path planning algorithm. c) Search tree resulting from the geometric exploration. d) Voxel map associated with the search tree.

The algorithmic variant used in this work is the ML-RRT [26]. This variant considers two sets of conformational parameters: *active* and *passive*. Active parameters are randomly sampled, while passive parameters are treated separately, and only when they hinder the motion of active parts. The main advantage of ML-RRT is a higher efficiency for dealing with high-dimensional problems thanks to the decoupled treatment of parameter subsets, which favors the exploration of active parameters. In the applications presented in this paper, the passive parameters correspond to the torsion angles of the protein side-chains, while the active parameters correspond to the other variables: the location and the internal torsions of the ligand, and the torsion angles of flexible protein backbone segments.

C. Putting search trees into voxel maps

Nodes and edges of the RRT search tree are embedded in a high-dimensional space (the conformational space of the molecular model). The idea is to arrange this information into a lower-dimensional data structure for facilitating further analysis. 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 three-dimensional, which permits a visual display.

The three dimensions of the voxel map correspond to three variables of interest for the particular problem. These three variables may be a subset of the conformational parameters of the mechanical model (e.g. position of the reference frame of a molecule, three selected bond torsions). They can also be chosen to encode information obtained from the conformation (e.g. position of one atom, center of mass of one domain). The voxel size also depends on the application. For instance, the resolution may be chosen depending on the motion amplitude, and on the cost of the ulterior treatment (geometric and/or energetic analysis). Note that increasing the voxel resolution does not affect computational efficiency.

The process for generating the voxel map from the RRT search tree is very simple. It is illustrated in Figure 1-d. Each time a node $\mathbf{q}_{\text{new}}$ is generated, it is added to the corresponding voxel. The procedure is slightly more complex for the edges. When an edge goes across voxels that do not contain the nodes it connects, an intermediate conformation along the edge is kept into each new voxel. Thus, each voxel contains a list of conformations corresponding to the RRT nodes and edges that are projected on it.

Voxels can be labeled in different ways. For instance, values can be assigned depending on the chronological order of generation. In this way, the voxel map can be used to display the regions of the space that are reached first during the conformational exploration (see Section III-B). Other labeling procedures can be devised based on conformational or geometric features, such as the distance between catalytic residues (see the protein loop example in Section III-C.1). The voxel labeling can also be made according to energy

evaluation. An energetic analysis of the conformations associated with voxels may provide very useful information about the conformational energy landscape (see the example of protein domain motions in Section III-C.2).

III. Applications

This section presents several examples illustrating voxel map computations for different types of problems involving molecular motions. The goal is to show the interest of the proposed approach. Therefore, the section does not provide in-depth explanations about the considered systems and the biological importance of the results.

The method has been implemented within our software prototype BioMove3D. PyMOL [29] has been used for viewing molecular models and voxel maps. The computing times given below correspond to tests run on a single AMD Opteron 148 processor at 2.6 GHz.

A. Findings channels into proteins

The most straightforward application of the method is the search and representation of channels into proteins. The channels are searched using the ML-RRT algorithm to explore feasible motions of a ball of arbitrary radius inside the protein model. This algorithm permits to directly treat all side-chain flexibility with a low computational cost. In this type of application, the three variables used for the voxel-map representation are the position parameters of the moving ball. The voxel resolution can be chosen in relation to the ball size.

The chosen benchmark for this application is cytochrome P450. The in/out channels of this enzyme have been recently characterized in [30]. The authors of the cited study applied the technique CAVER [31] for computing channels between the buried active site and the surface in different crystal structures of P450. CAVER is based on the construction of a vertex-weighted graph from a discrete three-dimensional grid model of the protein. The weights are computed from the distance to the protein atoms, the lowest weights corresponding to nodes with the highest clearance. A variant of Dijkstra's algorithm is applied to search the shortest low-cost paths. CAVER considers static structures, and performs systematic exploration of the protein interior. Molecular flexibility can only be indirectly treated by applying the technique to a set of structures (e.g. samples of a molecular dynamics simulation). The results described below show several advantages of the voxel map method over CAVER.

The structure represented in Figure 2 corresponds to bacterial P450-BM3 (PDBid 1JPZ). The voxel map in the figure represents the channels found by our method for a moving ball of radius 1.2 Å. Three of these channels, W, 2b and 2f, were also found by CAVER. However, channel 2d was not reported for this structure. Nevertheless, this channel was observed in a small number of other P450-BM3 structures [30]. A further analysis of channel 2d shows that the exit of the moving ball requires slight motions of some side-chains, in particular those of residues Leu20 and Leu29. This result shows the interest of considering side-chain motions when computing channels in proteins.

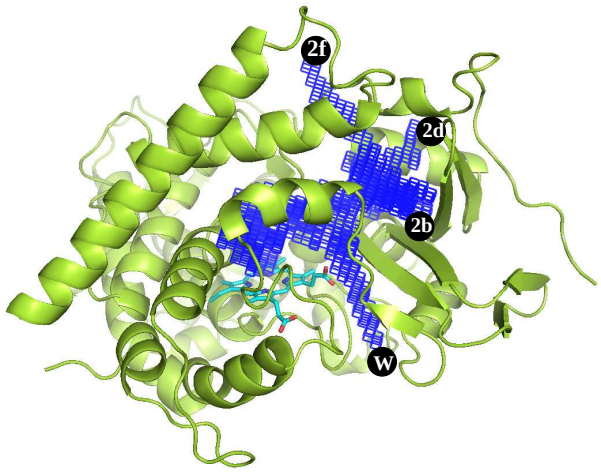

Fig. 2. Voxel-map representation of channels connecting the active site and the surface in bacterial P450-BM3. Channel identifiers follow the nomenclature in related work [30].

Also note that very similar channels were obtained from several runs of the voxel map computation with the moving ball initially located at different positions in the enzyme active site. Such a reliability shows the low sensibility to the starting point, which is another advantage over CAVER[1].

Finally note that, even considering side-chain flexibility, the computing time required to construct the voxel map representation of the channels remains low (only 5 minutes).

B. Analyzing ligand access/exit pathways

The proposed approach can be applied to carry out an analysis of access/exit pathways of ligands (or substrates/products) to the active site of a protein. Such pathways can play important roles in the protein activity and specificity, particularly for proteins presenting a deep binding pocket. For instance, they may be one factor in enzyme enantioselectivity [32].

The example illustrated in Figure 3 corresponds to the interaction between *Candida antarctica* lipase B (CALB) and (*R,S*)-enantiomers of 4-methyloctanoic acid [33]. Our method has been applied to explore possible motions of the *R* and *S* substrates from its catalytic position. The model considers the flexibility of the substrate and all protein side-chains. The voxel maps computed in this case represent the feasible positions reached by the substrate center of mass during the conformational exploration. Voxel resolution is 0.1 Å. This representation shows significant differences between both enantiomers. First, the bottom of the voxel map is narrower for the (*S*)-enantiomer than for the (*R*)-enantiomer, which shows that possible motions of the (*S*)-enantiomer are more constrained in this region. Voxels have been colored

[1]Results reported in [30] indicate that CAVER presents a significant sensibility to the staring point.

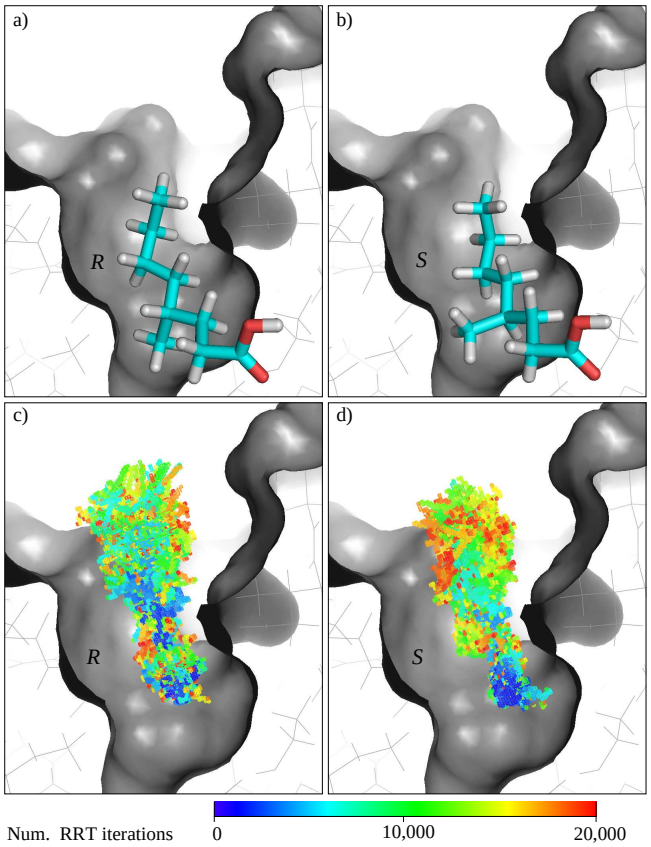

Fig. 3. a)-b) Models of the (*R,S*)-enantiomers of 4-methyloctanoic acid in the active site of CALB [33]. c)-d) Voxel maps representing locations of the center of mass of the (*R,S*)-enantiomers reachable from the catalytic position. Voxels size is 0.1 Å, and colors indicate the chronological order of generation.

depending on the chronological order of generation. Dark-blue voxels correspond to the positions that are reached first. For the (*R*)-enantiomer (Figure 3-c), dark-blue voxels reach the middle-part of the map. However, such voxels are concentrated in the bottom part of the map for the (*S*)-enantiomer (Figure 3-d). This means that, only considering geometric constraints, the access/exit of the (*R*)-enantiomer can be faster. These results tend to indicate that the topology of CALB active site is better suited for facilitating the reaction with (*R*)-4-methyloctanoic acid, which correlates with experimental results on the enzyme enantioselectivity showing a significant preference for this enantiomer [33].

Concerning computational performance, the construction of the voxel map (including the ML-RRT exploration process) took less than 7 minutes for each enantiomer. Note that, aiming to get a very good coverage of the space into the active site cavity, the ML-RRT search tree expansion process yielded voxel maps with more than 25,000 and 15,000 voxels respectively for the (*R*)- and (*S*) enantiomers. This result shows that even such an exhaustive exploration is computationally fast.

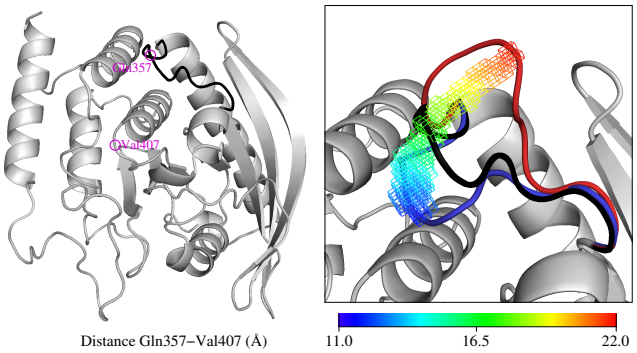

Fig. 4. (Right) Structure of PTPase (PDB ID: 1YPT), with the WDP loop in black color. (Left) Detail of the WDP loop and voxel map displaying the explored locations of the Cα atom of Glu357. Voxel colors indicate the distance between Cα atoms of Glu357 and Val407. The most open and closed loop conformations are represented in red and blue respectively.

C. Representing loop/domain motions

Our path-planning-based approach can be applied to compute large-amplitude internal molecular motions such as loop and domain motions [22], [18], [23]. Integrating the voxel-map representation in this approach provides a new tool for analyzing such deformations.

1) Loop motions: The example illustrated in Figure 4 concerns the "WDP loop" in *Yersinia* protein tyrosine phosphatase (PTPase). The movement of the WDP loop plays a central role in the PTPase-mediated catalytic process [34], [35]. An open conformation of this loop permits the substrate access to the protein active site. Then, the WDP loop is required to adopt a closed conformation that brings the catalytic residue Asp356 to a specific location for protein-substrate interaction. Starting from the open conformation of PTPase [34] (PDB ID: 1YPT), the ML-RRT algorithm was applied to explore the mobility of the WDP loop (residues 352-361). A voxel map obtained from this exploration is represented in the right part of Figure 4. It displays the positions reached by the Cα atom of the middle loop residue Glu357. Voxels size is 0.5 Å, and colors have been assigned depending on the distance between the referred atom and the Cα of Val407, which is located on the bottom of the binding pocket. These atoms were also chosen in [35] to measure the WDP loop gating during molecular dynamics simulations. The distance in the initial crystal structure is 17 Å. The minimum and maximum distances obtained through the conformational exploration are 11 Å and 22 Å respectively. The WDP loop reaches conformations very similar to the one in the closed structure [34] (PDB ID: 1YTN), with Cα RMSD below 1 Å. The voxel map shows that the loop can adopt more open conformations, as also suggested by molecular dynamic simulations [35]. Interestingly, the voxel map presents a marked pipe-like shape. Such a shape indicates that the WDP loop in PTPase is "mechanically" designed to perform opening-closure motions, while lateral-motions are not likely. This mechanical predisposition may explain the rapid opening-closure WDP loop motions reported in [35]. Finally note that constructing the voxel map required about

5000 iterations of the ML-RRT algorithm, with a computing time of 5 minutes.

2) Domain motions: Similarly, the proposed approach can be applied for analyzing protein domain motions. The example presented here concerns the N-Oct-3 POU domain. This molecule, represented in Figure 5, comprises two distinct, highly conserved sub-domains, termed POUs and POUh, connected by a flexible linker [36]. It has been shown that, due to its remarkable plasticity, the N-Oct-3 POU domain can adopt different configurations and corresponding homodimerization patterns, with the POUs and POUh sub-domains acting as sensors for the distinct underlying structures which characterize the respective DNA targets [37]. The conformational transitions of the N-Oct-3 POU domain are currently being investigated by combining experimental techniques (SAXS) and the computational approach presented in this paper. Figure 6 shows preliminary results of this work. The voxel map represents possible locations of POUh with respect to POUs (relative positions of the centers of mass) considering the flexibility of the long linker between them (18 residues are considered to be fully flexible and 13 have limited flexibility). All the protein side-chains are also potentially flexible (i.e. they move if they hinder backbone motions). Maximum and minimum values of the radius of gyration have been integrated as constraints for the conformational exploration. This structural information has been derived from SAXS experiments coupled to molecular mechanics [38]. The voxels size is 2 Å. The wideness of the explored region reflects the high flexibility of the linker. Voxels have been labeled using an energetic scoring. The conformations associated with each voxel have been clustered into significantly different sets. Then, one conformation of each cluster has been energy minimized[2] imposing constraints on the backbone atom positions in order to remain within the voxel. The voxel color is assigned according to the resulting lowest-energy conformation. The goal of this work is to compute a complete conformational map of the protein enabling the identification of low-energy conformations attainable from the initial structure, and possible transition pathways between them. Comparison with free and pre-bound models of N-Oct-3 POU constructed from experimental data and molecular modeling tools will allow us to validate this novel approach.

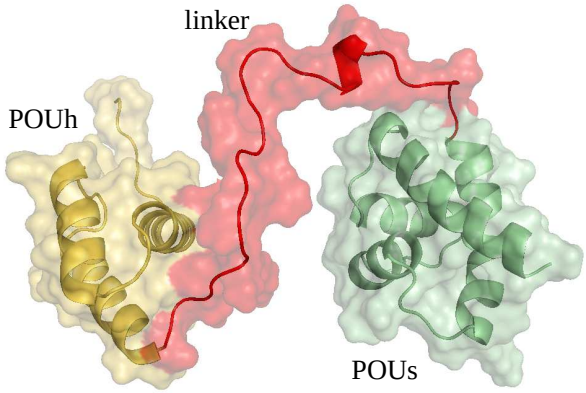

Fig. 5. Structure of N-Oct-3 POU domain.

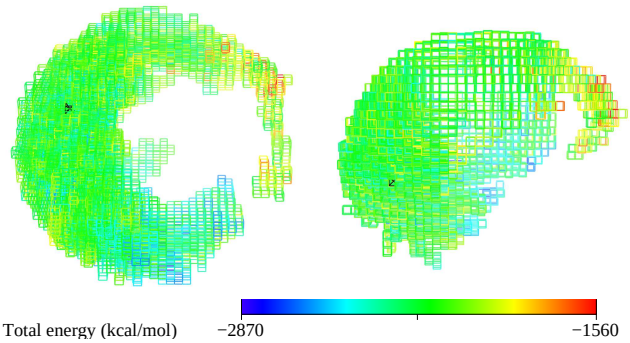

Fig. 6. Two views of the voxel map representing geometrically feasible motions of POUh with respect to POUs. The voxels display the relative position of the centers of mass of both domains. Voxels size is 2 Å, and colors have been assigned depending on the energies of the associated conformations. The black voxel indicates the initial conformation.

IV. Conclusions

We have presented a general framework for computing and representing molecular motions. The basic principle of the approach is to apply path-planning-based algorithms to explore feasible motions of mechanistic molecular models. The efficiency of such conformational exploration permits to attain large-amplitude motions with low computational cost. The voxel-map representation facilitates ulterior treatment of the explored conformations and permits direct visual interpretation of results. Besides, voxel maps could be used to bias or to focus the exploration to specific regions of the conformational space, and could permit to device more accurate metrics, considering motion feasibility, for improved path-planning algorithms.

Voxel maps can represent relative motions of two molecules. Such a representation displays the geometric suitability of a protein presenting a narrow, deep binding site for interacting with different ligands. When applied to explore molecular deformations, voxel maps can provide a global representation of the conformational space of protein loops and protein domains undergoing large-amplitude motions.

First results highlight the potential of the proposed approach. Voxel maps can be seen as a general tool that, combined with other computational and experimental methods, will help to investigate the importance of flexibility and motion in molecular interactions.

[2]AMBER ff03 force field [39] has been used for the energetic analysis.

Acknowledgments

The authors thank Maurice Franssen, David Guieysse and Magali Remaud for providing models and information about CALB. This work has been partially supported by the Région Midi-Pyrénées under projets ITAV ALMA and CTP AMOBIO, and by the French National Agency for Research (ANR) under project NanoBioMod.

REFERENCES

[1] H. Carlson, "Protein flexibility is an important component of structure-based drug discovery," *Curr. Pharm. Des.*, vol. 8, pp. 1571–1578, 2002.

[2] C. Cavasotto, A. Orry, and R. Abagyan, "The challenge of considering receptor flexibility in ligand docking and virtual screening," *Curr. Comput. Aided. Drug. Des.*, vol. 1, pp. 423–440, 2005.

[3] J. Janin, "Assessing predictions of protein-protein interaction: the CAPRI experiment," *Protein Sci.*, vol. 14, pp. 278–283, 2005.

[4] L. Ehrlich, M. Nilges, and R. Wade, "The impact of protein flexibility on protein-protein docking," *Proteins*, vol. 58, pp. 126–133, 2005.

[5] G. Katona, P. Carpentier, V. N. , P. Amara, V. Adam, J. Ohana, N. Tsanov, and D. Bourgeois, "Raman-assisted crystallography reveals end-on peroxide intermediates in a nonheme iron enzyme," *Science*, vol. 316, pp. 449–453, 2007.

[6] P. Schanda, V. Forge, and B. Brutscher, "Protein folding and unfolding studied at atomic resolution by fast two-dimensional nmr spectroscopy," *PNAS*, vol. 104, pp. 11 257–11 262, 2007.

[7] K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins," *Nature*, vol. 450, pp. 964–972, 2007.

[8] A. Leach, *Molecular Modeling: Principles and Applications*. Cambridge: Longman, 1996.

[9] T. Schlick, *Molecular Modeling and Simulation - An Interdisciplinary Guide*. New York: Springer, 2002.

[10] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten, "Steered molecular dynamics," in *Computational Molecular Dynamics: Challenges, Methods, Ideas. Vol. 4 of Lecture Notes in Computational Science and Engineering*, P. Deuflhard, J. Hermans, B. Leimkuhler, A. Mark, S. Reich, and R. Skeel, Eds. Berlin: Springer-Verlag, 1998, pp. 39–65.

[11] L. Mouawad and D. Perahia, "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.*, vol. 258, pp. 393–410, 1996.

[12] J. Jeong, E. Lattman, and G. Chirikjian, "A method for finding candidate conformations for molecular replacement using relative rotation between domains of a known structure," *Acta. Cryst.*, vol. D62, pp. 398–409, 2006.

[13] S. Wells, S. Menor, B. Hespenheide, and M. Thorpe, "Constrained geometric simulation of diffusive motion in proteins," *Phys. Biol.*, vol. 2, pp. 127–136, 2005.

[14] H. Choset, K. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L. Kavraki, and S. Thrun, *Principles of Robot Motion: Theory, Algorithms, and Implementations*. Cambridge: MIT Press, 2005.

[15] S. M. LaValle, *Planning Algorithms*. New York: Cambridge University Press, 2006.

[16] A. Singh, J.-C. Latombe, and D. Brutlag, "A motion planning approach to flexible ligand binding," *Proc. Conf. Intell. Syst. Mol. Biol. (ISMB)*, pp. 252–261, 1999.

[17] M. Apaydin, D. Brutlag, C. Guestrin, D. Hsu, and J.-C. Latombe, "Stochastic conformational roadmaps for computing ensemble properties of molecular motion," in *Algorithmic Foundations of Robotics V*, J.-D. Boissonnat, J. Burdick, K. Goldberg, and S. Hutchinson, Eds. Berlin: Springer-Verlag, 2004, pp. 131–147.

[18] J. Cortés, T. Siméon, V. Ruiz, D. Guieysse, M. Remaud, and V. Tran, "A path planning approach for computing large-amplitude motions of flexible molecules," *Bioinformatics*, vol. 21, pp. i116–i125, 2005.

[19] N. M. Amato, K. A. Dill, and G. Song, "Using motion planning to map protein folding landscapes and analyze folding kinetics of known native structures," *J. Comput. Biol.*, vol. 10, pp. 149–168, 2003.

[20] M. Apaydin, D. Brutlag, C. Guestrin, D. Hsu, J.-C. Latombe, and C. Varma, "Stochastic roadmap simulation: an efficient representation and algorithm for analyzing molecular motion," *J. Comput. Biol.*, vol. 10, pp. 257–281, 2003.

[21] X. Tang, B. Kirkpatrick, S. Thomas, G. Song, and N. Amato, "Using motion planning to study RNA folding kinetics," *J. Comput. Biol.*, vol. 12, pp. 862–881, 2005.

[22] J. Cortés, T. Siméon, M. Remaud-Siméon, and V. Tran, "Geometric algorithms for the conformational analysis of long protein loops," *J. Comput. Chem.*, vol. 25(7), pp. 956–967, 2004.

[23] S. Kirillova, J. Cortés, A. Stefaniu, and T. Siméon, "An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins," *Proteins*, vol. 70, pp. 131–143, 2008.

[24] A. Enosh, S. Fleishman, N. Ben-Tal, and D. Halperin, "Prediction and simulation of motion in pairs of transmembrane α-helices," *Bioinformatics*, vol. 23, pp. e212–e218, 2007.

[25] 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. snd G. Sukhatme, S. Schaal, and O. Brock, Eds. Cambridge: MIT Press, 2005, pp. 6–11.

[26] J. Cortés, L. Jaillet, and T. Siméon, "Disassembly path planning for complex articulated objects," *IEEE Transactions on Robotics*, vol. 24, pp. 475–481, 2008.

[27] M. Zhang and L. Kavraki, "A new method for fast and accurate derivation of molecular conformations," *J. Chem. Inf. Comput. Sci.*, vol. 42(1), pp. 64–70, 2002.

[28] S. M. LaValle and J. J. Kuffner, "Rapidly-exploring random trees: Progress and prospects," in *Algorithmic and Computational Robotics: New Directions (WAFR2000)*, B. Donald, K. Lynch, and D. Rus, Eds. Boston: A.K. Peters, 2001, pp. 293–308.

[29] W. DeLano, "The PyMOL molecular graphics system," 2002, http://www.pymol.org.

[30] V. Cojocaru, P. Winn, and R. Wade, "The ins and outs of cytochrome P450s," *Biochim. Biophys. Acta*, vol. 1770, pp. 390–401, 2007.

[31] M. Petřek, M. Otyepka, P. Banás, P. Košinová, J. Koča, and J. Damborský, "CAVER: A new tool to explore routes from protein clefts, pockets and cavities," *BMC Bioinfo.*, vol. 7, pp. 316–324, 2006.

[32] D. Guieysse, J. Cortés, S. Puech-Guenot, S. Barbe, V. Lafaquière, P. Monsan, T. Siméon, I. André, and M. Remaud-Siméon, "A structure-controlled lipase enantioselectivity investigated by a path planning approach," *ChemBioChem*, vol. 9, pp. 1308–1317, 2008.

[33] M. Franssen, M. Kamp, L. Alessandrini, M. Huibers, and J. Vervoort, "The interaction between candida antarctica lipase and branched chain fatty acids: a kinetic and molecular modelling study," *Proc. Int. Symp. Biocatalysis and Biotransformations (Biotrans)*, 2005.

[34] J. Stuckey, H. Schubert, E. Fauman, Z.-Y. Zhang, J. E. Dixon, and M. Saper, "Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 Å and the complex with tungstate," *Nature*, vol. 370, pp. 571–575, 1994.

[35] X. Hu and C. Stebbins, "Dynamics of the WPD loop of the Yersinia protein tyrosine phosphatase," *Biophys. J.*, vol. 91, pp. 948–956, 2006.

[36] D. Latchman, "POU family transcription factors in the nervous system," *J. Cell. Physiol.*, vol. 179, pp. 126–133, 1999.

[37] R. Alazard, M. Blaud, S. Elbaz, C. Vossen, G. Icre, G. Joseph, L. Nieto, and M. Erard, "Identification of the 'NORE' (N-Oct-3 responsive element), a novel structural motif and composite element," *Nucleic Acids Res.*, vol. 33, pp. 1513–1523, 2005.

[38] R. Alazard, L. Mourey, C. Ebel, P. Konarev, M. Petoukhov, D. Svergun, and M. Erard, "Fine-tuning of intrinsic N-Oct-3 POU domain allostery by regulatory DNA targets," *Nucleic Acids Res.*, vol. 35, pp. 4420–4432, 2007.

[39] Y. Duan, C. Wu, S. Chowdhury, M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins," *J. Comput. Chem.*, vol. 24, pp. 1999–2012, 2003.