

## Review Article

# Considering Global and Local Conformational Changes in Molecular Docking

Philot EA<sup>1,4†</sup>, Lima AN<sup>2†</sup>, Resende-Lara PT<sup>1,3†</sup>, Tuffaneto PC<sup>1,3</sup>, and Scott LPB<sup>1,4†\*</sup>

<sup>1</sup>Laboratório de Biologia Computacional e Bioinformática, Universidade Federal do ABC, Brazil

<sup>2</sup>Programa de Pós-Graduação em Ciências Fisiológicas, Universidade Federal do Triângulo Mineiro, Brazil

<sup>3</sup>Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Brazil

<sup>4</sup>Centro de Matemática Computação e Cognição, Universidade Federal do ABC, Brazil

<sup>†</sup>These authors contributed equally to this work

**\*Corresponding author**

Scott LPB, Laboratório de Biologia Computacional e Bioinformática, Universidade Federal do ABC, Av. dos Estados 5001, Bangu, Santo André - SP - Brasil, Tel: 551149960129; Email: luisp37@gmail.com

Submitted: 28 March 2016

Accepted: 22 September 2016

Published: 24 September 2016

ISSN: 2333-6633

**Copyright**

© 2016 Philot et al.

**OPEN ACCESS****Keywords**

- Conformational selection
- Induced fit
- Protein flexibility
- Molecular docking

**Abstract**

Proteins undergo changes in their form (conformational changes) upon interaction with compounds/substrates. Molecular docking is an important tool used in the study of correlations between structure and function, aiding the understanding of several biological processes, shedding light on drug development. Structural rearrangements can occur during molecular recognition in order to optimize interactions in the complex, leading to local and global conformational changes. Conformational selection and induced fit are models that attempt to explain structural variation effects in molecular recognition. In this review we discuss the different strategies employed for global and local conformational changes, in both protein-ligand and protein-protein docking.

**ABBREVIATIONS**

ANM: Anisotropic Network Models; CS: Conformational Selection; IF: Induced Fit; NMA: Normal Modes Analysis

**INTRODUCTION**

are inherently dynamic systems with various internal motions that result in different shapes (conformational changes) upon interactions with substrate and/or compounds [1,2]. These conformational changes allow proteins to function properly, as well as to modulate other protein's functions [1]. Since structure and function are intrinsically related, their understanding is essential for drug design [1]. During the molecular recognition process (either protein-ligand or protein-protein) conformational motions can occur in various degrees, ranging from small local changes (vibrational motions) to collective global changes (domain and allosteric motions). This protein dynamics plays an essential role in their functions, which is determined by the flexibility pattern in wild type [3], mutant [4] or allosteric sites [5] [6,7]. In this context, flexibility is essential in the study of protein collective motions, interface formation of the proteins complexes, and protein function as a whole.

Molecular docking is a computational technique that simulates interactions between biomolecules, which can be two proteins (protein-protein docking), or a protein and a compound

(protein-ligand docking). These approaches are important for drug development and the understanding of several biological processes.

Docking methods can be classified in three classes, according to flexibility: rigid, semi-flexible, and flexible. Most methods/software falls into the two first categories due to the high computational cost for flexible docking. The different levels of approximations for docking simulations depend on the type and number of freedom degrees. The fully flexible - which considers all degrees of freedom - is the most interesting case where both molecules (ligand and receptor) change their conformations [8], however, this approach has high computational cost and still poses challenges.

Since flexible docking is highly demanding from both time and computational viewpoints, a number of residues with backbone flexibility and/or side-chain free torsions should be accounted in protein-protein docking protocols in order to access various conformational set [9]. These procedures attempt to mimic conformational selection and induced fit, respectively. In both cases, docking simulations must be accurate in regards to two main parameters: the search for binding configurations and result classification through scoring functions.

There are two different models to explain protein dynamics in molecular recognition: conformational selection (CS) and induced

fit (IF). In CS model, the receptor undergoes global changes with a large conformation set presenting and equilibrium of both active and inactive forms. The active conformer is preferred by the ligand, which interacts and forms the interface. Induced fit (IF) models consider structural changes that receptors undergo after ligand binding, that are generally local sidechain rearrangements.

several computational approaches are being used to simulate CS and IF. Among them, elastic network models like Anisotropic Network Models (ANM) and Normal Modes Analysis (NMA) provide intrinsic flexibility data about receptor structure giving an overview of protein collective motions [10–13]. Rotamer libraries are also used to adjust and refine results after docking calculations. There are many different methods proposed to consider protein conformational change (both local and global) in molecular recognition. Some of the procedures applied to protein-ligand and protein-protein docking are described in the following sections.

## PROTEIN-LIGAND DOCKING

Protein-ligand docking consists in predicting conformations of small molecules (ligands) in the binding sites of macromolecules (receptors). This is a relevant technique in the context of inhibitor/modulator drug design for disease related proteins. Currently there are numerous programs that employ this methodology. The most cited protein-ligand docking programs are Autodock [14], GOLD [15] and Glide [16]. Until 2013, 25,87% of works published in this area used the Autodock program 16,69% used GOLD and 11,38% Glide of the main computational challenges for protein-ligand docking is protein flexibility, due to the macromolecule's number of freedom degrees. Protein flexibility can be approached by four methodologies: (i) soft docking, (ii) selective docking, (iii) ensemble docking and (iv) on-the-fly docking. The simplest way to introduce partial flexibility to protein is soft docking. The technique consists in a soft potential in the receptor-ligand interface, i.e., a smoothing of the Lennard-Jones potential in this region [17,18]. Small adjustments in the receptor-ligand interface are allowed through a closer approximation of the same. The programs Dock [19] and Glide use a scaling factor of the Lennard-Jones radii. Others examples are the Lennard-Jones 8–4 potential in GOLD and smooth potential in AutoDock 3.0. docking also interpreters the receptor as partially flexible, considering specific regions of the receptor, like side chains of active site, as flexible [20]. Rotamer libraries are one of the main tools used for this purpose. GOLD AutoDock 4 and FITTED [21] use rotamer libraries as a set of energetically accessible conformations of side chains for selected residues - ICM [22] uses rotamer libraries combined with the Monte Carlo search algorithm. In 2012 Lima et al developed a methodology denominated GANM that combines genetic algorithms, NMA and rotamer libraries with protein-ligand docking simulations. The use of rotamer libraries allows side chains of the active site residues flexible [23].

Ensemble and on-the-fly docking are approaches that can comprise conformational changes of large amplitude for the receptor. DOCK and FlexE [24] are examples of programs that use ensemble docking - which consists of a set of protein conformations, obtained either experimentally or computationally (through the use of techniques like NMA, molecular dynamics and principal components analysis), as opposed to the traditional approach, in

which only one receptor structure is targeted for docking [25,26]. et al used experimentally derived conformations to perform flexible docking to the cytochrome c peroxidase - the ensemble of the receptor structure was built from multiple conformations present in the electron-density map of the cavity site of an apo structure of cytochrome. They used 583,363 compounds in 16 conformations of this protein, resulting in new ligand for the cytochrome c peroxidase [27]. Sperandio et al., used NMA to create a structure ensemble of cyclin-dependent kinase 2. Their protocol selected several receptor conformations suitable for docking [28]. Philot et al., also identified binding sites to human thioredoxin 1 from an ensemble of structures obtained through NMA to investigate the docking with three phenothiazine drugs [29,30].

Lastly, on-the-fly docking treats the receptor as fully flexible, i.e., it changes the receptor conformations during the docking. Due to the large amount of freedom degrees present in this type of simulation, other strategies are applied in order to reduce the computational cost. The first strategy is to dock the ligand in a rigid conformation receptor, later changing the side chains of proteins using rotamer library and minimizing the complex [17]. This methodology is applied by ROSETTALIGAND [31].

## PROTEIN-PROTEIN DOCKING

In the protein-protein docking approach the receptor protein is considered to be fixed, while the ligand protein moves (rotation and translation) in the binding mode search. The protein flexibility protocols in protein-protein docking are very similar to the previously discussed, and can be treated implicitly or explicitly. Implicit flexibility is performed through the use of soft scoring functions and/or representation of protein structure, which allows for a penetration degree between specific portions of the proteins, or the softening of their molecular surfaces [32,33]. Therefore an optimization stage is required to remove steric clashes - Hex [34] uses spherical polar Fourier correlations to describe the protein structure and soft potential to protein flexibility, for instance. Soft docking is also able to describe local flexibility of side chains and small amplitude rearrangements for backbone and loops [32,33].

The explicit approach to protein flexibility is subdivided in partially flexible or fully flexibility methodologies [35]. Partially flexible methodologies are employed in several ways: some software's apply rigid body protein-protein docking followed by subsequent refinements in complex interface to provide partial flexibility. These refinement steps may allow conformational changes of side-chains, loops and local interface backbone. ICM-DISCO [36] combines soft docking - rigid-body motions and soft potentials followed by optimization of side-chain at complex interface. This method provides good results for targets that display small amplitude conformational change.

Bastard et al. [37], proposed a successful method to predict complex geometry considering loop flexibility in protein-protein docking. They build an ensemble of possible loop arrangements (multi-copy), minimized the energy of ligand starting configuration and selected the best loop configuration, according to RMSD and average energy. HADDOCK uses different approaches to consider protein flexibility in docking protocol:

(i) rigid body with energy minimization, (ii) semi-flexible refinement of interface residues by side-chain and backbone and (iii) Cartesian refinement in solution [10].

Soft and semi-flexible docking cannot accurately model large amplitude conformational changes (global backbone and domain motions) when using one single conformation of complex partners. These global motions have been simulated with ensemble docking and on-the-fly protocols. In ensemble docking a set of pre-generated conformations obtained from different sources (simulation and/or experimental) [38] can be used instead of a single conformation. Trellet et al. [39], used protein-peptide ensemble flexible docking methodology with a set of three distinct peptide conformations - extended,  $\alpha$ -helix and polyproline-II. Similarly, Sahu et al. [40], used protein-protein ensemble docking to investigate the stability of  $\alpha$ -synuclein dimers - they used two totally distinct conformations of a  $\alpha$ -synuclein monomer: one with experimental  $\alpha$ -helix structure, and the other with modeled  $\beta$ -structure.

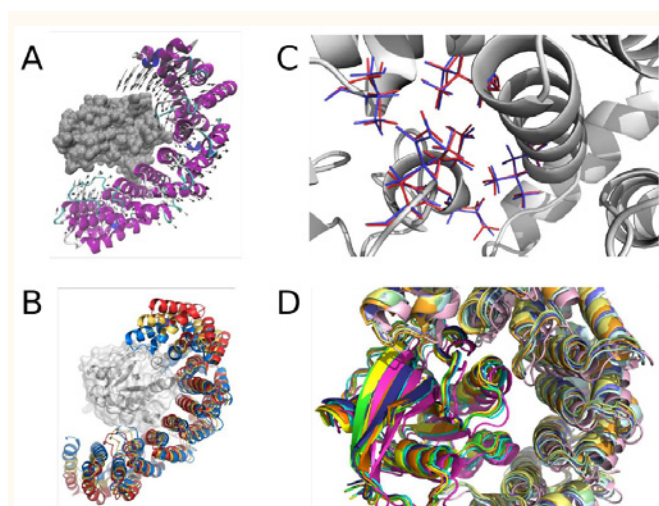
Several protein-protein docking softwares of on-the-fly scheme have used NMA to account for global changes, due to the ability of such method in describing large amplitude motions [41]. Venkatraman and Ritchie Eigen Hex [13] developed an algorithm that uses normal modes analysis of a simple elastic network of a protein flexibility - initially performing rigid docking with later soft docking approaches. They used the swarm optimization algorithm to perturb docked conformation using NMA data in a pose-dependent way which allows moderate flexibility.

Fiber Dock [42] is yet another methodology that uses normal modes to account for backbone flexibility. This software interactively chooses relevant modes and minimizes the flexible protein structure along each one of them, achieving a flexibility refinement that allows side-chain and backbone conformational changes. As described in this work, rotamer libraries are used to side-chain flexibility and normal modes to backbone. Other methodologies as cNMA [12], Swarm Dock [43], etc, also use NMA to account for global changes during molecular recognition.

Figure (1) illustrates the most common ways to sample conformational changes and a few of the methods used to describe them, summarizing the methodologies discussed in this topic.

## DISCUSSION & CONCLUSION

Due the importance of protein-protein and protein-ligand docking in the drug design process, new methods and tools that consider the global and local conformational changes have been more and more investigated and perfected during the past years. The methodologies used to confer flexibility in protein-ligand and protein-protein docking is similar and attempt to describe molecular recognition models. Conformational selection is often necessary to molecular binding relaxation, and induced fit is not mandatory but sufficient in certain cases. Therefore, robust methods must employ techniques that simulate mainly CS followed by optimization steps that simulate IF. Different approaches treat the flexibility explicitly, involving local and global structural rearrangements that consider side-chain optimization and backbone displacement. Ensemble docking and on-the-fly strategies have been used to describe global



**Figure 1** Conformation sampling in molecular docking of protein RAN and Importin  $\beta$  through various methods. A) Importin  $\beta$  (cartoon) adjusts in RAN surface, increasing the contact area and, thus, promoting more specificity in binding. Elastic networks models allows to sample receptor conformers with large displacements through global flexibility data. This data can be achieved by performing NMA or ANM. B) On-the-fly adjustment of importin  $\beta$  in RAN. Accounting global flexibility while docking is performed increase the accuracy, but also with the increase of the computational cost. C) Importin  $\beta$  and RAN side chain adjustments. Side chain rotamers provide local optimization of molecular contacts, refining final solutions. This approach can be employed by a rotamer library or molecular dynamics. D) Importin  $\beta$  and RAN multiple sampling with molecular dynamics. This approach can be used to generate multiple conformers of both molecules and perform rigid ensemble docking.

conformational changes, being employed in combination with NMA, Molecular Dynamics, among others, to achieve this goal. Another strategy is to perform rigid docking as a first step, later optimizing the best solutions with post-processing algorithms that consider protein motions and flexibility. This strategy combines different tools and methods with sophisticated scoring functions to improve the accuracy of solutions

Regardless of the methods used, we conclude that it is advantageous to take into account local and global structural changes in protein-ligand and protein-protein docking. Different sampling methodologies have improved the prediction capacity of molecular docking software's in recent years. In spite of the resolution level (coarse or atomistic), elastic network models (mainly ANM and NMA) emerge as promising methods to compute protein flexibility, allowing better interface adjustments and overall results. Rotamer library is still a good approach to minor, but important, refinements. However, the field remains challenging - the improvement of the accuracy /computational cost ratio while maintaining robustness in molecular sampling can shed some light to new methods in following years.

## ACKNOWLEDGEMENTS

The authors would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Fundação de Amparo à Pesquisa do Estado de São Paulo and Universidade Federal do ABC for funding.

## REFERENCES

- Moree B, Connell K, Mortensen RB, Liu CT, Benkovic SJ, Salafsky J. Protein Conformational Changes Are Detected and Resolved Site Specifically by Second-Harmonic Generation. *Biophys J*. 2015; 109: 806-815.
- Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature*. 2007; 450: 964-972.
- Uyar A, Kurkcuoglu O, Nilsson L, Doruker P. The elastic network model reveals a consistent picture on intrinsic functional dynamics of type II restriction endonucleases. *Phys Biol*. 2011; 8: 056001.
- De Conto V, Braz AS, Perahia D, Scott LP. Recovery of the wild type atomic flexibility in the HIV-1 protease double mutants. *J Mol Graph Model*. 2015; 59: 107-116.
- Yao XQ, Grant BJ. Domain-opening and dynamic coupling in the  $\alpha$ -subunit of heterotrimeric G proteins. *Biophys J*. 2013; 105: L08-10.
- Karplus M, Petsko GA. Molecular dynamics simulations in biology. *Nature*. 1990; 347: 631-639.
- Zhou H, Li S, Badger J, Nalivaika E, Cai Y, Foulkes-Murzycki J, et al. Modulation of HIV protease flexibility by the T80N mutation. *Proteins*. 2015; 83: 1929-1939.
- May A, Zacharias M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim Biophys Acta*. 2005; 1754: 225-231.
- Kilambi KP, Reddy K, Gray JJ. Protein-protein docking with dynamic residue protonation states. *PLoS Comput Biol*. 2014; 10: e1004018.
- de Vries SJ, van Dijk AD, Krzeminski M, van Dijk M, Thureau A, Hsu V, et al. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*. 2007; 69: 726-733.
- Frezza E, Lavery R. Internal Normal Mode Analysis (iNMA) Applied to Protein Conformational Flexibility. *J Chem Theory Comput*. 2015; 11: 5503-5512.
- Oliwa T, Shen Y. cNMA: a framework of encounter complex-based normal mode analysis to model conformational changes in protein interactions. *Bioinformatics*. 2015; 31: i151-160.
- Venkatraman V, Ritchie DW. Flexible protein docking refinement using pose-dependent normal mode analysis. *Proteins*. 2012; 80: 2262-2274.
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009; 30: 2785-2791.
- Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol*. 1995; 245: 43-53.
- Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J Med Chem*. 2006; 49: 6177-6196.
- Antunes DA, Devaurs D, Kavraki LE. Understanding the challenges of protein flexibility in drug design. *Expert Opin Drug Discov*. 2015; 10: 1301-1313.
- Jiang F, Kim SH. "Soft docking": matching of molecular surface cubes. *J Mol Biol*. 1991; 219: 79-102.
- Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, et al. DOCK 6: Impact of new features and current docking performance. *J Comput Chem*. 2015; 36: 1132-1156.
- Guedes IA, de Magalhães CS, Dardenne LE. Receptor ligand molecular docking. *Biophys Rev*. 2013; 6: 75-87.
- Corbeil CR, Englebienne P, Moitessier N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J Chem Inf Model*. 2007; 47: 435-449.
- Neves MA, Totrov M, Abagyan R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J Comput Aided Mol Des*. 2012; 26: 675-686.
- Lima AN, Philot EA, Perahia D, Braz ASK, Scott LPB. GANM: A protein-ligand docking approach based on genetic algorithm and normal modes. *Appl Math Comput*. 2012; 219: 511-520.
- Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol*. 2001; 308: 377-395.
- Kokh DB, Wade RC, Wenzel W. Receptor flexibility in small-molecule docking calculations. *Wiley Interdiscip Rev Comput Mol Sci*. 2011; 1: 298-314.
- Sousa SF, Ribeiro AJ, Coimbra JT, Neves RP, Martins SA, Moorthy NS, et al. Protein-ligand docking in the new millennium--a retrospective of 10 years in the field. *Curr Med Chem*. 2013; 20: 2296-2314.
- Fischer M, Coleman RG, Fraser JS, Shoichet BK. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat Chem*. 2014; 6: 575-583.
- Sperandio O, Mouawad L, Pinto E, Villoutreix BO, Perahia D, Miteva MA. How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis. *Eur Biophys J*. 2010; 39: 1365-1372.
- Philot EA, Perahia D, Braz AS, Costa MG, Scott LP. Binding sites and hydrophobic pockets in Human Thioredoxin 1 determined by normal mode analysis. *J Struct Biol*. 2013; 184: 293-300.
- Philot EA, da Mata Lopes D, de Souza AT, Braz AS, Nantes IL, Rodrigues T, et al. Binding of phenothiazines into allosteric hydrophobic pocket of human thioredoxin 1. *Eur Biophys J*. 2016; 45: 279-286.
- Meiler J, Baker D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins*. 2006; 65: 538-548.
- Andrusier N, Mashiah E, Nussinov R, Wolfson HJ. Principles of flexible protein-protein docking. *Proteins*. 2008; 73: 271-289.
- Huang SY. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. *Drug Discov Today*. 2014; 19: 1081-1096.
- Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. *Proteins*. 2000; 39: 178-194.
- Park H, Lee H, Seok C. High-resolution protein-protein docking by global optimization: Recent advances and future challenges. *Curr Opin Struct Biol*. 2015; 35: 24-31.
- Fernández-Recio J, Totrov M, Abagyan R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins*. 2003; 52: 113-117.
- Bastard K, Prévost C, Zacharias M. Accounting for loop flexibility during protein-protein docking. *Proteins*. 2006; 62: 956-969.
- Chaudhury S, Gray JJ. Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles. *J Mol Biol*. 2008; 381: 1068-1087.
- Trellet M, Melquiond AS, Bonvin AM. A unified conformational selection and induced fit approach to protein-peptide docking. *PLoS One*. 2013; 8: e58769.
- Sahu KK, Woodside MT, Tuszynski JA.  $\alpha$ -Synuclein dimer structures

- found from computational simulations. *Biochimie*. 2015; 116: 133-140.
41. Dobbins SE, Lesk VI, Sternberg MJ. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proc Natl Acad Sci U S A*. 2008; 105: 10390-10395.
42. Mashiach E, Nussinov R, Wolfson HJ. FiberDock: Flexible induced-fit backbone refinement in molecular docking. *Proteins*. 2010; 78: 1503-1519.
43. Moal IH, Bates PA. SwarmDock and the use of normal modes in protein-protein docking. *Int J Mol Sci*. 2010; 11: 3623-3648.

#### Cite this article

Philot EA, Lima AN, Resende-Lara PT, Tuffaneto PC, Scott LPB (2016) Considering Global and Local Conformational Changes in Molecular Docking. *JSM Chem* 4(3): 1031.