

Chapter 42

Protein–Protein Docking: Are We There Yet?

Horia Jalily Hasani

University of Alberta, Canada

Khaled H. Barakat

University of Alberta, Canada

ABSTRACT

Protein-protein docking algorithms are powerful computational tools, capable of analyzing the protein-protein interactions at the atomic-level. In this chapter, we will review the theoretical concepts behind different protein-protein docking algorithms, highlighting their strengths as well as their limitations and pointing to important case studies for each method. The methods we intend to cover in this chapter include various search strategies and scoring techniques. This includes exhaustive global search, fast Fourier transform search, spherical Fourier transform-based search, direct search in Cartesian space, local shape feature matching, geometric hashing, genetic algorithm, randomized search, and Monte Carlo search. We will also discuss the different ways that have been used to incorporate protein flexibility within the docking procedure and some other future directions in this field, suggesting possible ways to improve the different methods.

INTRODUCTION

Protein-protein interactions play key roles in several biological processes. These processes involve many essential mechanisms ranging from signal transduction and cellular transport to gene expression and immune responses. All these processes are mediated by selective and potent protein-protein interactions (Waksman & Sansom, 2005). Furthermore, many diseases have been associated with either an over-activated or an under-regulated protein-protein interaction and the cure for these diseases has been focused on either inhibiting or stimulating these interactions, respectively. For example, the p53-MDM2 interaction is associated with a severe down regulation of the p53 pathway. An inhibitor for this interaction (e.g. nutlin3) can reactivate the p53 pathway, forcing cancer cells to undergo apoptosis (Barakat, Gajewski, & Tuszynski, 2012; Barakat, Issack, Stepanova, & Tuszynski, 2011; Barakat, Mane, Friesen,

DOI: 10.4018/978-1-5225-1762-7.ch042

Protein-Protein Docking

& Tuszynski, 2010; Chène, 2003; Kojima et al., 2006). The more we know about such crucial interactions, the more we can build vital protein networks and apply this knowledge to identify treatments for many diseases. Moreover, characterizing these interactions at the atomic level can help in rationally designing new therapeutic agents that can either enhance or inhibit these interactions. Constructing a three dimensional structure of such protein complexes is an essential step toward identifying their binding interface and recognizing any hot spots that can be targeted for their regulation (Elcock et al. 2001; Kann, 2007; Kortemme & Baker, 2004).

For the last few decades X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and electron microscopy have been the main source to predict such structures. Despite their accuracy, efficiency and the huge amount of details they can provide, they are expensive and very labour and skill demanding. A simple comparison between protein structure and gene sequence databases would simply reveal the great discrepancy between the two. That is, although hundreds of thousands of gene expressions have been characterized, only less than thirty thousand protein structures have been determined so far and most of these structures are either redundant or describe only apo (unbound) proteins (Villoutreix et al., 2014). Moreover, protein complexes are more difficult to crystallize than the individual proteins, consequently, they are less represented in Protein Data Bank (PDB) (Berman et al., 2000) and constitute only a small fraction of the experimentally determined structures. This huge discrepancy and lack of structural details motivated many computational groups to fill this gap and suggest a new, rapid and cheap way to predict these interactions (Barakat & Tuszynski, 2011; Barakat, Houghton, Tyrrell, & Tuszynski, 2014; Barakat, Mane, & Tuszynski, 2011; Gógl et al., 2015; Nillegoda et al., 2015; Pedotti, Simonelli, Livoti, & Varani, 2011; Taylor et al., 2015). One solution they provided which is also the focus of this chapter is protein-protein docking.

Protein-protein docking simulations date back to the early 90s (Katchalski-Katzir et al., 1992), soon after the development of many protein-ligand docking techniques. Although the concept of docking in both cases is similar, there are still key differences that prevented the direct technology transfer between the two problems. For example, since proteins are continually dynamic and usually undergo huge structural changes, their modeling is not as easy as that of a small molecule and requires more precise predictions. Furthermore, proteins are large and have complex shapes making them harder to simulate than ligand-protein interactions (Ravikant & Elber, 2011; Sheinerman et al. 2000). On the other hand, the concept of a binding site is usually inapplicable to protein-protein interactions and in many cases proteins interact through ‘hot spot’ areas on their surfaces. Identifying these binding locations is particularly difficult and requires highly innovative methods. In this context, the search methods in protein-protein docking simulations are somehow different from those being used in small molecule docking (Keskin, Ma, & Nussinov, 2005). However, the reader should keep in mind that although protein-protein docking is in many ways different from protein-ligand docking, some of its reflections might be mentioned in this chapter, if applicable.

In this chapter, we will first outline the concept of docking in general, with an emphasis on the different approaches. We then briefly review the theoretical concepts, principles and specific features of the search strategies and scoring functions implemented in different programs, highlighting their strengths as well as their limitations. We will also discuss the different methods of incorporating protein flexibility within the docking procedure. After these descriptive review sections, we will guide the reader through a number of case studies and recent successes that illustrate some of the significant applications of protein-protein docking algorithms. We will also discuss the existing challenges and limitations in current algorithms and highlight some potential future directions and possible improvements in this

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/protein-protein-docking/174162

Related Content

Enzymatic Research Having Pharmaceutical Significance

Ishan H. Raval and Arvind Kumar Singh Chandel (2018). *Research Advancements in Pharmaceutical, Nutritional, and Industrial Enzymology* (pp. 141-158).

www.irma-international.org/chapter/enzymatic-research-having-pharmaceutical-significance/203814

Pharmaceutical and Medical Applications of Nanofibers

Khosro Adibkia, Shadi Yaqoubi and Solmaz Maleki Dizaj (2017). *Novel Approaches for Drug Delivery* (pp. 338-363).

www.irma-international.org/chapter/pharmaceutical-and-medical-applications-of-nanofibers/159672

Colonic Bacterial Enzymes: Pharmaceutical Significance and Applications

Srushti M. Tambe and Namita D. Desai (2018). *Research Advancements in Pharmaceutical, Nutritional, and Industrial Enzymology* (pp. 71-99).

www.irma-international.org/chapter/colonic-bacterial-enzymes/203811

Industrial Enzyme Technology: Potential Applications

Michael Bamitale Osho (2018). *Research Advancements in Pharmaceutical, Nutritional, and Industrial Enzymology* (pp. 375-394).

www.irma-international.org/chapter/industrial-enzyme-technology/203824

Cluster Origin of Solvation Features of C-Nanostructures in Organic Solvents

Francisco Torrens and Gloria Castellano (2016). *Advancing Pharmaceutical Processes and Tools for Improved Health Outcomes* (pp. 189-293).

www.irma-international.org/chapter/cluster-origin-of-solvation-features-of-c-nanostructures-in-organic-solvents/150020