Progress in Biophysics and Molecular Biology 114 (2014) 123–136

Contents lists available at SciVerse ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Original research

Modelling and enhanced molecular dynamics to steer structure-based drug discovery

Subha Kalyaanamoorthy, Yi-Ping Phoebe Chen*

Department of Computer Science and Computer Engineering, Faculty of Science, Technology and Engineering, La Trobe University, Melbourne, VIC 3086, Australia

ARTICLE INFO

Article history:
Available online 1 July 2013

Keywords:
Modelling
Molecular dynamics
Enhanced sampling molecular dynamics
Structure based drug discovery

ABSTRACT

The ever-increasing gap between the availabilities of the genome sequences and the crystal structures of proteins remains one of the significant challenges to the modern drug discovery efforts. The knowledge of structure-dynamics-functionalities of proteins is important in order to understand several key aspects of structure-based drug discovery, such as drug–protein interactions, drug binding and unbinding mechanisms and protein–protein interactions. This review presents a brief overview on the different state of the art computational approaches that are applied for protein structure modelling and molecular dynamics simulations of biological systems. We give an essence of how different enhanced sampling molecular dynamics approaches, together with regular molecular dynamics methods, assist in steering the structure based drug discovery processes.

1. Introduction

Enormous biological data, from raw genome sequences to high quality three dimensional structures of proteins, are becoming increasingly available. Drawing relationships between the sequences to structures and then to functions remains an important step in drug design efforts (Krissinel, 2007). A couple of advanced computational approaches, structural bioinformatics and molecular dynamics, go hand-in-hand with experiments to decipher these relationships. Structural bioinformatics applies modern techniques to develop biological insights from protein structures, while molecular dynamics studies help in gaining relevant biophysical insights from the structures. In either case, the focus is to unravel the relationships between the protein structures and functionalities in order to accelerate drug design and discovery processes.

Knowledge of the 3D structure of proteins is a major prerequisite for structure-based drug design (SBDD) (Henry, 2001; Kalyaanamoorthy and Chen, 2011). The Protein Data Bank (Berman et al., 2000) (PDB, accessible at www.pdb.org) is one of the most popular online databases, holding structural information on proteins. According to recent statistics published in 2012, PDB contains the structures of almost 79,120 proteins (Berman et al., 2012), however, the number of amino acid sequences reported in the UniProtKB/Swiss-Prot database (http://www.uniprot.org/) of European bioinformatics institute (EBI), for instance exceeds 537,000. This huge gap (Flohil et al., 2002) between the annotated genome sequences and the experimental structures remains a challenge in modern SBDD efforts (Cavasotto and Phatak, 2009).

Large-scale genome sequencing technologies are seeing rapid advancements and it is now possible to sequence a complete human genome in about eight days and at a cost of approximately $10,000. With the advent of next generation ultra-high-throughput sequencing methods, the nanopore sequencing technique, for instance, researchers are aiming towards a 15-min genome sequencing capability at a cost of only $1000.[1] On the other hand, resolving the structures of protein sequences through experimental techniques continues to fall behind. Although common experimental techniques such as X-ray crystallography and nuclear magnetic resonance (NMR), which are used to determine the 3D structures of proteins, are also advancing consistently, the speed of genome sequencing cannot be matched. This, in turn, demands the use of computational methods to bridge the sequence-structure gap and assist in SBDD efforts (Rost and Sander, 1996).

Amino acids are known as basic molecular building blocks that determine the tertiary structures and functions of proteins, and therefore, it should be possible to predict the 3D structures of proteins with amino acid sequences (Berg et al., 2002). Different computational approaches such as homology modelling or comparative modelling, threading and *de novo* modelling, are used

* Corresponding author.
E-mail address: phoebe.chen@latrobe.edu.au (Y.-P.P. Chen).

[1] http://www.economist.com/node/18304268.

0079-6107/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pbiomolbio.2013.06.004

to predict the tertiary structures of unknown proteins (Dalton and Jackson, 2007). Fig. 1 presents a conceptual overview of the different methods that are employed in protein structure modelling. In this review, we discuss briefly the different computational approaches, from molecular modelling to state-of-the-art enhanced molecular dynamics approaches, which are useful for accelerating the lead discovery in SBDD.

The homology modelling approach is based on the principle that evolutionary-related proteins could share a similar 3D structure (Marti-Renom et al., 2001; Pitman et al., 2006) This approach is applicable when the target (i.e. unknown protein) and the template (i.e. the evolutionarily related protein with experimental structure) shares >30% of sequence identity. On the other hand, when the target sequence does not have an evolutionarily related template with a sequence identity of >30%, other methods such as threading and *de novo* modelling can be useful. Most proteins are commonly believed to share similar folding patterns, irrespective of their evolutionary relationship. As a result, the threading method uses the folds of the known proteins to construct the tertiary structure of the target protein sequence. In this approach the target sequence is used to search the structure fold databases, Structural Classification of Proteins (SCOP)(Murzin et al., 1995) for instance, in order to identify the proteins that have similar folding patterns. The sequence − structure alignments in the threading approach is then evaluated using a scoring function. THREADER (Jones et al., 1998) is one of the most popular program for constructing protein models using the threading approach. However, the threading approach may be inappropriate when a target sequence possesses a novel folding pattern, which does not match with any of the known protein folds.

When both homology modelling and threading approaches are not applicable, then *de novo* modelling can be employed. *De novo* methods are based on the physical principles that native proteins generally prefer global energy-minimum states. According this hypothesis the *de novo* modelling approach performs a large scale conformational search on the energy landscape to identify low energy structural conformations for the given target sequence (Bonneau and Baker, 2001). Methods such as Monte Carlo simulations and molecular dynamic approaches are used to search the energy landscape. Recently, there have been considerable advancements in the *de novo* structure prediction programs including ROSETTA (Huang et al., 2011) and I-TASSER (Zhang, 2008). The need for more computational time for the extensive sampling of the energy space remains one of the significant challenges in the *de novo* modelling approaches.

2. Homology modelling

Homology modelling approach has gained considerable popularity in bioinformatics and SBDD (Ginalski, 2006; Marti-Renom et al., 2001). The 3D model of a protein of interest (known as 'target') is constructed using the known structures of its closely related proteins (known as 'homologs'). As shown in Fig. 2(Eswar et al., 2001), the homology modelling approach involves four major steps (Marti-Renom et al., 2001) in constructing a 3D model of the unknown protein such as: (i) template identification, (ii) target-template sequence alignment, (iii) model construction and (iv) model evaluation.

(i) Template identification

The homology modelling approach begins with the identification of a suitable template for the query sequence. The target sequence (whose structure is to be modelled) is used as a query to

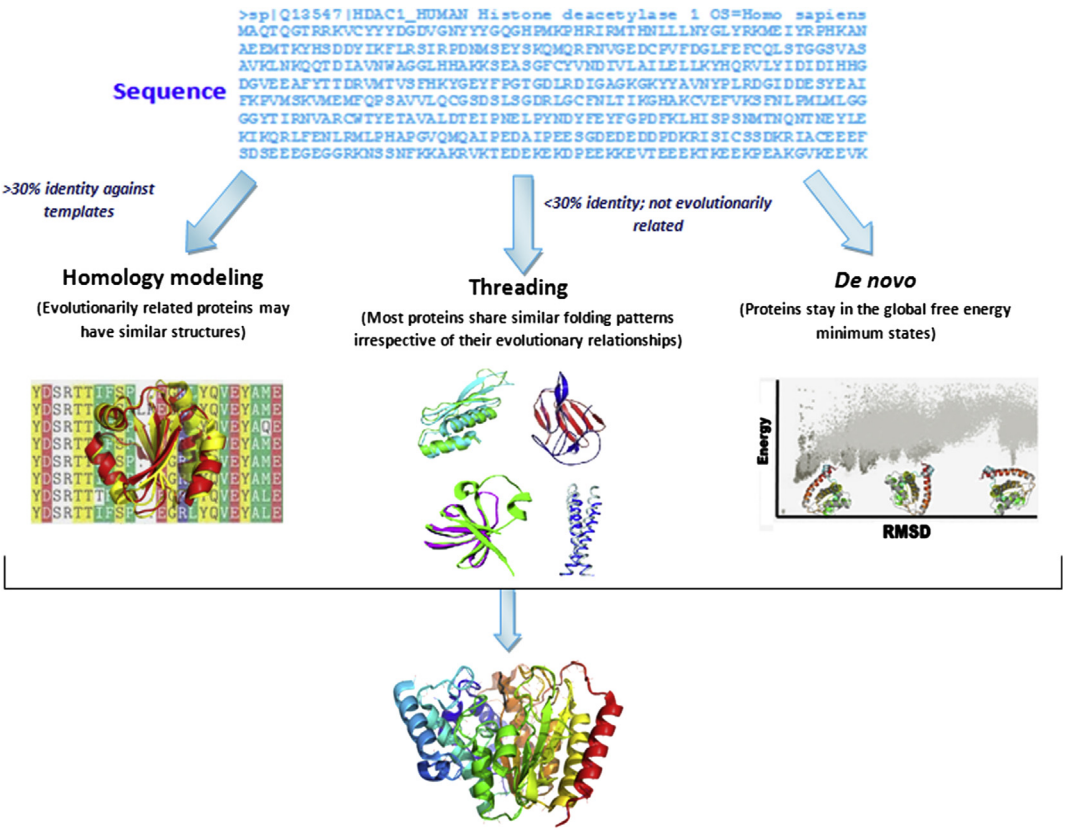

Fig. 1. Protein structure prediction methods.