ELSEVIER

Modelling three-dimensional protein structures for applications in drug design

Tobias Schmidt[1,2], Andreas Bergner[1,2,3] and Torsten Schwede[1,2]

[1] Biozentrum, University of Basel, Klingelbergstrasse 50-70, 4056 Basel, Switzerland
[2] SIB Swiss Institute of Bioinformatics, 4056 Basel, Switzerland

A structural perspective of drug target and anti-target proteins, and their molecular interactions with biologically active molecules, largely advances many areas of drug discovery, including target validation, hit and lead finding and lead optimisation. In the absence of experimental 3D structures, protein structure prediction often offers a suitable alternative to facilitate structure-based studies. This review outlines recent methodical advances in homology modelling, with a focus on those techniques that necessitate consideration of ligand binding. In this context, model quality estimation deserves special attention because the accuracy and reliability of different structure prediction techniques vary considerably, and the quality of a model ultimately determines its usefulness for structure-based drug discovery. Examples of G-protein-coupled receptors (GPCRs) and ADMET-related proteins were selected to illustrate recent progress and current limitations of protein structure prediction. Basic guidelines for good modelling practice are also provided.

Introduction

The goal of drug discovery is to contrive bioactive molecules that efficaciously modify a disease in a way that is beneficial to the patient, while keeping adverse effects such as toxic responses controllable. On the molecular level these requirements translate into a picture where a drug molecule binds to one or more target proteins that are implicated in the pathophysiology of a disease and act as, for example, inhibitors, agonists or modulators. At the same time, binding to proteins that have a negative impact on efficacy, or cause unwanted side effects, has to be avoided. In this sense, drug design is an enterprise that aims to engineer molecules with a controlled interaction profile against a multitude of different target and off-target proteins in an organism. During the initial target validation and hit finding phases of a drug discovery programme the focus is usually on the main target and then, as a programme progresses into lead optimisation, the attention shifts to the interplay of the drug candidate with an increasing number of proteins. Obviously, a full characterisation of these interactions down to the 3D structural details would constitute a profound structural perspective of the mode of action (MOA) of a drug molecule, and hence greatly facilitate drug design. Nowadays, a vast amount of experimental structural data, mainly generated by X-ray crystallography, is available [1]. Yet, the number of known protein sequences vastly exceeds the number of corresponding 3D structures. This so-called sequence–structure gap implies that for many important proteins there are no structures available. Fortunately, 3D protein structure prediction often offers an appropriate remedy in such situations [2,3]. In this review, we discuss the current status, applicability and limitations of protein models derived from protein structure prediction methods. We briefly introduce the prevailing prediction methods, with a focus on their relevance in drug discovery. Using selected examples, we also demonstrate typical applications at various stages of the drug discovery process.

Although it might appear trivial, it is worth emphasising one of the most important achievements of protein 3D structure modelling: the transformation, integration and contextualisation of heterogeneous information, such as mutation and SAR data, in a 3D model. Numerous visualisation tools have been developed for inspecting, analysing and annotating such models [4]. Visualisation of valid models is not merely a decorative offshoot of

Corresponding author: Schwede, T. (Torsten.Schwede@unibas.ch)
[3] Present address: Boehringer Ingelheim, Vienna, Austria.

1359-6446/06/$ - see front matter http://dx.doi.org/10.1016/j.drudis.2013.10.027

Please cite this article in press as: Schmidt, T. et al., Modelling three-dimensional protein structures for applications in drug design, Drug Discov Today (2013), http://dx.doi.org/10.1016/j.drudis.2013.10.027

modelling, rather a focal point where disparate facets of research efforts can amalgamate and converge into a detailed view of the underlying mechanistic basis, which in turn can become the driving force for further advances. It should be kept in mind that, even at relatively low resolution, 'any level of physical characterisation of a protein, as opposed to its absence, is valuable' [5].

Protein and binding site flexibility

Proteins are intrinsically dynamic systems that can exhibit significant flexibility and structural plasticity, also in their drug binding sites. A single structural model embodies only a static snapshot, regardless of whether it is an experimental or a predicted structure, and can therefore not always capture all the relevant characteristics of a protein. In essence, there is no entity such as 'the' structure of a protein, and this principle also applies to experimental structures for which, in addition to the issues associated with protein flexibility, experimental conditions, structural errors [6] and crystal packing effects [7] must be taken into account. Consideration of target and binding site flexibility is of paramount importance in computer-aided drug design (CADD), and disregarding them can dramatically hamper its success. Consequently, appropriate treatment of protein flexibility has become a major effort [8,9]. The ligand-steered modelling approaches outlined below have emerged as a result of these challenges.

Methods for protein modelling

Computational methods for predicting 3D protein models are widely used in the pharmaceutical industry, and much effort has been invested in improving model accuracy, and in expanding the scope of these methods (Table 1). Methods are generally categorised into template-based (i.e. homology) modelling and *de novo* modelling [10,11]. Traditional homology modelling (or comparative modelling) is considered to be the most accurate of these methods, and is thus most commonly applied in drug discovery research [12]. Homology modelling is based on the fundamental observation that all members of a protein family persistently exhibit the same fold, characterised by a core structure that is robust against sequence modifications [13]. It relies on experimentally determined structures of homologous proteins (templates), and enables the generation of models starting from given protein sequences (targets). The most accurate models can be obtained from close homologue structures; however, even with low sequence similarity (~20%) suitable models can be obtained [14,15].

A homology modelling pipeline generally comprises the following steps which can be repeated until a suitable model is obtained: (i) template selection for identifying the most suitable experimentally determined structures; (ii) target–template sequence alignment; (iii) 3D model structure building; (iv) model refinement; and (v) model quality estimation. Model refinement usually involves clash removal and geometrical regularisation of bond lengths and angles, but can also involve additional more sophisticated structural amendments. As a rule of thumb, most attention should be devoted to steps (i), (ii), (iii) and (v), whereas global model refinement (iv) typically has a disappointing return on investment [16].

LSM: ligand-steered modelling

As mentioned above, appropriate modelling of the binding site and correct ligand placement are of the utmost importance in CADD. However, native protein ligands such as enzyme substrates or signalling molecules often exhibit only weak binding affinities and are therefore often lost during purification procedures. As a result, protein structures are often determined experimentally in the absence of ligands. Additionally, template selection procedures in traditional homology modelling are often based on sequence similarity as the only criterion, neglecting ligand information in the template structures. As a consequence of this, the resulting protein models often represent an unliganded state of the binding site.

Classically, docking approaches have been used to place the ligands into the binding sites of the final homology models as a post-processing step [17–19]. The shortcomings of this practice have been addressed by developing more ligand-aware approaches that treat ligands as an integral part of a model throughout the entire modelling process. Generally, two strategies can currently be distinguished. First, ligand-guided (or steered) receptor modelling (LSM) directly incorporates ligands in the modelling process for guiding the protein conformation sampling procedure. One pioneering approach is binding site remodelling, which uses restraints obtained from initially modelled complex structures to build a second refined model [20]. Such approaches often require expert knowledge and time-consuming manual intervention, and hence call for the development of fully automatic homology modelling pipelines. Dalton and Jackson [21] have developed and assessed two variants of LSM, both yielding significantly more accurate complex models than docking into static homology models, regardless of whether or not the ligand had been incorporated into the modelling process. The most successful

TABLE 1

Frequently used servers and tools for protein structure homology modelling

Resource	Refs and URL	
Protein Model Portal	[94]	http://www.proteinmodelportal.org
HHpred	[95]	http://toolkit.tuebingen.mpg.de/hhpred
ICM	[96]	http://www.molsoft.com/
IntFOLD	[97]	http://www.reading.ac.uk/bioinf/IntFOLD/
Modeller, ModWeb	[98]	http://salilab.org/modeller/
Phyre2	[99]	http://www.sbg.bio.ic.ac.uk/phyre2/
Robetta	[100]	http://robetta.bakerlab.org/
SWISS-MODEL	[101]	http://swissmodel.expasy.org

Please cite this article in press as: Schmidt, T. et al., Modelling three-dimensional protein structures for applications in drug design, Drug Discov Today (2013), http://dx.doi.org/10.1016/j.drudis.2013.10.027