The International Journal of Biochemistry & Cell Biology 77 (2016) 264–274

Contents lists available at ScienceDirect

The International Journal of Biochemistry & Cell Biology

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

Structural modeling of G-protein coupled receptors: An overview on automatic web-servers

Mirko Busato[a], Alejandro Giorgetti[a,b,*]

[a] Department of Biotechnology, University of Verona, Strada le Grazie 15, 37134 Verona, Italy
[b] Computational Biomedicine, Institute for Advanced Simulation IAS-5 and Computational Biomedicine, Institute of Neuroscience and Medicine INM-9, Forschungszentrum Jülich, Germany

ARTICLE INFO

Article history:
Received 5 February 2016
Received in revised form 9 April 2016
Accepted 15 April 2016
Available online 19 April 2016

Keywords:
Homology modeling
G Protein-coupled receptor
Structural bioinformatics
Database
Web server

ABSTRACT

Despite the significant efforts and discoveries during the last few years in G protein-coupled receptor (GPCR) expression and crystallization, the receptors with known structures to date are limited only to a small fraction of human GPCRs. The lack of experimental three-dimensional structures of the receptors represents a strong limitation that hampers a deep understanding of their function. Computational techniques are thus a valid alternative strategy to model three-dimensional structures. Indeed, recent advances in the field, together with extraordinary developments in crystallography, in particular due to its ability to capture GPCRs in different activation states, have led to encouraging results in the generation of accurate models. This, prompted the community of modelers to render their methods publicly available through dedicated databases and web-servers. Here, we present an extensive overview on these services, focusing on their advantages, drawbacks and their role in successful applications. Future challenges in the field of GPCR modeling, such as the predictions of long loop regions and the modeling of receptor activation states are presented as well.

1. Introduction

G protein-coupled receptors (GPCRs) constitute the largest membrane-bound receptor family in the human genome (Schöneberg et al., 2004). GPCRs are the molecular messengers of the cell, and transduce stimuli from outside the cell into intracellular signals causing a signaling cascade. The about 800 human GPCRs can be grouped into five main sub-families named: glutamate, rhodopsin, adhesion, frizzled/taste2 and secretin, according to the GRAFS system (Fredriksson et al., 2003; Krishnan et al., 2012). The knowledge of the experimental three-dimensional (3D) structures of GPCRs is decisive to gain insights into the molecular mechanisms underlying their function. Unfortunately, membrane proteins are often unmanageable to crystallize. In this regard, it has to be noted that the first experimental 3D structure, the bovine rhodopsin, was released in 2000 (Palczewski et al., 2000) and only with the recent efforts, after the human β2-adrenoceptor solved in 2007 (Cherezov et al., 2007; Rasmussen et al., 2007), GPCR crystallography enabled the structural characterization of other receptor proteins and their interaction with ligands. Although the low sequence identity (SI) (<30%) between the members of the family, they share a similar core structure. The latter consists of seven transmembrane (TM) domains connected by six loops, three on the intracellular side (IL) and three on the extracellular side (EL) and an intra-cellular C-terminal loop and an extracellular N-terminal loop (Venkatakrishnan et al., 2013). So far, there are 136 (as of April 2016) rhodopsin-like GPCR X-ray structures solved, two secretin-like [corticotropin-releasing factor receptor 1 and glucagon receptor], four glutamate [the metabotropic glutamate receptor 1 and the metabotropic glutamate receptor 5], and five frizzled/taste2 [the smoothened receptor] (as reported in http://blanco.biomol.uci.edu/mpstruct). The data here reported comprise receptors in various functional states, stabilized by either a heterotrimeric G protein or a G protein-mimetic nanobody (Steyaert and Kobilka, 2011). For an excellent review on this topic,see: http://departments.agri.huji.ac.il/biochemfoodsci722/teachers/niv_masha/research.htm in the section "Summary of GPCR structures solved by X-ray crystallography (2015)".

For the cases in which the structure of a GPCR of interest has not yet been solved and structural insights are anyway needed for characterizing the interaction of the receptor with its cognate ligands, computational methods are a valid alternative strategy to

* Corresponding author at: Department of Biotechnology, University of Verona, Strada le Grazie 15, 37134 Verona, Italy.
E-mail addresses: mirko.busato@univr.it (M. Busato), alejandro.giorgetti@univr.it (A. Giorgetti).

http://dx.doi.org/10.1016/j.biocel.2016.04.004
1357-2725/

model their structures.[1] Recently, a very large number of computational models have been generated for GPCRs. In an outstanding work (Cavasotto and Palomba, 2015), identified and summarized at least twenty-five novel applications of GPCR homology models in the field of structure-based drug design. In the case of GPCRs, recent decoys of crystallized structures, mostly due to technological breakthroughs in membrane protein crystallization, opened to the development of more accurate comparative GPCRs models. As most of the recently modeled GPCRs were generated by using homology modeling techniques, we dedicated the next section to an extensive description of this methodology.

Homology or comparative modeling methods are based on the idea that evolutionary related proteins share a similar structure (Chotia and Lesk, 1986). The quality of the predicted models is thus correlated with the evolutionary distance between the template (member of the family with known structure) and the target proteins (Tramontano et al., 2008). Homology model techniques are the most reliable and accurate methods to generate structures (Meier and Söding, 2015; Tramontano et al., 2008). It is commonly accepted that this procedure involves the following steps (Fig. 1): (1) identification of proteins evolutionary related to the target sequence, whose 3D structures are solved (templates); (2) target-template sequence alignment; (3) generation of an initial model for the target based on the template selected and the sequence alignment; (4) refinement of the model at the side chain, loops and backbone level by using molecular dynamics (MD) simulations and incorporation of additional experimental structural information; (5) validation and evaluation of the model. These steps must be iterated to achieve an acceptable model.[2]

Broadly speaking, comparative modeling can be considered as a cost-effective alternative tool when experimental structures are absent. Although homology modeling can boast a number of successes in many applications, it should be noted that the generated models are predictions and could present some inaccuracies. Homology modeling methods depend strongly on both the SI between the target and the templates and the accuracy of alignment. SI < 50% generally leads to structural divergence between the models and the actual experimental structure, measured as Cα atom RMSD, larger than 1 Å (Chotia and Lesk, 1986). Actually, two proteins with SI > 35% were shown to share the same fold (Orengo et al., 1997). Finally, with low target-template SI (between 10% and 25%), the comparative models might contain serious errors, thus, it is strongly recommended the introduction of experimental information such as ligand information, site-directed mutagenesis, and other experimental restraints to improve the accuracy of the model (Yarnitzky et al. 2010). These very general observations are often sustainable but can vary depending on the protein of interest. Indeed, it is well-established that some "superfolds" dominate the fold space. This implies that even very distant proteins, *e.g.* GPCR proteins, can share a similar 3D structure (Magner et al., 2015).

The limitations/errors in homology modeling can be grouped into five categories (Fiser, 2010; Palomba and Cavasotto, 2015): (1) Errors in side-chains modeling. These errors can introduce drastic changes in the side-chains involved in the ligand binding (Rodrigues et al., 2013). The latter highlights the importance of using existing ligand information in the homology modeling protocol (Yarnitzky et al., 2010). (2) Structural Deviations of a target region that is aligned correctly with the template. Even if the aligned segments of the model are correct; the target protein could present local structural differences than the template structure, indeed these divergences could be due to artifacts in structure determination of the template in different environments, and not to errors in the alignment. Therefore, models might be improved by using unrestrained MD simulation in order to arrange the misfolded regions (Schlegel et al., 2005) or by using multiple-template approaches (Srinivan and Blundell, 1993). (3) Inaccuracies of target regions that do not have an equivalent segment in the template structures. Indels are an important issue in the homology modeling procedure. For instance, residues in incorrect positions or missing residue in the comparative models could be due to gapped residues in the core region. (4) Distortions or shifts of a region that is aligned incorrectly with the template structures. The quality of the alignment is one the major problems in homology modeling, especially when the SI falls below 20%. Misaligned regions correspond to errors in positioning the target residues on the template fold; resulting in an unreliable model. Multiple sequence alignment (MSA) and hidden Markov models (HMM) profiles approaches combined with manual inspection and curation of the alignments are strongly recommended to investigate possible errors and adjust key motifs in the alignment. (5) Templates are fragmentary or incorrect. The misfolded structure resulting from using an erroneous template or fragmentary template is the problem typically arises when SI is below 30%. Moreover, for large protein (>200 residues) the number of full-length templates is still sparse comparing with the many fragments of full-length proteins.

Other modeling techniques offered by bioinformatics, *i.e.* the *Non Homology-based methods,* include:

- *Ab-initio modeling techniques*: When no appropriate homologous receptors with known structure (templates) can be found, *ab initio* methods can be used to predict the protein structure just from the sequence. Hence, topology-based techniques have shown great advances in the GPCR field. The seven TM α–helices are built up and then folded together to form the common scaffold (typical of all GPCRs) by using the physical forces acting on the atoms of the protein. This approach was the first used to generate GPCR models based on bacteriorhodopsin and rhodopsin structures with MembStruck (Shacham et al., 2001; Vaidehi et al., 2002).
- *Fold recognition methods or threading*: This technique incrementally replaces the sequence of a known 3D protein structure with a target sequence of unknown structure. The generated model is then evaluated using a scoring function that measure the fitness between the target and a given fold. The process is repeated against all known 3D structures until an optimal fit is found. Then, iterative structural assembly simulations are carried out to generate the full-length 3D model. Eventually, the target protein function is estimated by matching the predicted structures with already solved structures.

The need of GPCR structure predictions for gaining insights into the determinants underlying their biology and pharmacology, has paved the way to the development of tools mainly devoted to GPCR structure prediction. In the last years, a great variety of successful modeling methodologies have been automatized and rendered available to the entire scientific community through web-services.

This review has the principal aim of offering an overview on the automated web servers dedicated to the modeling of GPCR structures and their related models' databases. These user-friendly tools can help researchers to overcome the difficulties of setting up the

[1] Several features often help in the modeling procedure, *i.e.* remarkably, in the rhodopsine-subfamiliy, there is at least one highly conserved residue in each of the TM helices (Mirzadegan et al., 2003). This, was used to create the Ballesteros and Weinstein scheme (Ballesteros and Weinstein, 1995) that was recently generalized "generalized scheme" in (Isberg et al., 2016). By taking into account the bulges and constrictions, it ensures that the residues aligned in sequence are those that align in structure.

[2] Historically, several software have been developed for carrying out the modeling procedure (Dolan et al., 2012; Martí-Renom et al., 2000; Nayeem et al., 2006). Some of the most famous programs that automate the homology modeling protocol are listed in (Supplementary material: SM Table 1).