The International Journal of Biochemistry & Cell Biology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

The International Journal of Biochemistry & Cell Biology

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

Review

Molecular modelling approaches for cystic fibrosis transmembrane conductance regulator studies☆

Norbert Odolczyk[a], Piotr Zielenkiewicz[a,b,*]

[a] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warszawa, Poland
[b] Faculty of Biology, Warsaw University, 02-106 Warszawa, Poland

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form 1 April 2014
Accepted 4 April 2014
Available online xxx

Keywords:
Cystic fibrosis
CFTR
Molecular modelling
In silico studies

ABSTRACT

Cystic fibrosis (CF) is one of the most common genetic disorders, caused by loss of function mutations in the gene encoding the CF transmembrane conductance regulator (CFTR) protein. CFTR is a member of ATP-binding cassette (ABC) transporters superfamily and functions as an ATP-gated anion channel. This review summarises the vast majority of the efforts which utilised molecular modelling approaches to gain insight into the various aspects of CFTR protein, related to its structure, dynamic properties, function and interactions with other protein partners, or drug-like compounds, with emphasis to its relation to CF disease.

This article is part of a Directed Issue entitled: Cystic Fibrosis: From o-mics to cell biology, physiology, and therapeutic advances.

Contents

1. Introduction

Cystic fibrosis (CF) is one of the most common genetic disorders and is caused by loss-of-function mutations in the *CFTR* gene, which encodes the CF transmembrane conductance regulator (CFTR) protein (Riordan et al., 1989). CFTR, a PKA-activated Cl^- channel, is a rate-limiting factor for fluid absorption in numerous epithelia (Robert et al., 2008).

Molecular modelling techniques have made significant advances in recent years due to several reasons: (1) increasing knowledge and better understanding of life phenomena in atomic scale; (2) development of more sophisticate and accurate algorithms with combined integrative strategies; (3) continuous increase of the computational power. All above aspects contributed to the fact, that molecular modelling tools are becoming essential

☆ This article is part of a Directed Issue entitled: Cystic Fibrosis: From o-mics to cell biology, physiology, and therapeutic advances.

* Corresponding author at: Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5a, 02-106 Warszawa, Poland.
Tel.: +48 225922145; fax: +48 226584636.
E-mail address: piotr@ibb.waw.pl (P. Zielenkiewicz).

http://dx.doi.org/10.1016/j.biocel.2014.04.004
1357-2725/

components of biological research, and are widely used in the simulation of biomolecular systems.

The many biophysical and biochemical processes that have been successfully investigated by molecular modelling tools include protein folding, protein-protein and protein-DNA recognition, conformational changes of macromolecules, and enzyme catalysis.

In this review, we would like to summarise most of the efforts that have been using molecular modelling approaches to gain insight into the various structural aspects of the CFTR protein and its relationship to CF disease.

2. Insights into CFTR structure

Protein structure is important due to its intimate connection with protein function (Laskowski et al., 2005). Thus, it is extremely difficult to fully understand the importance of a particular protein in biological phenomena without determining its three-dimensional structure. Since the first isolation, cloning and characterisation of the *CFTR* gene, there has been an incessant endeavour to characterise the structure of its protein product (Dalton et al., 2012). Experimental techniques, such as protein crystallography (Hess and Rupley, 1971) and NMR spectroscopy (Knowles, 1972), are invaluable for protein tertiary structure determination (Berman et al., 2000), but they also have limitations (Acharya and Lloyd, 2005, Marion, 2013), especially in the context of CFTR protein, which is high molecular mass membrane protein, difficult to express at high levels, purify, and reconstitute in a functional form. Despite significant progress in the X-ray crystallography of membrane proteins (Kang et al., 2013), there is still limited atomic-resolution information on the full-length CFTR channel. All current knowledge of the spatial organisation of CFTR has been derived from the following mixed approaches: X-ray crystallography of individual domains NBD1 (Lewis et al., 2004, 2005, 2010; Thibodeau et al., 2005; Atwell et al., 2010; Mendoza et al., 2012), NBD2 (Zhao et al., unpubl., PDB ID: 3GD7), and homodimer of NBD1s (Atwell et al., 2010); NMR studies on NBD1 with or without RD (Baker et al., 2007; Kanelis et al., 2010; Hudson et al., 2012); low-resolution experiments of full length CFTR structure (Awayn et al., 2005; Ford et al., 2011; Mio et al., 2008; Rosenberg et al., 2004; Rosenberg et al., 2011; Zhang et al., 2009, 2011), SAXS studies of NBDs hetero dimer (Galeno et al., 2011; Galfre et al., 2012) or RD (Marasini et al., 2013), as also various sequence analyses, and molecular modelling approaches. The latter techniques have been intensively developed in recent years to provide alternative methods to time-consuming experimental procedures. The theoretical methods for prediction of 3D-protein structures include first principle methods (*de novo*), fold recognition and homology (or comparative) modelling, which is currently the most accurate and, therefore, the most widely used approach for protein structure prediction (Dahl and Sylte, 2005). Homology modelling is based on the empirical observation that evolutionarily related proteins tend to have similar three-dimensional (3D) structures (Al-Lazikani et al., 2001).

The first insights into the general topology of CFTR came from simple evolutionary sequence analysis of putative protein sequence, defining its similarities with protein products of homologous genes, and calculating hydropathy profile (Riordan et al., 1989). CFTR shares its evolved domain organisation with other members of the ATP-binding cassette (ABC) transporter protein superfamily (Hyde et al., 1990). Its 1480 amino acid long polypeptide chain consists of two nucleotide-binding domains (NBD1 and NBD2) and two transmembrane domains (TMD1 and TMD2), which are arranged alternately and are separated into two symmetrical fragments by a regulatory domain (RD) (Riordan, 2008) (Fig. 1). The RD is unique among the superfamily members and is responsible for the regulation of CFTR activity (Seibert et al., 1999; Ostedgaard et al., 2001).

2.1. Modelling of individual domains

Because the crystal structures of human NBD1 and NBD2 domains are already known, this paragraph is presented here mostly for historical reasons, rather than knowledgeable aspects. The pioneering efforts to propose a 3D-structural model of the cytoplasmic domains were undertaken in the early 1990s. However, due to the lack of an appropriate template crystal structure, the models were constructed on non-homologous proteins using less reliable approaches based on fold recognition methods (Godzik, 2003; Peng and Xu, 2010). The prediction and analysis of secondary structure by different algorithms inclined two groups to select the structure of adenylate kinase as a template for NBD1 modelling (Hyde et al., 1990; Mimura et al., 1991), the threading approach resulted in constructing the NBD1 model based on aspartate aminotransferase (Hoedemaeker et al., 1998). According to the sequence similarity, others decided to used as a template bovine heart mitochondrial (Annereau et al., 1997) or rat liver mitochondrial F1-ATPase (Bianchet et al., 1997) structures. Bianchet et al. (1997) also proposed the structural model of the NBD2 domain as also the NBD1–NBD2 heterodimer. Unfortunately, all of the above template structures came from evolutionarily unrelated proteins, and the models had little or no information value.

The only one study, which attempted to characterise the structural properties of the isolated R domain by *in silico* approaches, has employed discrete MD (DMD) simulations (Dokholyan et al., 1998, Dokholyan et al., 2000) and the all-atom force field Medusa (Ding and Dokholyan, 2006) to generate an ensemble of 3D-structures of the R domain at the atomic level (Hegedus et al., 2008). However the recently published experimental measurements contradicted the above theoretical studies (Marasini et al., 2013).

2.2. Assembly of domains

Constructing models of a membrane integral protein composed of several domains is a challenging task and usually requires different approaches for cytosolic and membrane spanning fragments as well as different templates for each domain (Frishman, 2010). Moreover, the templates often come from crystals of individual domains in which the native domain-domain interaction interfaces are unsettled and at least require refinement of side chain conformations to correctly predict a relative domain position to each other (Fernandez-Fuentes et al., 2007; Wollacott et al., 2007).

2.2.1. NBDs heterodimer

The first reliable homology model proposed for the NBD1–NBD2 dimer (Callebaut et al., 2004; Eudes et al., 2005) was constructed on a template of the experimentally resolved dimeric structure of the bacterial ABC transporter, MJ0796 (Smith et al., 2002). This model was supported by evolutionary information gained from other ABC-like domains, such as BtuCD (Locher et al., 2002), HisP (Hung et al., 1998), MJ1267 (Yuan et al., 2003), TAP1 (Gaudet and Wiley, 2001) and MalK (Chen et al., 2003), as well as by hydrophobic cluster analysis (HCA) (Gaboriaud et al., 1987; Callebaut et al., 1997). The model provided the following important structural insights: (a) the heterodimer formation and its "head-to-tail" orientation; (b) the location of both nucleotide-binding sites on the dimer interface; (c) description of the important contribution of residues from both subunits into each active site; (d) and precisely identified the spatial organisation of functional sequence motifs with their characteristically asymmetric features. Indeed, the nucleotide-binding sites of NBD1 and NBD2 are differentiated by non-canonical residues in Walker-B (Ser instead of Glu) and by switch (Ser instead of His) motifs for NBD1 as well as by a signature sequence in NBD2 (LSHGH instead of LSGGQ) (Gadsby et al., 2006). Such asymmetry reflects on NBD1 catalytic activity, which, in contrast to NBD2, is not able to

Please cite this article in press as: Odolczyk N, Zielenkiewicz P. Molecular modelling approaches for cystic fibrosis transmembrane conductance regulator studies. Int J Biochem Cell Biol (2014), http://dx.doi.org/10.1016/j.biocel.2014.04.004