

Knowledge Based Membrane Protein Structure Prediction: From X-Ray Crystallography to Bioinformatics and Back to Molecular Biology

Alejandro Giorgetti^{1,2} and Stefano Piccoli¹

¹*Applied Bioinformatics Group, Dept. of Biotechnology, University of Verona,*

²*German Research School for Simulation Sciences, Jülich Research Center and RWTH-Aachen University, Jülich*

¹*Italy*

²*Germany*

1. Introduction

Life, or at least as we know it, would not exist without the ability of living organisms to communicate with their surroundings and respond to changes within them. Cells are able to capture and decode environmental stimuli into biologically signals. Indeed, communicating mechanisms, able to detect stimuli coming from the outside world are of fundamental importance for the survival of living beings. A deep understanding of the molecular mechanisms underlying signal transduction is thus needed for a complete characterization of the way our cells communicate with the rest of the world.

Living cells are surrounded by a plasma membrane that forms a boundary between the cell interior and the external physical world. As a consequence, the cellular plasma membrane presents a major target for environmental stimuli acting upon a living cell. The membrane contains protein molecules that confer various functions on it.

Integral membrane proteins play a key role in detecting and conveying outside signals into cells, allowing them to interact and respond to their environment in a specific manner. They are involved as main players in several signaling pathways and therefore, the majority of drug targets are associated to the cell's membrane. The original human genome sequence project estimated 20% of the total gene count of 31,778 genes to code for membrane proteins (Almen et al., 2009). Thus membrane proteins constitute a very large set of yet-to be-characterized proteins mediating all the relevant life-related functions both in prokaryotes and eukaryotes. Moreover, the total amount of membrane proteins for which the three-dimensional structure is known, is just about 842, corresponding to 298 unique proteins, as included in the Membrane Proteins with Known Structure database (<http://blanco.biomol.uci.edu/mpstruc/listAll/list>).

This extremely low number of membrane proteins with known structure is due to the fact that membrane proteins are very difficult to study because they are inserted into lipid bilayers surrounding the cell and its sub-compartments. In these conditions they expose to the polar outer and inner environments portions of different sizes, completely changing the biophysics with respect to soluble proteins. Thus, when isolated from membranes, membrane proteins are

generally less stable than globular ones. It is therefore difficult to purify them in the native, functional form, and more difficult to crystallize them. Thus, crystallization of this type of proteins is yet a very difficult process, given the fact that they expose two different chemical-physical surfaces to the environment: water- and lipid-like. On the other hand, the lipid environment constraints the membrane protein stable folding: it is indeed evident from the structures that have been deposited so far that only all-alpha and beta-barrel structural organizations are present in nature. Indeed most of membrane proteins in the Protein Data Bank (PDB: <http://www.pdb.org>), i.e. 67%, consist of bundles of transmembrane helices with different tilting with respect to the membrane plane and to each other.

In the last few years, several efforts have been carried out for the determination of new crystal structures of membrane proteins. Although the improvements in the technologies allowed the determination of several structures, the gap between the known sequences and the solved structures is still enormous. Furthermore, albeit their extreme importance, crystal structure only provide a static image of a protein under conditions that sometimes are far from being physiological.

Thus, the combination of existing crystal structures, computational biology techniques and molecular biology validating experiments, may be the key to face the challenges of bridging the gap between the characterized membrane proteins with and without solved structure.

This and other issues may be resolved in the post-genomic era. To do this we should take advantage of all the theoretical efforts aiming at developing tools based on our present knowledge that are capable of extracting selected structural/functional features from known sequences/structures and of computing the likelihood of their presence in never-seen before sequences/structures. Moreover, once generated, the models can be considered as hypothesis to be tested. Thus, it is of fundamental importance to accurately validate the models and the conclusions outlined from their analysis.

Here we will review some of the efforts of the last years aimed at the characterization at the structural level of different membrane proteins for which the crystal structure is not known.

We will introduce state-of-art modelling techniques that use the most recent membrane proteins crystal structures as input for the modelling of membrane proteins in different activation states. For each of the studied cases we will review also the experiments carried out in order to validate the proposed hypothesis.

2. Protein structural bioinformatics

Proteins, as defined in the glossary of Molecular Biology of the cell, are just 'the major macromolecular constituent of cells. A linear polymer of amino acids linked together by peptide bonds in a specific sequence' (Alberts et al., 2002). The latter definition of a protein, albeit technically correct, lacks of the most important part of the definition, that is, proteins are the product of evolution. In fact, proteins are indeed polypeptides, but formed following a very precise relation of sequence-structure-function, modelled by events of random variation and natural selection along millions of years of evolution, that introduced small changes in the protein sequences passing from one generation to the other. The structure/fold of a protein is then determined by its amino acid sequence and, therefore, upon mutation of one amino acid with another may affect the structure in different ways. Taking into account the folded-protein energy landscape, it was observed that, although summing up big numbers coming from a multitude of weak interactions between their atoms, functional proteins have a limited stability, equivalent to just a few hydrogen bonds

(H-bonds). This suggests that this extremely weak folding energy landscape might be easily altered by mutations, giving place to unfolded proteins. However, as all the living beings that we know today are indeed 'alive', it means that during evolution, the structure-function relationship has been preserved, so all the proteins that we observe today can only contain mutations that did not alter too much the global folding energy landscape with respect to their ancestor sequence. From those considerations it was concluded that evolutionarily related proteins, that diverged from a common ancestor via the accumulation of small changes, cannot but have similar structure, where mutations have been accommodated only causing small local rearrangements. If the number of changes, i.e. the evolutionary distance, is high, these local rearrangements can cumulatively affect the protein structure and produce relevant distortions, but the general architecture, i.e. the fold, of the protein has to be conserved. On the other hand, if two proteins have evolved from a common ancestor, it is likely that a sufficient proportion of their sequences has remained unchanged so that an evolutionary relationship can be deduced by their comparative analysis. Therefore if we can ensure that two proteins are homologous, that is evolutionary related, the structure of one can be used as a template for the building up of the structure of the other. This forms the basis of the technique known as comparative or homology modeling (Tramontano, 2006).

In this regard, 25 years ago, Chothia and Lesk, in a seminal article (Chothia & Lesk, 1986) have aligned the sequences and structures of all the proteins with known structure, finding a correlation between the evolutionary distance and structural divergence between evolutionary related proteins. This work settled up the basis of the comparative (or homology) modeling technique, a method that allows the prediction of protein structures using as a template a member of the family for which the three-dimensional structure (3D-structure) is known. So, if we assume that the sequence alignment between two protein sequences, one of unknown (the target) and one of known (the template) structure, reflects the evolutionary relationship between their amino acids, we can assume that most of them have conserved the same relative position in the structure and use the coordinates of the backbone of the template as first approximations of the coordinates of the backbone of the target (Tramontano, 2006). We must then model the conformation of the side chains and the local rearrangements of the structure brought about by the amino acid substitutions (Tramontano, 2006). Templates for a comparative model would be found among the structures present in databases, i.e. PDB and can be extracted by searching the database for proteins putatively homologous to the target protein.

All these ideas were shown to be valid through the years for soluble proteins. Indeed, the comparative modeling technique has not always been considered valid when applied to membrane proteins (Floriano et al., 2006). The main criticism regarded the low amount of membrane proteins with known three-dimensional structure and the enormous evolutionary gap that must be filled in order to produce models of the most studied membrane proteins. However, in the last few years X-ray crystallography reached very high levels of applicability in the field of membrane proteins and a lot of new and more refined structures were solved by several groups. In the applications section we will review some of the most relevant structures and how they allowed a better characterization of the mechanisms underlying the function. Indeed, this great advancement in the crystallographic techniques produced as of today 298 structures of unique membrane proteins for a total of 842 structures. At this point, we have checked for the existence of correlation between evolutionary relationship and structure similarity. We have recently (not published)

followed the same protocol as Chothia and Lesk (Chothia & Lesk, 1986) but considering just membrane proteins. Indeed, by the use of the LGA server (<http://proteinmodel.org/>) we have aligned (the structures and sequences) present in the core of all the membrane proteins with known three-dimensional structure and produced the graph of Figure 1. The structural divergence between two evolutionary correlated proteins is measured as their Root Mean Square Deviation (RMSD).

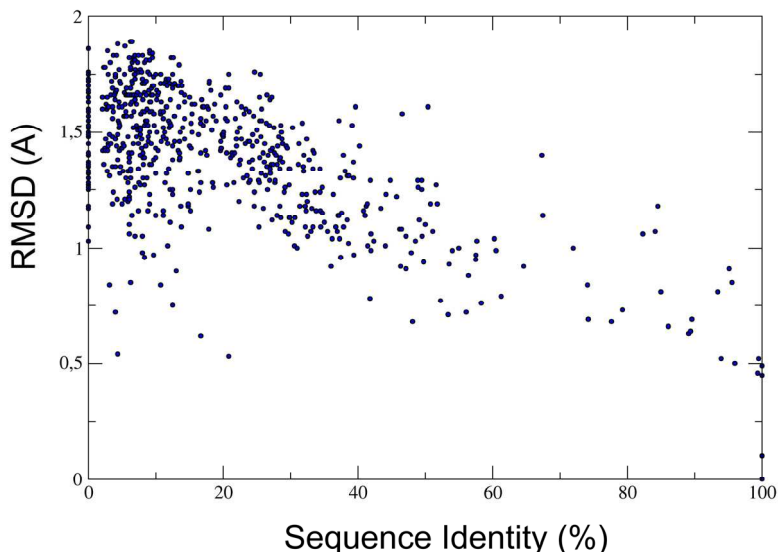


Fig. 1. RMSD versus Percent Sequence Identity of membrane proteins. The core of all the membrane proteins found in the Membrane Proteins with Known Structure Database (<http://blanco.biomol.uci.edu/mpstruc/listAll/list>) was aligned at the structural level and at the sequence level.

Figure 1 shows the RMSD of the backbone of the core of pairs of evolutionarily related proteins as a function of the percent of identity between their amino acid sequences. The definition of the “core” of the structure differs in different methods. It can be intuitively seen as the internal, closely packed, evolutionary conserved part of the structure that contains most of the repetitive secondary structure elements (Tramontano, 2006). For practical purposes we considered as the core of the proteins, all the amino acids present in secondary structure elements and those regions not diverging for more than 3 Å, as Chothia and Lesk did (Chothia & Lesk, 1986).

As stated before, Comparative Modeling is based on the idea that evolutionarily correlated proteins share similar three-dimensional structures. That is, if we want to predict the structure of a protein we can look in database for an evolutionarily correlated protein with known structure, and use the latter as template for building up the structural model of our preferred protein (Tramontano, 2006). The important thing is that, based on Figure 1 the procedure is valid also for membrane proteins. In this regard, the astonishing improvements in membrane proteins crystallography together with comparative modeling techniques will allow the characterization of an enormous amount of membrane protein in the near future.

3. Applications

In the following sections we will present two cases in which homology modelling has been applied on membrane proteins, i.e. ion channels and G-protein-coupled receptors. Both cases are representative icons of the difficulties found in the structure solving and modelling of membrane proteins for many years. Fortunately, in the last few years there was an explosion of newly solved crystal structures that completely revolutionised the field. Indeed, several mechanisms were understood and functional features could be extended to several members of the families.

We will review in both cases the advancements in X-ray crystallography and how we have used the recently solved crystal structures combined with homology modelling and molecular biology experiments to characterize functional mechanisms.

3.1 Ion channels

Ion channels are integral membrane proteins that function as molecular sensors of physical and chemical stimuli and convert these stimuli into biological signals vital for the existence of every living organism. In other words, ion channels represent the doors and windows of the cell, that open and close following precise stimuli and leave the entrance/exit of very accurately selected 'visitors'. As molecular transducers of mechanical, electrical, chemical, thermal or electromagnetic (light) stimuli, ion channels contribute to changes in electrical, chemical or osmotic activity within cells by gating between the two basic conformations in which they exist - open and closed. Through the gating mechanisms, i.e. opening and closing, ion channels regulate the permeation of ions (in some cases also other solutes), allowing ions to cross the hydrophobic core of the cell membrane, affecting its activity. Because of the well-known difficulties in obtaining high resolution 3D structures by X-ray crystallography of ion channels, alternative strategies based on computational biology tools are currently used to investigate their biophysical properties (for a review about ion channel modelling see: Giorgetti & Carloni, 2003).

The last two decades have been exceptionally exciting for research in the field of ion channels. Astonishing progress has resulted from the use of multidisciplinary approaches to gain insight into the structure and function of ion channels and their role in various aspects of cell physiology and signal transduction. Molecular biology and genetics have provided the sequences of a very large number of ion channel proteins and have helped identify their contribution to various cellular functions. The patch clamp technique has provided the means to study the functional properties of single ion channels with unprecedented precision. X-ray and electron crystallography have provided structural snapshots of a number of ion channel molecules at near atomic resolution, whereas magnetic resonance spectroscopy and fluorescence spectroscopy have provided means to access the dynamics of these molecules. In detail, as of today we count with 14 unique ion channel structures (<http://blanco.biomol.uci.edu/mpstruc/listAll/list>) for a total of about 45 crystal structures of ion channels solved in different activation states and co-crystallized with different ligands and ions. Moreover, more than 2/3 of the solved structures were obtained in the last two years, showing an exponential development of the field.

Using the structural and functional information obtained by these experimental techniques, computer-assisted molecular modeling has brought ion channels to life by allowing the features underlying the molecular events that shape their function. Indeed, the

multidisciplinary approach to the study of ion channels has yielded an unprecedented wealth of new data.

In the next sections examples of applications on a specific ion channel will be illustrated.

3.1.1 CNG channels

Cyclic nucleotide-gated (CNG) channels are ion channels that generally express in several sensory and non-sensory cells (Kaupp & Seifert, 2002; Matulef & Zagotta, 2003). The most characterized members of the family are those involved in sensory transduction in vertebrate photoreceptors and in olfactory sensory neurons (Kaupp & Seifert, 2002; Matulef & Zagotta, 2003). In their native forms, CNG channels are heterotetramers (Kaupp & Seifert, 2002; Matulef & Zagotta, 2003). CNG channels are differentially sensitive to cyclic nucleotides (CNs): CNG channels from vertebrate rod photoreceptors are selectively activated by cGMP binding, whereas CNG from olfactory neurons are activated by either cAMP or cGMP binding.

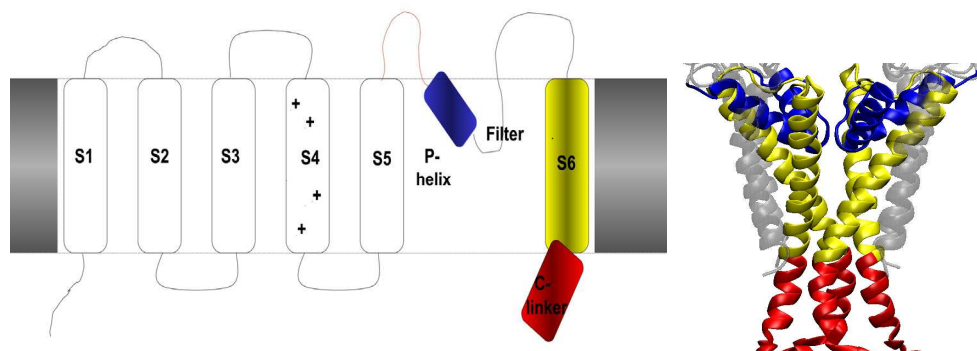


Fig. 2. Left panel: Topology of CNG channels. The plot shows the topological relevant elements of CNG channels analyzed in the main text. In particular P-helix was depicted in blue while the S6 helix and the C-linker, are shown in yellow and red, respectively. The C-linker domain, connects the transmembrane domain with the cyclic nucleotide binding domain. S4 is the positively-charged voltage sensor of the channel. The conformational changes proposed regards C-linker, S6 transmembrane helix and the filter P-helix. Right panel: the figure shows the three dimensional configuration of the tetrameric filter region of CNG channels, as obtained through the modeling procedure.

Several very important groups worked for years, accumulating electrophysiological data for mutants and wild-type CNG channels in different activation states, in order to characterize the biophysical properties underlying the functioning of these channels. In particular, most of the experimental available information regard electrophysiological investigations with mutant channels, performed on the CNGA1 channels from bovine rods (Craven & Zagotta, 2006; Kaupp & Seifert, 2002; Matulef & Zagotta, 2003), a channel with a primary structure composed of 690 residues (Kaupp et al. 1989). Although the analysis of accessibility, based on Cysteine Scanning Mutagenesis (CSM) (Akabas et al., 1992; Karlin & Akabas, 1998) has shown that CNG and potassium channels share the same gross topology (Bechetti et al. 1999; Flynn & Zagotta, 2001; Giorgetti et al., 2005; Liu & Siegelbaum, 2000; Matulef & Zagotta, 2003; Mazzolini et al., 2009; Nair et al., 2009), these two families of channels have

different functional properties. Indeed, while voltage-gated potassium channels are extremely selective (MacKinnon, 2003) and their gating strongly depends on membrane voltage (Bezanilla, 2005; Bezanilla, 2008; Swartz, 2004; Swartz, 2008), CNG channels have a low ion selectivity and their gating is only poorly voltage dependent (Kaupp & Seifert, 2002).

Here, we review some of the recent years results of an extensive combined experimental/computational structural study on the widely characterized homotetrameric cyclic nucleotide-gated channel (CNG), from *bovine rod*. As stated before, CNG channels are tetrameric and each subunit consists of two domains arranged as shown in Figure 2: (i) a transmembrane domain formed by six transmembrane helices (S1–S6) and a pore helix (P-helix) with the same topology of voltage-gated potassium channels (Becchetti et al., 1999; Sesti et al., 1995); (ii) A cytoplasmic domain formed by the cyclic nucleotide binding domain (CNBD) which is linked to the transmembrane domain through the so called C-linker region. The pore, un-selective for sodium and potassium, is believed to gate via a conformational change of S6 transmembrane helix (TMH) initiated by the binding of cyclic nucleotides to the binding domains. This conformational change is then transmitted to the pore via coupling with the four P-helices (Johnson & Zagotta, 2001; Matulef et al., 1999). During the last 5 years, we and our collaborators have provided the molecular basis for the characterization of the mechanisms underlying the functioning of these channels by constructing homology models of the transmembrane region of the CNGA1 channel. The models include the S6, P-helix-loop (P-helix+pore wall or filter), along with the C-linker N-terminal sections. Indeed, all the modeled regions were extensively characterized by a great amount of experimental data.

Models of P-helix-loop and S6 are based on the KcsA X-ray structure, whose topology has been suggested to be similar to that of CNG channels (Becchetti et al., 1999). On the other hand, the C-linker domain was modeled using the C-linker of the *mouse* Hyperpolarization-activated and Cyclic-nucleotide-modulated (mHCN) channel in its ligand bound state, for which the X-ray structure has been recently solved (Zagotta et al., 2003).

The homology models, were then refined by the inclusion of an extensive dataset of spatial constraints inferred by electrophysiological measurements on cysteine mutants. A large set (about 50) of structural constraints among C α atoms were inferred from measurements of the electrophysiological properties of the channel in the presence of metal ions (Becchetti et al., 1999; Becchetti & Roncaglia, 2000; Flynn & Zagotta, 2001; Johnson & Zagotta, 2001; Liu & Siegelbaum, 2000; Mazzolini et al., 2010; Nair et al., 2009). For example (i) Cadmium, which can block the channel when it binds to, at least, two cysteine residues. (ii) The mild-oxidizing agent copper phenanthroline (CuP) favors disulfide bridge formation between two cysteines separated by a distance going from 6 to 11 Å. The electrophysiological results were then converted into distance constraints by a statistical analysis of the PDB. Indeed, we have looked for all the proteins co-crystallized with cadmium atoms and we have extracted all the mean distances from cysteines bound to the cadmium atoms. Then, taking into account that the channels are homotetramers, the reversible/irreversible condition of the cadmium blockage was converted into distance restraints by just geometrical considerations. These and other agents were included in the solutions to characterize different features of the channel mechanisms (Mazzolini et al., 2010; Nair et al., 2009). The procedure followed by us and by our experimental collaborators, consisted in a series of iterative steps that extended for more than three years of successive cycles of modeling followed by experiments and vice versa.

This experiment-guided computational model allowed us to gain insights into the structural basis of CNG channel gating mechanism. We have suggested several mechanical features underlying channel functioning. For example we have hypothesized the bending and the counterclockwise rotation of the C-linker N-terminal section. Indeed, this motion is suggested to be transmitted upwards to cause the upper part of S6 to rotate counterclockwise producing the conformational changes needed for the opening and closing of the filter region (Giorgetti et al., 2005). The procedure that allowed to unravel the functional gating mechanisms was characterized by an iterative theoretical/experimental work that permitted not only the hypotheses generation but also their experimental validation. Indeed, on the basis of our models, and using cysteine-scanning mutagenesis, our experimental collaborators were also able to show that in the presence of a mild oxidizing agent, copper phenanthroline (CuP), certain cysteine mutations were able to lock the channel in either the closed or open state, depending on whatever state they happened to be in at the time of CuP application (Nair et al., 2006).

This kind of work has very few precedents due to the general difficulties found in the expression of different mutants. Indeed, some of the suggested mutants included double and triple mutations for a single protein.

3.2 GPCRs

G protein-coupled receptors (GPCRs), or 7 transmembrane helix receptors (Figure 3), are membrane embedded proteins, responsible for the communication between the cell and the environment (Sakmar et al. 2002). Malfunction of these receptors are generally involved in many major diseases, thus making GPCR receptors one of the most exploited targets for the pharmaceutical industry (Schertler, 1998). About 5500 GPCR sequences are publicly available. The total number of GPCRs with and without introns in the human genome has been estimated to be approximately 900, of which 500 are odorant or taste receptors and 450 are receptors for endogenous ligands (Takeda et al., 2002). Binding constants are available for approximately 30000 ligand-receptor combinations (Horn et al., 1998).

This wealth of sequences, ligands, and mutations are in net contrast with the small amount of structural information available. Indeed, till a few years ago, rhodopsin was the only structurally characterized GPCR and is still considered a prototypical member of the superfamily. The first X-ray structure of rhodopsin reflected the dark adapted ground-state of the bovine receptor, captured in four different crystals of 2.8 Å (Palezewski et al., 2000), 2.65 Å (Li et al., 2004) and 2.2 Å (Okada et al., 2004), 3.4 Å (Standfus et al., 2007), 4.15 Å (Salom et al., 2006). Very recently the structure of a GPCR in its empty state, Opsin, has been crystallized (Park et al., 2008), opening in this way an exciting new era in the studies of GPCRs. This new structure of rhodopsin followed another very recent key event, the crystallization of a second GPCR, the beta2-adrenergic receptor (Bokoch et al., 2010; Cherezov et al., 2007; Hanson et al., 2008; Rasmussen 2007).

Nearly all medicines are discovered by trial and error. Nevertheless, most pharmaceutical companies have large research departments that use every imaginable technique to design drugs. Homology modelling, as a tool to obtain structural information, is one of those techniques. In the past, bacteriorhodopsin (Henderson & Schertler, 1990; Luecke et al., 1998; Pebay-Peyroula et al., 1997; Takeda et al., 1998) was often used as a modelling template, but from 2000 the three-dimensional coordinates (Palezewski et al., 2000) of bovine rhodopsin have become available. Along several years, it was shown to be a much better template for GPCR homology modelling than bacteriorhodopsin.

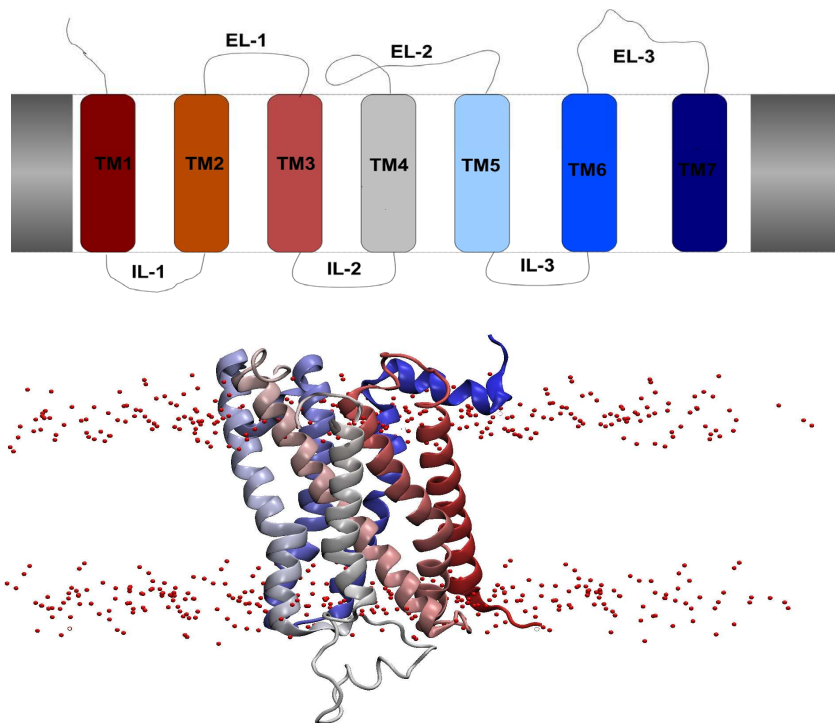


Fig. 3. Upper panel. Topology of GPCR receptors. TM, EL and IL are transmembrane helix, extracellular loops and intracellular loops, respectively. Lower panel: The three dimensional distribution of a GPCR: model of the TAS2R38 receptor.

As can be appreciated from Figure 3, structurally GPCRs are characterized by an extracellular N-terminus, followed by seven transmembrane (7-TM) α -helices (TM-1 to TM-7) connected by three intracellular (IL-1 to IL-3) and three extracellular loops (EL-1 to EL-3), and finally an intracellular C-terminus. The GPCR arranges itself into a tertiary structure resembling a barrel, with the seven transmembrane helices forming a cavity within the plasma membrane that serves a ligand-binding domain that is often covered by EL-2.

Rhodopsin has been for several years an extraordinarily valuable system for understanding the structure and mechanism of activation of G-protein-coupled receptors (GPCRs). Rhodopsin is highly specialized for the detection of light, exhibiting functional and biochemical characteristics that differentiate it from GPCRs expressed in other tissues such as those specialized in detecting diffusible hormones and neurotransmitters. Crystal structures have recently been determined also for the human β 2 adrenoreceptor (β 2AR) (Bokoch et al., 2010; Cherezov et al., 2007; Hanson et al., 2008; Rasmussen 2007), a receptor for adrenalin and noradrenalin that is involved in the regulation of cardiovascular and pulmonary function by the sympathetic nervous system. β 2AR was the first non-rhodopsin GPCR to be cloned and is one of the most extensively studied members of this family (Lefcowitz, 2000). These structures provide the keys for a highly expected way for compare and contrast with rhodopsin.

The structures of the β 2AR provide new exciting insights into the mechanisms of activation in several ways. On the other hand, there exist a huge amount of mutagenesis, biophysical and computational data for the β 2AR and closely related receptors. The new structures provided us with a structural scaffold that may allow the interpretation and validation/rejection of these studies and for generating testable hypothesis for future studies.

One of the most important achievements of crystallography in the last few years, was to provide the community with a way of comparing rhodopsin against other members of the superfamily with different functions. Indeed, rhodopsin evolved for the efficient detection of light: it is present in only one organ and serves only one purpose. In the dark it has almost no activity toward its corresponding G protein, but just one photon can photoisomerize its covalently bound ligand, retinal, changing it from an inverse agonist to a full agonist. By contrast, the β 2AR, like many other GPCRs, has a broader range of signalling behaviour: coupling to more than one G protein and to G protein independent pathways, and responding to a large spectrum of cognate molecules (Lefcowitz & Shenoy, 2005). Comparing the structures can help in gaining insights into the structural basis for these functional differences. In rhodopsin, the binding pocket, specialized in covalently binding of retinal, is hindered by a β sheet lid formed by the second extracellular loop (ECL2), and a small domain formed by the N terminus, protecting *cis*-retinal from hydrolysis. This structure would limit access for diffusible agonists and is not present in the β 2AR. By contrast, in the β 2AR ECL2 forms a helix that is constrained by two disulfide bonds such that there is open access to the ligand-binding pocket.

Very recently the structures of two new GPCR were solved, i.e. the human adenosine A2 receptor (Jakola et al., 2008; Lebon et al., 2011; Xu et al., 2011;) and the human beta1-adrenergic receptor, β 1AR, (Moukhametzianov et al. 2011; Warne et al., 2008; Warne et al., 2011), giving also valuable information that allowed the generalization of several structural/functional features conserved along the families. Indeed, all the available structural information, along with experiments coming from the molecular biology and functional assays combined with extensive computational calculations were applied by us and collaborators in order to unravel the binding site and the gating mechanisms of a very particular family of GPCRs, i.e. the bitter taste receptors (Biarnés et al., 2010).

3.2.1 Bitter taste receptors

Mammals, have been prevented, during evolution, from ingesting toxic compounds because of their strong bitter taste (Behrens & Meyerhof, 2009; Meyerhof, 2005; Mueller et al. 2005; Soranzo et al., 2005). This protection mechanism has been carried out for millions of years by a family of about 30 bitter taste receptors (TAS2Rs) expressed in taste receptor cells (Adler et al., 2000; Behrens et al., 2007; Chandrashekar et al., 2000; Matsunami et al. 2000; Shi & Zhang, 2006). TAS2Rs was shown to belong to the super family of GPCR receptors, albeit their low sequence identity with rhodopsin, for example (Adler et al., 2000; Chandrashekar et al., 2000; Matsunami et al., 2000). The binding of a bitter compound to its cognate target TAS2R, is able to fire a downstream cascade of events inside the cell, typical of GPCRs signaling pathways (Chandrashekar et al., 2006), leading to the production of an electrical signal, i.e. bitter taste perception (Behrens & Meyerhof, 2009). Thus, albeit natural selection decreased its constraints in this sense, i.e. we are not walking around tasting plants that we

do not know, the bitter taste is becoming more and more important regarding food taste, cuisine and pharmaceuticals matters. Thus, a complete characterization of the events giving rise to taste perception is needed. One of the most interesting bitter taste receptors, also from the evolutionary point of view, is the human receptor for phenylthiocarbamide (PTC) and propylthiouracil (PROP) molecules, i.e. TAS2R38 receptor. Indeed, within the polymorphisms present in this receptor, the most pronounced ones affect its perception of the PROP/PTC. In fact, differences in the perception of PROP and PTC has divided the human population into tasters and non-tasters. Albeit these astonishing results were reported in the early nineties, polymorphisms in the hTAS2R38 gene underlying the observed phenotype were identified only recently by Kim and coworkers (Kim et al., 2003). Indeed, the taster/non-taster quality originate in three hTAS2R38 non-synonymous polymorphisms, i.e. the haplotypes code for either the amino acids PAV (P49, A262, I296) constituting the taster variant of hTAS2R38 or AVI in the corresponding positions for the non-taster variant (Bufe et al., 2005; Kim et al., 2003). Up to now, the molecular/structural basis of bitter taste sensing were analyzed by very few studies, i.e. on hTAS2R16 and hTAS2R38 relied on computations only (Floriano et al., 2006; Miguet et al., 2006).

In addition, three experimentally guided structure-activity studies are available now, which all addressed hTAS2Rs distantly related to hTAS2R38 (Brockhoff et al., 2010; Pronin et al., 2004; Sakurai et al., 2010). First principle (Floriano et al., 2006) and homology modeling approaches based on bovine rhodopsin (Miguet et al., 2006) have been used to predict the structure of the widely studied bitter taste receptor hTAS2R38 (Bufe et al., 2005; Khafizov et al., 2007; Kim et al., 2003; Kleinau et al., 2007). Both works coincide in the fact that more computational refinement and/or experimental validations are needed.

Very recently (Biarnés et al., 2010) we have used a combined experimental/computational iterative approach with the aim at identifying hTAS2R38 residues involved in binding to one of its main agonists, i.e. PTC, as well as in receptor activation. We used state-of-the-art bioinformatics approaches based on multiple sequence alignment across the whole family of GPCRs combined with structural bioinformatics tools; we also used the homology modeling techniques because sequences at their own were not likely to be sufficient to identify residues in the binding site, as ligands pockets vary largely in position and orientation across this family (Jaakola et al., 2008). Furthermore extensive virtual docking experiments were carried out to predict the putative binding cavities for PTC. In fact, homology modeling and molecular docking has been shown to guide satisfactorily the design of site-directed mutagenesis experiments, in spite of the little power of the structural predictions (Ballesteros & Weinstein, 1992). Indeed, the proposed receptor positions were then studied and validated/rejected by site-directed mutagenesis experiments and measurements of receptor activation by recording intracellular calcium levels following agonist administration.

We thus have proposed that hTAS2R38 activation upon PTC binding is reminiscent of the transition of the G-protein/opsin complex to free rhodopsin (Scheerer et al., 2008). Indeed, we were able to identify some of the residues directly involved in the interaction with the ligand, those that define the shape of the binding cavity and, more important, we were able to identify the residues participating in receptor activation. In our model, TMs 5, 6 and 7 change conformation upon ligand binding, in particular TM6 tilts around the helical bundle upon G-protein binding. Similar sequences of events also have been suggested to play a role for activation of all GPCRs (Altenbach et al., 2008).

4. Conclusion

Membrane proteins, as we saw, are of fundamental importance for the survival of any living being. A deep insight into the molecular mechanisms underlying their function is thus needed for a complete characterization of the way our cells communicate with the rest of the world. Obviously, a complete characterization implies the passage through the structural features of the membrane proteins. Unfortunately, as of today, the structural biology scientific community still finds several inconvenients to systematically solve the structure of membrane proteins, albeit giant steps forward were carried out in the last few years. Moreover, we have to consider that the main challenges for the near future will include the development and application of methods that permit the full description at the molecular/structural level of large protein complexes, most of all including membrane proteins with unknown structure.

In this chapter we have described two examples for which an experimental/computational multidisciplinary approach was shown to be the key for the gaining of insights into complex systems. In both cases not only 'static' structural elements have been identified, but also putative dynamical mechanisms, comprising large conformational changes. Indeed, we have demonstrated that advancements in experimental structural biology, extensively combined with state-of-art computational biology tools and model-guided molecular biology experiments may become extremely effective for the characterization of complex molecular mechanisms including membrane proteins. These observations make us confident that more difficult challenges of structural/functional characterization can be undertaken in short time and that these approaches may provide a great improvement to our understanding of cell and molecular biology events.

5. Acknowledgements

We deeply acknowledge M. Garonzi, M. Zucchelli, A. Pizzolato, C. Grigoli, A. Turati, M. Denitto, A. Atzeni, A. Bazaj, V. Marino, S. Compri, M. I. Muddei and G. Tosadori for the excellent work done for obtaining Figure 1.

6. References

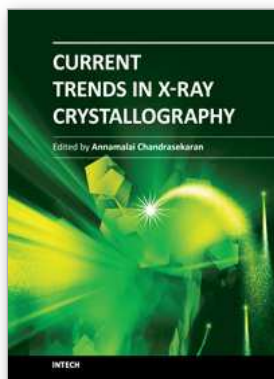
- Adler, E. et al. (2000) A novel family of mammalian taste receptors. *Cell*, 100, 693-702.
- Akabas, M.H. et al. (1992) Acetylcholine receptor channel structure probed in cysteine-substitution mutants. *Science*, 258, 307-310.
- Alberts B., Johnson A., Lewis J., Raff M., Roberts K., and Walter P. (2002) *Molecular Biology of the Cell (4th edition)*, Garland Science, ISBN-10: 0-8153-3218-1, New York.
- Almén, M.S. et al. (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol*, 7, 50.
- Altenbach, C. et al. (2008) High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc. Natl. Acad. Sci. U.S.A.*, 105, 7439-7444.
- Ballesteros, J. A and Weinstein, H. (1992) Analysis and refinement of criteria for predicting the structure and relative orientations of transmembranal helical domains. *Biophys. J.*, 62, 107-109.

- Becchetti,A. et al. (1999) Cyclic nucleotide-gated channels. Pore topology studied through the accessibility of reporter cysteines. *J. Gen. Physiol*, 114, 377-392.
- Becchetti,A. and Roncaglia,P. (2000) Cyclic nucleotide-gated channels: Intra- and extracellular accessibility to Cd²⁺ of substituted cysteine residues within the P-loop. *Pflugers Archiv European Journal of Physiology*, 440, 556-565.
- Behrens, M. and Meyerhof,W. (2009) Mammalian bitter taste perception. *Results Probl Cell Differ*, 47, 203-220.
- Behrens, M. et al. (2007) Gustatory expression pattern of the human TAS2R bitter receptor gene family reveals a heterogenous population of bitter responsive taste receptor cells. *J. Neurosci*, 27, 12630-12640.
- Bezanilla,F. (2008) How membrane proteins sense voltage. *Nat. Rev. Mol. Cell Biol*, 9, 323-332.
- Bezanilla,F. (2005) Voltage-gated ion channels. *IEEE Trans Nanobioscience*, 4, 34-48.
- Biarnés,X. et al. (2010) Insights into the binding of Phenylthiocarbamide (PTC) agonist to its target human TAS2R38 bitter receptor. *PLoS ONE*, 5, e12394.
- Bokoch,M.P. et al. (2010) Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature*, 463, 108-112.
- Brockhoff,A. et al. (2010) Structural requirements of bitter taste receptor activation. *Proc. Natl. Acad. Sci. U.S.A.*, 107, 11110-11115.
- Bufe,B. et al. (2005) The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. *Curr. Biol*, 15, 322-327.
- Chandrashekar,J et al. (2000) T2Rs function as bitter taste receptors. *Cell*, 100, 703-711.
- Chandrashekar,J et al. (2006) The receptors and cells for mammalian taste. *Nature*, 444, 288-294.
- Cherezov,V. et al. (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*, 318, 1258-1265.
- Chothia,C. and Lesk,A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J*, 5, 823-826.
- Craven,K.B. and Zagotta,W. N. (2006) CNG and HCN channels: two peas, one pod. *Annu. Rev. Physiol*, 68, 375-401.
- Floriano,W.B. et al. (2006) Modeling the human PTC bitter-taste receptor interactions with bitter tastants. *J Mol Model*, 12, 931-941.
- Flynn,G.E. and Zagotta,W. N. (2001) Conformational changes in S6 coupled to the opening of cyclic nucleotide-gated channels. *Neuron*, 30, 689-698.
- Giorgetti,A. and Carloni,P. (2003) Molecular modeling of ion channels: structural predictions. *Curr Opin Chem Biol*, 7, 150-156.
- Giorgetti,A., Nair,A.V., et al. (2005) Structural basis of gating of CNG channels. *FEBS Lett*, 579, 1968-1972.
- Giorgetti,A., Raimondo,D., et al. (2005) Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics*, 21 Suppl 2, ii72-76.
- Hanson,M.A. et al. (2008) A specific cholesterol binding site is established by the 2.8 Å structure of the human beta2-adrenergic receptor. *Structure*, 16, 897-905.
- Henderson,R. and Schertler,G.F. (1990) The structure of bacteriorhodopsin and its relevance to the visual opsins and other seven-helix G-protein coupled receptors. *Philosophical transactions of the Royal Society of London. Series B: Biological sciences*, 326, 379-389.
- Horn,F. et al. (1998) GPCRDB: An information system for G protein-coupled receptors. *Nucleic Acids Research*, 26, 275-279.

- Jaakola, V. P. et al. (2008) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*, 322, 1211-1217.
- Johnson, J.P. and Zagotta, William N. (2001) Rotational movement during cyclic nucleotide-gated channel opening. *Nature*, 412, 917-921.
- Karlin, A. and Akabas, M.H. (1998) Substituted-cysteine accessibility method. *Meth. Enzymol*, 293, 123-145.
- Kaupp, U.B. et al. (1989) Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel. *Nature*, 342, 762-766.
- Kaupp, U.B. and Seifert, R. (2002) Cyclic nucleotide-gated ion channels. *Physiol. Rev*, 82, 769-824.
- Khafizov, K. et al. (2007) Ligand specificity of odorant receptors. *J Mol Model*, 13, 401-409.
- Kim, U. K. et al. (2003) Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science*, 299, 1221-1225.
- Kleinau, G. et al. (2007) Implications for molecular mechanisms of glycoprotein hormone receptors using a new sequence-structure-function analysis resource. *Mol. Endocrinol*, 21, 574-580.
- Lebon, G. et al. (2011) Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature*, 474, 521-525.
- Lefkowitz, R.J. (2000) The superfamily of heptahelical receptors. *Nature Cell Biology*, 2, E133-E136.
- Lefkowitz, R.J. and Shenoy, S.K. (2005) Transduction of receptor signals by β -arrestins. *Science*, 308, 512-517.
- Li, J. et al. (2004) Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol*, 343, 1409-1438.
- Liu, J. and Siegelbaum, S.A. (2000) Change of pore helix conformational state upon opening of cyclic nucleotide-gated channels. *Neuron*, 28, 899-909.
- Luecke, H. et al. (1998) Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. *Science*, 280, 1934-1937.
- MacKinnon, R. (2003) Potassium channels. *FEBS Lett*, 555, 62-65.
- Matsunami, H. et al. (2000) A family of candidate taste receptors in human and mouse. *Nature*, 404, 601-604.
- Matulef, K. et al. (1999) Molecular rearrangements in the ligand-binding domain of cyclic nucleotide-gated channels. *Neuron*, 24, 443-452.
- Matulef, K. and Zagotta, W. N. (2003) Cyclic nucleotide-gated ion channels. *Annu. Rev. Cell Dev. Biol*, 19, 23-44.
- Mazzolini, M., Anselmi, C., and Torre, V. (2009a) The analysis of desensitizing CNGA1 channels reveals molecular interactions essential for normal gating. *J. Gen. Physiol*, 133, 375-386.
- Mazzolini, M., Marchesi, A., Giorgetti, A., and Torre, Vincent (2010a) Gating in CNGA1 channels. *Pflugers Arch*, 459, 547-555.
- Meyerhof, W. (2005) Elucidation of mammalian bitter taste. *Rev. Physiol. Biochem. Pharmacol*, 154, 37-72.
- Miguet, L. et al. (2006) Computational studies of ligand-receptor interactions in bitter taste receptors. *J. Recept. Signal Transduct. Res*, 26, 611-630.

- Moukhametzianov,R. et al. (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a beta1-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.*, 108, 8228-8232.
- Mueller,K. L. et al. (2005) The receptors and coding logic for bitter taste. *Nature*, 434, 225-229.
- Nair,A.V. et al. (2006) Locking CNGA1 channels in the open and closed state. *Biophys. J.*, 90, 3599-3607.
- Nair,A.V. et al. (2009) Conformational rearrangements in the S6 domain and C-linker during gating in CNGA1 channels. *Eur. Biophys. J.*, 38, 993-1002.
- Okada,T. et al. (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.*, 342, 571-583.
- Palczewski,K et al. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, 289, 739-745.
- Park,J.H. et al. (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*, 454, 183-187.
- Pebay-Peyroula,E. et al. (1997) X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science*, 277, 1676-1681.
- Pronin,A.N. et al. (2004) Identification of ligands for two human bitter T2R receptors. *Chem. Senses*, 29, 583-593.
- Rasmussen,S.G.F. et al. (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature*, 450, 383-387.
- Sakmar,T.P. (2002) Structure of rhodopsin and the superfamily of seven-helical receptors: the same and not the same. *Curr. Opin. Cell Biol.*, 14, 189-195.
- Sakurai,T. et al. (2010) Characterization of the beta-D-glucopyranoside binding site of the human bitter taste receptor hTAS2R16. *J. Biol. Chem.*, 285, 28373-28378.
- Salom,D. et al. (2006) Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 16123-16128.
- Scheerer,P. et al. (2008) Crystal structure of opsin in its G-protein-interacting conformation. *Nature*, 455, 497-502.
- Schertler,G.F. (1998) Structure of rhodopsin. *Eye (Lond)*, 12 (Pt 3b), 504-510.
- Sesti,F. et al. (1995) The multi-ion nature of the cGMP-gated channel from vertebrate rods. *Journal of Physiology*, 487, 17-36.
- Shi,P. and Zhang,J. (2006) Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol. Biol. Evol.*, 23, 292-300.
- Soranzo,N. et al. (2005) Positive selection on a high-sensitivity allele of the human bitter-taste receptor TAS2R16. *Curr. Biol.*, 15, 1257-1265.
- Standfuss,J. et al. (2007) Crystal structure of a thermally stable rhodopsin mutant. *J. Mol. Biol.*, 372, 1179-1188.
- Swartz,K.J. (2008) Sensing voltage across lipid membranes. *Nature*, 456, 891-897.
- Swartz,K.J. (2004) Towards a structural view of gating in potassium channels. *Nat. Rev. Neurosci.*, 5, 905-916.
- Takeda,K. et al. (1998) A novel three-dimensional crystal of bacteriorhodopsin obtained by successive fusion of the vesicular assemblies. *Journal of Molecular Biology*, 283, 463-474.
- Takeda,S. et al. (2002) Identification of G protein-coupled receptor genes from the human genome sequence. *FEBS Lett.*, 520, 97-101.

- Tramontano,A. (2006) The role of molecular modelling in biomedical research. *FEBS Letters*, 580, 2928-2934.
- Warne,T. et al. (2011) The structural basis for agonist and partial agonist action on a $\beta(1)$ -adrenergic receptor. *Nature*, 469, 241-244.
- Warne,T. et al. (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature*, 454, 486-491.
- Xu,F. et al. (2011) Structure of an agonist-bound human A2A adenosine receptor. *Science*, 332, 322-327.
- Zagotta,W.N. et al. (2003) Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature*, 425, 200-205.



Current Trends in X-Ray Crystallography

Edited by Dr. Annamalai Chandrasekaran

ISBN 978-953-307-754-3

Hard cover, 436 pages

Publisher InTech

Published online 16, December, 2011

Published in print edition December, 2011

This book on X-ray Crystallography is a compilation of current trends in the use of X-ray crystallography and related structural determination methods in various fields. The methods covered here include single crystal small-molecule X-ray crystallography, macromolecular (protein) single crystal X-ray crystallography, and scattering and spectroscopic complimentary methods. The fields range from simple organic compounds, metal complexes to proteins, and also cover the meta-analyses of the database for weak interactions.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Alejandro Giorgetti and Stefano Piccoli (2011). Knowledge Based Membrane Protein Structure Prediction: From X-Ray Crystallography to Bioinformatics and Back to Molecular Biology, Current Trends in X-Ray Crystallography, Dr. Annamalai Chandrasekaran (Ed.), ISBN: 978-953-307-754-3, InTech, Available from: <http://www.intechopen.com/books/current-trends-in-x-ray-crystallography/knowledge-based-membrane-protein-structure-prediction-from-x-ray-crystallography-to-bioinformatics-a>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.