Załącznik nr 3
Wersja angielska

Appendix no 3

Dorota Latek, PhD
University of Warsaw
Faculty of Chemistry

SUMMARY OF PROFESSIONAL ACCOMPLISHMENTS

Determination of structures and mechanisms of action of selected transmembrane proteins considering their evolutionary diversity

Warsaw, April 2019

Table of Contents

1. Name 3
2. Information about diplomas and degrees 3
3. Information on hitherto employment in scientific units 3
4. Description of the achievement resulting from art. 16 sec. 2 of the Act of 14 March 2003 on academic degrees and academic title (Journal of Laws of 2016, item 882, as amended in Journal of Laws of 2016, item 1311.):
    a) Title of the scientific achievement 4
    b) List of scientific publications that constitute the scientific achievement 4
    c) Description of the scientific objective referring to the above publications and obtained results with description of their possible applications 5
        I. Research objective 5
        II. The state of the art of knowledge 7
        III. Research methodology 12
        IV. Detailed description of the achievement and its applications 17
        V. Summary 28
        VI. References 29
5. Description of the remaining scientific and research achievements 36

1. Name

Dorota Latek

2. Information about diplomas and degrees

2009	**PhD in Chemistry** Thesis title: "Modeling of protein structure based on sparse and inaccurate experimental data", promoter: Prof. dr hab. Andrzej Koliński, Faculty of Chemistry, University of Warsaw
2004	**MSc in Chemistry** Thesis title: "Prediction of globular protein structure based on sparse NMR data", promoter: Prof. dr hab. Andrzej Koliński, Faculty of Chemistry, University of Warsaw

3. Information on hitherto employment in scientific units

2004-2009	Faculty of Chemistry, University of Warsaw, PhD studies
2009-2010	Institute of Physical Chemistry of the Polish Academy of Sciences, post-doc (contract)
2010-2013	International Institute of Molecular and Cell Biology, post-doc (long-term contract)
2011	European Molecular Biology Laboratory, Heidelberg, Germany, post-doc (contract for five months)
2014	University of California San Francisco, United States, post-doc (EMBO short-term fellowship for three months)
Od 2013 r.	Faculty of Chemistry, University of Warsaw, adjunct (contract)

4. Description of the achievement resulting from art. 16 sec. 2 of the Act of 14 March 2003 on academic degrees and academic title and degrees and title in the field of art (Journal of Laws of 2018, item 1789):

a) Title of the scientific achievement

Determination of structures and mechanisms of action
of selected transmembrane proteins
considering their evolutionary diversity

b) List of scientific publications that constitute the scientific achievement

Item no.	Author / authors, publishing year, publication title, journal name, volume, pages	Impact Factor[1]	Times cited[2]
H1	**Latek D**, Kolinski M, Ghoshdastider U, Debinski A, Bombolewski R, Plazinska A, Jozwiak K & Filipek S* (2011) *Modeling of ligand binding to G protein coupled receptors: cannabinoid CB1, CB2 and adrenergic beta 2 AR.* **J Mol Model**, 17, 2353-2366	1.797	14
H2	Dreisigacker S, **Latek D**, Bockelmann S, Huss M, Wieczorek H, Filipek S, Gohlke H, Menche D & Carlomagno T* (2012) *Understanding the inhibitory effect of highly potent and selective archazolides binding to the vacuolar ATPase.* **J Chem Inf Model**, 52, 2265-2272	4.304	11
H3	**Latek D**, Modzelewska A, Trzaskowski B, Palczewski K & Filipek S* (2012) *G protein-coupled receptors--recent advances.* **Acta Biochim Pol**, 59, 515-529	1.185	48
H4	Trzaskowski B, **Latek D**, Yuan S, Ghoshdastider U, Debinski A & Filipek S* (2012) *Action of molecular switches in GPCRs--theoretical and experimental studies.* **Curr Med Chem**, 19, 1090-1109	4.07	157
H5	**Latek D***, Pasznik P, Carlomagno T & Filipek S* (2013) *Towards Improved Quality of GPCR Models by Usage of Multiple Templates and Profile-Profile Comparison.* **PLOS ONE**, 8, e56742	3.534	34
H6	**Latek D***, Bajda M, Filipek S* (2016) *A Hybrid Approach to Structure and Function Modeling of G Protein-Coupled Receptors.* **J Chem Inf Model**, 56(4), 630-641	3.76	10
H7	**Latek D*** (2017) *Rosetta Broker for membrane protein structure prediction: concentrative nucleoside transporter 3 and corticotropin-releasing factor receptor 1 test cases.*	1.308	2

	BMC Structural Biology, 17(1), 8		
H8	Miszta P, Pasznik P, Jakowiecki J, Sztyler A, **Latek D**, Filipek S* (2018) *GPCRM – a homology modeling web service with triple membrane-fitted quality assessment of GPCR models.* **Nucleic Acid Research**, 46(W1), W387-W395	11.561[3]	1
H9	Pasznik P, Rutkowska E, Niewieczerzal S, Cielecka-Piontek J, **Latek D*** (2019) *Potential off-target effects of beta-blockers on gut hormone receptors: in silico study including GUT-DOCK – a web service for small-molecule docking.* **PLOS ONE**, 14(1), e0210705	2.766[3]	1
H10	**Latek D***, Rutkowska E, Niewieczerzal S, Cielecka-Piontek J (2019) *Drug-induced diabetes type 2: In silico study involving class B GPCRs.* **PLOS ONE**, 14(1), e0208892	2.766[3]	0

* - corresponding author

[1]according to the publishing year

[2]according to Web of Science, date: 08/04/2019

[3]no data from the publishing year, 2017 impact factor

c) Description of the scientific objective referring to the above publications and obtained results with description of their possible applications

I. Research objective

The structure and function of proteins of a given organism is formed during evolution by a number of environmental factors leading to the occurrence of a specific phenotype, often intentionally induced [1]. In the microscale, these changes are observed as the variability of nucleotide and amino acid sequences [2]. The main objective of my research described in the current summary and conducted considering evolutionary diversity of transmembrane proteins was:

- to develop a complete modeling procedure focused on both, structure and function of selected families of transmembrane proteins: G protein-coupled receptors (GPCRs) [3], transporters (solute carriers, SLCs) [4] and ATPases [5], that was described in subsections: **GPCRM – a web application focused on a GPCR receptor structure**, **GUT-DOCK – a web application focused on a GPCR receptor function** and

Structural characterization of selected transmembrane proteins (the last two protein families)

- to determine functionally relevant structural features of selected representatives of the above listed protein families (rhodopsin-like GPCR receptors, secretin receptors, class F GPCR receptors, vacuolar V-ATPase, concentrative nucleoside transporter CNT3) by using the developed modeling procedure and associated applications, that was described in the subsection: **Structural characterization of selected transmembrane proteins**
- to describe the molecular mechanisms of action of selected transmembrane proteins associated with their natural activation process and drug or disease-induced inhibition of their function, subsections: **Study of mechanism of activation of GPCR receptors**, **Design of transmembrane protein inhibitors** and **Molecular bases of drug-induced reactions**
- to study the effect of single nucleotide polymorphisms on the structure and function of selected transmembrane proteins and to explain the associated specific phenotypes observed in clinical trials, that was described in the subsection: **Molecular basis of pharmacologically significant polymorphisms**

Evolutionary information referring to sequence variability (e.g., protein sequence profiles, protein sequence motifs) [2], similarity of structures, cofactors and substrates of sequence homologs [6] together with statistical potentials [7] was mainly used by me in modeling of protein structures **[H1-2] [H5-10]**. Evolutionary information referring to similarity of protein function **[H3-4]** was used by me in studies on molecular mechanisms of activation of transmembrane proteins that have not been yet crystallized. Models of transmembrane proteins which were prepared by me considering genetic variants identified in patients [8], have also found biomedical applications **[H9-10]** in the field of personalized medicine [9].

As a research objective I chose integral transmembrane proteins due to their important role not only in cell signaling pathways, but also in the transport and metabolism of endo and exogenous substances [10]. In addition, the selected protein families are extremely difficult to study with experimental methods, not only in the case of appropriate substitution of their native non-polar environment during the experiment [11], but also in the case of data acquisition and analysis [3, 4, 5].

Development of a consistent and complete modeling procedure for these proteins required achieving a balance between its optimization towards the maximum global similarity of a protein model to its corresponding native structure [12] and its optimization towards the most accurate reconstruction of a local protein active site. Development of the modeling procedure also required repeatability of the modeling process, which is necessary, e.g., for drug design studies [13, 14]. Implementation of the modeling procedure in the form of web applications provided simple tools to conduct the basic theoretical research on any GPCR receptor of an unknown structure **[H5-6] [H8-10]**. As a result, not only exact structural characteristics of many GPCRs became possible, constituting the basis for experimental research, but also a detailed study of their activation mechanisms, which I described in **[H1] [H3-4]**.

My aim was also to compare structural features and mechanisms of activation of GPCR receptors from two different subfamilies: rhodopsin-like receptors (commonly used name: class A) and secretin receptors (class B). Class A GPCR receptors were described in **[H1] [H3-6] [H8]**, class B receptors were described in **[H9-10]**, class F receptor SMO – a part of **[H6]** and all GPCR receptors classes were described in detail in **[H3-4]**.

A significant part of drugs currently used in pharmacotherapy was designed to modify activity of biological targets with unknown structures [15]. Knowledge of a structure of a biological target of a given drug allows to improve its specificity and selectivity in treatment, which significantly reduces the likelihood of adverse drug reactions. Therefore, the aim of my work was also to study the molecular basis of drug-induced type II diabetes and to propose a way to avoid it **[H9-10]**. In addition, I performed a study on the role of genetic diversity of patients in the effectiveness of treatment with nucleoside analogues **[H7]**.

II. The state of the art of knowledge

A particularly important role of theoretical methods in studies on transmembrane proteins is to provide an initial protein model (most often a homology model) while resolving its structure based on X-ray crystallography or electron cryo-microscopy data [16]. In the second case, an accurate analysis of structural data reflected by a microscope image requires building of a starting conformation of a given protein, which is then optimized on the basis of experimental data [17]. A similar approach has been used for many years in X-ray crystallography, in which a protein model is fitted to an electron density map [18]. Differences

between transmembrane and globular proteins in experimental structure determination depend on the special, non-polar environment imitating the cell membrane that has to be reconstituted during the experiment, e.g., by means of detergents, micelles [19] or lipid nanodiscs [11]. In the case of crystallographic studies of membrane proteins reflecting their distinct conformations, e.g., observed during the activation process (GPCR receptors), it is necessary to additionally stabilize a given conformational state of a protein. Recent advances in X-ray crystallography enabled to perform such studies due to a discovery of stabilizing nanoparticles (recombinant antigens) [20]. In 2011, for the first time, Brian Kobilka together with collaborators from United States and Europe had obtained the crystal structure of a GPCR receptor (β2-adrenergic) in an active conformation in a complex with G protein [21], after 11 years from the first X-ray determination of a GPCR receptor structure (rhodopsin) [19] and its dimers [22] by Krzysztof Palczewski and his collaborators (Sławomir Filipek). Another solution used for the A2A adenosine receptor structure was implemented by Heptares Therapeutics and relies on a thermostabilization of an active state of GPCR receptor by point mutations (StaR® technology) [23]. A recently released structure of an active GPCR receptor conformation (calcitonin receptor CT-R belonging to the secretin subfamily) of 3.3 Å resolution was obtained with cryo-EM using a novel method for data acquisition to obtain a high-resolution protein structure (Volta phase plate) [24]. Some of GPCRs crystal structures determined so far represent conformations of intermediate states (not fully active) instead of actives states despite binding of full or partial agonists by receptors [25]. That confirms experimental difficulties in stabilizing active states of GPCR receptors.

Similar difficulties in experimental structure determination are observed in the case of the other two families of transmembrane proteins that I studied. V-ATPase rotor structures illustrating its mechanism of action had been solved by cryo-EM using DDM detergent (n-Dodecyl β-D-maltoside) [26] and lipid nanodiscs [27] for the non-polar membrane environment reconstitution after 3 and 6 years, respectively, since the publication of my theoretical studies **[H2]**. Intermediate states of a rotor action were shown in [26] (see Fig. 1).

A structure of concentrative nucleoside transporter vCNT3 of the SLC28 family in a closed (ligand bound), inward-occluded conformation was obtained by X-ray crystallography in 2012, also using a DDM detergent as a solubilizer [28]. The structure of vCNT3 was used in my study as a structural template for homology modeling. Quite recently, in 2017, using X-ray crystallography, a full transportation cycle (from an outward-open to an inward-open conformation) of another transporter belonging to the SLC29 family has been determined [29].

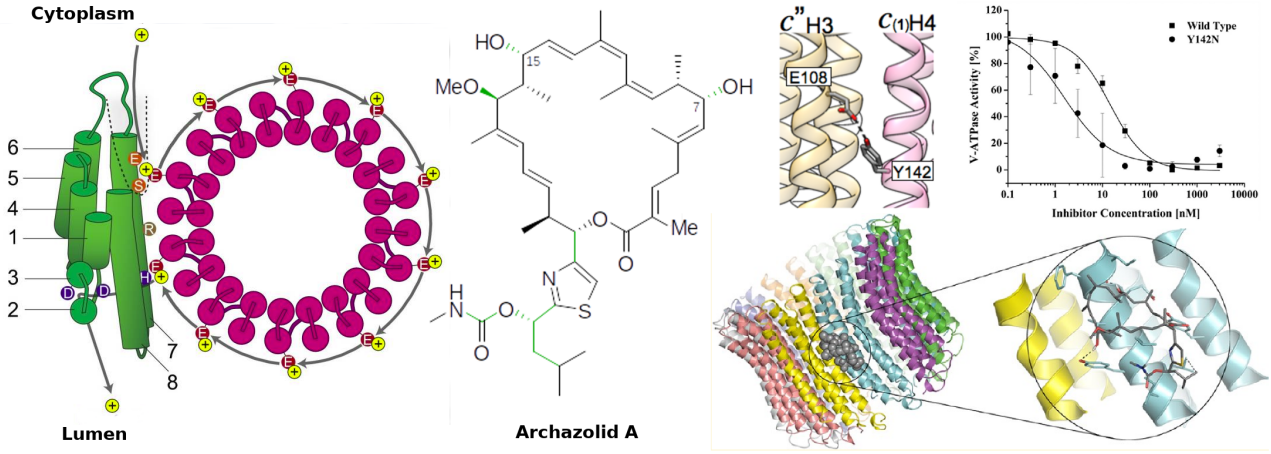

Figure 1. A mechanism of action of a transmembrane domain of a molecular machine – V-ATPase rotor (left panel). ATP hydrolysis-induced rotation of helical transmembrane ring enables protons (yellow) transport through cell membrane in which glutamic acid residues (red) are involved. A middle panel – a chemical structure of archazolid A, V-ATPase inhibitor which blocks Y142. The ligand activity and its binding mode (right panel) was confirmed experimentally in **[H2]**. Based on: Shep DG et al. PNAS 2016, 113 (12) 3245-3250, Roh SH et al. Mol Cell 2018, 69(6):993-1004.e3 and **[H2]**.

A small number of transmembrane protein structures solved so far by experimental methods illustrates difficulties and at the same time is the justification for my theoretical studies described in this summary. Experimental difficulties also apply to the correct identification of potential biological targets for novel, pharmacologically active substances, which in turn may lead to the occurrence of adverse drug reactions [30], e.g. due to the modulation of other than expected signaling pathways as I described in **[H9-10]**. Functional tests, which are massively used to assess the actual (not only affinity) biological activity of the ligand [30], have limitations resulting from the fact that the experiment is not performed in biological conditions (*in vivo*) [31].

A ligand ability to activate several cell signaling pathways has been confirmed, for example, for beta-blockers (carvedilol and nebivolol), which were considered as antagonists of β-adrenergic receptor but were proved to actually activate as agonists the β-arrestin signaling pathway of this receptor [30]. Even phenotypic screening [30], due to its standardization, is not able to detect weak off-target ligand-receptor interactions associated with less severe side effects of pharmacotherapy which are often observed after many years of treatment in clinical trials. Automated protocols used for, e.g., ADME-Tox tests, allow to exclude compounds with high cytotoxicity, but often do not exclude compounds with relatively low toxicity [32]. For

example, in **[H9-10]** I described the impact of various drug classes on the likelihood of inducing diabetes during the treatment. Type II diabetes is an example of a long-term side effect of pharmacotherapy in contrast to immediate adverse drug reactions.

In recent years, as a result of a significant number of reported cases of adverse drug reactions [33], a growing interest in molecular mechanisms of such reactions has been observed and several attempts has been made to develop protocols for their prediction [34]. Notably, such studies on the mechanisms of side effects occurrence are significantly hampered by incomplete and often contradictory [35] results of clinical trials, that was described in my work **[H10]** in the case of two independent clinical trials referring to the statin drug class. Other complications are associated with increasing reports the patient's individual response to treatment related to his genetic profile [36]. Nevertheless, there are attempts to gather information on adverse drug reactions in a more standard and public way, e.g., in the form of public databases such as SIDER (Side Effect Resources), released by European Molecular Biology Laboratory [36]. Also, it should be mentioned that some of the observed side effects of drugs are often used in drug repositioning, which is an alternative way to design of completely new active substances that in other ways require long and tedious clinical tests before being used in treatment.

The above described experimental difficulties associated with structure determination and investigation of transmembrane proteins have been overcome over the years by a number of different theoretical methods. These include: comparative modeling (including homology modeling), *de novo* modeling (without any structural template of a homologous protein), fragment assembly, evolutionary information from correlated mutations, threading, fold recognition and many other methods with a smaller range of use. In the field of drug discovery, standard theoretical methods involve virtual screening (structure-based or ligand-based) and lead optimization in order to improve drug specificity and selectivity. Methods listed above have been described in detail in all my publications **[H1-10]** and additionally in **[B1-9]** and 8 book chapters. Publications on *de novo* protein structure prediction by Andrzej Koliński [37] and protein modeling supported by evolutionary information [38, 39] deserves, in my opinion, the most attention. Development of Rosetta – an algorithm for protein modeling using fragment assembly [40] by David Baker and co-workers was a breakthrough discovery not only in the field of computational methods but also in the field of the whole proteomics. Development of an optimal method for conversion of three-dimensional structural information from templates into spatial restraints for a protein of an unknown structure by Andrej Sali (MODELLER [41])

enabled to characterize structurally a large number of proteins (deposited in the MODBASE database), similarly to another tool for comparative modeling (SWISS-MODEL [42]).

For many years, limited population of transmembrane protein structures in the Protein Data Bank database (PDB) was a serious obstacle not only in development of statistical potentials for *de novo* modeling but also in typical comparative modeling studies. Therefore, publications from the last few years referring to the use of evolutionary information in the modeling of transmembrane protein structures seem to be extremely important. Such evolutionary information in the form of correlated mutations was used to predict inter-residue contacts [43, 44] and in the form of sequence profiles was used in secondary structure and topology prediction [45] and was supplemented with statistical potentials derived specifically for transmembrane proteins [46]. Another solution in theoretical studies of transmembrane proteins is a modification of an existing tool developed originally for globular proteins. For example, based on Rosetta another tool was created (Rosetta-MP [47]). Original force fields used for molecular dynamics simulations (e.g. CHARMM [48], AMBER [49]) were supplemented with additional parameterization of lipid molecules (an explicit membrane environment) or additional potentials characterizing interactions of a protein and ligands with lipids in an approximate way (e.g. the IMM1 method [49] – an implicit membrane environment). As for all-atom molecular dynamics (GROMACS [51], NAMD [52]), it is often used to refine a homology model of a transmembrane protein and to observe its mechanism of action in the microseconds scale [53]. I described it in publications: **[H1-2] [H7]**.

In recent years computational methods imitating the evolutionary selection of organisms (genetic algorithms) or neural networks (machine learning methods) have brought attention. They were used by me in ligand docking (Autodock **[H1-2]** and GOLD **[H6]**). In these methods, a solution to a biological problem (e.g. a protein structure or a ligand binding mode) is found by imitating evolution, i.e. choosing a development pathway in a given external environment that is the most beneficial from the energetic point of view (e.g., the genetic algorithm in the GOLD docking program [54]).

III. Research methodology

Development of a new procedure for protein structure modeling, especially for GPCRs modeling, its optimization and applications in research was described by me in publications: **[H1] [H3-6] [H8-10]** (GPCRs) and in: **[H2] [H7]** (other transmembrane proteins). The

developed procedure included, i.a., model refinement in all-atom molecular dynamics simulations **[H2] [H7]** and multiple templates modeling, described in **[H5]**, which was an innovative solution to the problem of the lack of an appropriate structural template that could be used in, so-called, 'hard' homology modeling of GPCR receptors, when the identity of target and template sequences is well below 30%. My research in this field had been started by, described in **[H3]**, an analysis of sequence similarity of rhodopsin-like GPCRs belonging to four branches of this subfamily: α, β, γ and δ. It showed that in general rhodopsin-like GPCRs are not very similar to each other in the matter of sequences (sequence identity below 20%). An exception of this rule is the heterogeneous branch δ, that includes olfactory receptors of sequence identity ca. 40%. The described above rhodopsin-like GPCRs are an example of transmembrane proteins that sequences were hardly conserved during evolution in contrast to their structure (seven transmembrane helical topology) and function (signal transduction).

I described a solution to the problem of 'hard' homology modeling of GPCR receptors in detail in **[H5]**. It is based on weighted averaging, depending on similarity of whole sequences or their fragments, of available template structures which were deposited in PDB. The averaged multiple templates are then used in protein model building using a modified method of conjugated gradients supplemented with molecular dynamics (MODELLER [41]). I implemented this concept, that allows to build a model of any GPCR receptor of unknown structure using multiple templates modeling (see Fig. 2), in a web application GPCRM described in **[H5-6]** and **[H8]**. Template structures, in other words: three-dimensional protein structures derived from PDB, are converted into a set of restraints describing inter-atomic distances and dihedral angles between bonds. Additionally, a protein model is characterized by a set of stereochemical restraints (bond lengths, bond angles), that was obtained from molecular mechanics computations (CHARMM-22 [48]), and by values of dihedral angles and inter-atomic distances derived from the PDB statistics. In the next step, mentioned above structural data is used to build a protein models by fitting its amino acid chain to previously generated restraints with probability depending on the sequence similarity and the Boltzman distribution, that was described in the original publication of MODELLER's author (Sali et al. [55]).

Due to the fact that all GPCR structures determined so far consist of seven transmembrane helices (they have the same topology) it is possible to use more than 2-3 structural templates without losing the resolution of the final protein model **[H5]**. In contrast, exceeding the limit of 2-3 templates in the case of globular proteins of more diverse topologies may cause decrease in

the accuracy of the final model [56]. It is a result of using too diverse and contradictory structural restraints which cannot be fulfilled at the same time. The described above multiple templates modeling procedure was especially useful in reconstruction of helical distortions (kinks, bulges, grooves) which are typical for membrane proteins. I described it in **[H5]** by taking A_{2A} and κ-opioid receptor (in Supplementary) as examples.

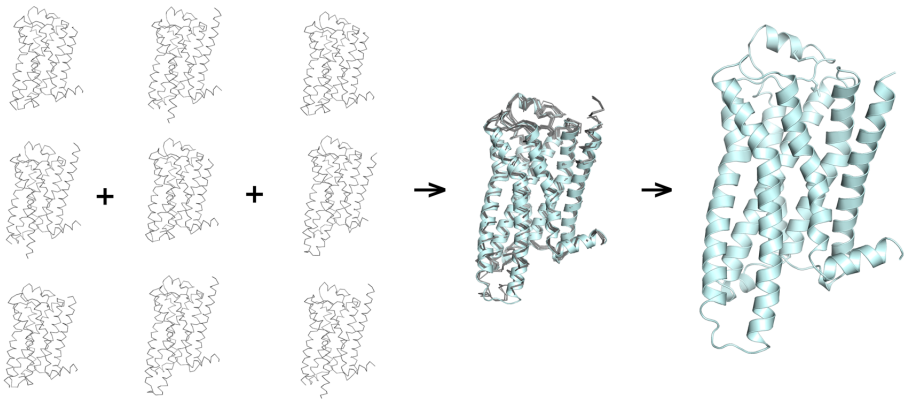

Figure 2. GPCR receptor model building using multiple structural templates of homologous proteins.

My research on class B GPCRs resulted in development of a procedure for an allosteric active site modeling (corticotropin-releasing hormone receptor CRF_1) **[H7]**. In this case, a helical distortion observed in a crystal structure of this receptor was associated with a specific interaction between the receptor and its active ligand (antagonist CP-376395), that had never been observed in any of GPCR structures deposited in PDB before. For the allosteric active site modeling I used an algorithm for *de novo* membrane protein modeling that included statistical potentials (Rosetta Broker [57]). This tool enabled to reconstruct helix VI in such way that it moved away from the active site leaving the space for the antagonist docking **[H7]**.

Protein model building using the multiple templates approach is preceded by, developed by me and described in **[H5]**, a procedure of target-template alignment generation. In addition to commonly used methods (pairwise sequence alignment, multiple sequence alignment), other two methods have been implemented, previously not used in automated GPCR modeling procedures. The first method, that was described in **[H5]**, is using target and template sequence profiles to generate target-template sequence alignment. Sequence profiles contain evolutionary information about the frequency of occurrence of a given amino acid at a given position in a group of homologous sequences (Gribskov et al. [58]). Sequence profiles are generated with

BLAST (Altschul et al. [59]) using the non-redundant database of sequences. Alignment of sequence profiles was used for improvement of sequence alignment in globular protein modeling, e.g., by polish researchers [60], but not in a distinct case of GPCR receptors. In my research I have observed that generation of sequence alignment based on profile-profile alignment enables to detect gaps in alignments, that are associated with helical distortions: grooves (gap in a target sequence) and bulges (gap in a template sequence). I described it in the publication **[H5]** (also in Supplementary).

The second method for the sequence alignment generation, that was designed and described by me in the publication **[H8]**, is based on the initial structural alignment of all templates used for the modeling. Based on this structural alignment, a multiple sequence alignment is generated. This method is particularly useful in 'hard' homology modeling, e.g., class F GPCR receptor structure building using a class A GPCR receptor template (low target-template sequence similarity) **[H6]**, that is too difficult for standard alignment generation methods. I used this method successfully in modeling of the SMO receptor structure during the GPCR Dock 2013 competition [61], which I described in detail in **[H6]** (see also: **5. Description of the remaining scientific and research achievements** and **[B1]**).

In the next step, GPCR models built with the multiple templates approach described in **[H5]** were assessed by two distinct criteria depending on the modeling purpose: either studying interactions with other protein domains or drug design. In the first case, I had to optimize the modeling procedure in such way that the global RMSD (or TM-score) value of a model with respect to its corresponding native structure from PDB was the lowest. To achieve this, I used statistical potentials, that was described in **[H6]** and **[H8]**. Statistical potentials for local and long-distance interactions were derived based on membrane protein structures deposited in PDB. I used statistical potentials implemented in BCL::Score [46], that was described in **[H6]**. I also implemented this model quality assessment method in GPCRM **[H8]**. The second criterium for the GPCR receptor model assessment, that was particularly useful in drug design or studying mechanisms of small-molecule ligands interactions, was described in publications: **[H6]** and **[H9-10]**. It was associated with the optimization of the modeling procedure towards the most accurate reconstruction of the ligand-receptor interactions inside the binding site. In **[H6]** I described the model quality assessment method, that was based on two-step ligand docking. In the first step, ligand docking with GOLD [54] was performed and models with the most converged results were selected for the next step – high precision ligand docking with

Glide. In **[H9-10]** I described another approach, based on enrichment factors. For a few selected class B GPCRs, their active ligands (experimentally confirmed) and a DUD-E-generated decoy set [13] I computed enrichment factors EF and other measures (i.a., ROC, BEDROC, AUC). Models of the highest EFs were selected for ligand docking. Due to this selection criterium I obtained GPCRs models with proven effectiveness in virtual screening, recognizing the active ligands among the inactive for a given receptor. Results of this study were made public in the form of a web application GUT-DOCK **[H9]**. GUT-DOCK enables to perform docking of small-molecule ligands and short peptides (via Autodock VINA [62]) to the mentioned above models of gut hormone receptors. Among the currently available assessment functions used in ligand docking (e.g., force field-based, empirical, based on statistical potentials or machine learning), I chose an empirical function, that was derived based on PDBbind containing experimental data for ligand affinities in complexes deposited in PDB and that was implemented in Autodock VINA.

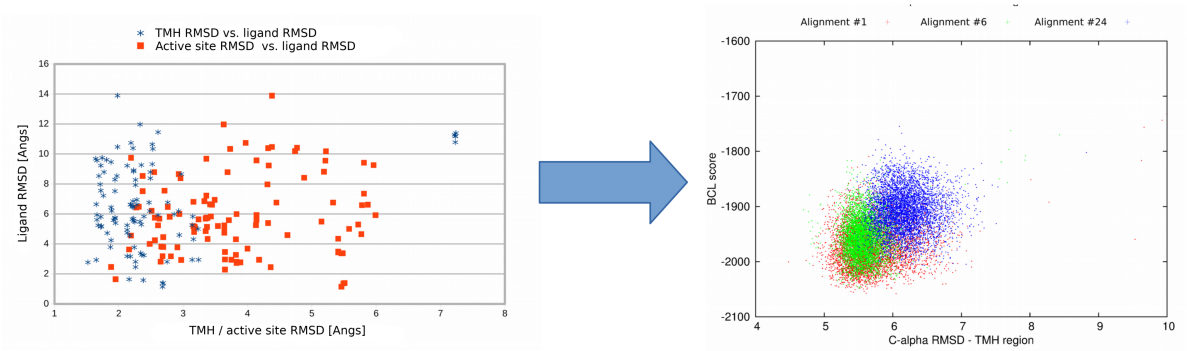

Figure 3. The lack of correlation between RMSD values (with respect to the whole native structure) and results of ligand docking (ligand RMSD) is a major obstacle in theoretical studies of GPCRs (left panel). A change in a position of a single amino acid, e.g., in a consequence of target-template misalignment, may completely change a result of a ligand docking. Therefore, one of GPCRM modules, that was designed by me, enables to improve the accuracy of the active site reconstruction. This is done by improving target-template sequence alignment due to the usage of statistical potentials (right panel). Based on my publication **[H6]** (right panel) and results of all research groups in GPCR Dock 2013 in the 5-HT_{2B} category (left panel) [61].

In rare cases, both mentioned above criteria of model quality assessment can produce consistent results (e.g., results of Stockholm-Carlsson and Copenhagen-Gloriam research groups in the case of 5-HT_{1B} in GPCR Dock 2013 [61] – see: **5. Description of the remaining scientific and research achievements** and **[B1]**). However, in majority of cases (see Fig. 3, left panel, blue) protein models that perform the best in virtual screening are not the models which are the most similar globally to their native structures (the lowest global RMSD with respect to the native structure). It was described in my publication **[H6]** and in previous publications referring to

globular proteins, e.g., in [14]. A slightly better correlation with performance in ligand docking is observed for the accuracy of the structure prediction only in the active site region (see Fig. 3, left panel, red). Noteworthy, results of GPCR Dock 2013 constitute a heterogenious data set, that is difficult for interpretation. I mentioned that in **[H6]** and instead I performed computations for homogenous data sets generated by our group (Warsaw) during the competition. Nevertheless, in my opinion, GPCR Dock 2013 results give an idea of the current difficulties in GPCRs complexes modeling.

The GPCRs modeling procedure (GPCRM) developed by me **[H5-6] [H8]** together with the procedure for modeling of small-molecules GPCR complexes (GUT-DOCK) **[H9-10]** allows to conduct a complete *in silico* study for any GPCR receptor from class A and B. For example, in **[H1]** and **[H3-4]**, I described mechanisms of activation of cannabinoid receptors and other class A GPCRs. These studies had been conducted using the described above procedure, though non-automated yet. In the last stage of my research described in the current summary I used this modeling procedure for a study of drug side effects involving GPCRs **[H9-10]**. In this case I studied molecular mechanisms of drug reactions referring to disruption of glucose metabolism leading to type II diabetes.

The described above methodology for studying GPCR receptors was supplemented with molecular dynamics studies. Molecular dynamics allowed to validate obtained homology models, on one hand, and allowed to observe the mechanism of action of selected proteins, on the other. In **[H2]** I used all-atom molecular dynamics simulations (YASARA [63], AMBER [49]) to refine of a homology model of V-ATPase rotor (wild-type and mutant) and to study the stability of its complexes with archazolid A derivatives. Based on this, the best affinity compound had been selected. In **[H9-10]** I described usage of molecular dynamics (AMBER) for generation of conformational ensembles of glucagon receptors. These conformational ensembles proved to be more useful for virtual screening than the starting crystal structures of GCGR and GLP1R, especially in the latter case **[H9]**.

In the case of a homotrimer of CNT3 concentrative nucleoside transporter from the SLC28 family I used molecular dynamics for model refinement, that was described in **[H7]**. Molecular dynamics improved the fitness of homotrimer subunits and was used for quality assessment for a subunit model, that was built using the hybrid approach (homology modeling combined with *de novo*). In the case of CNT3 transporter evolutionary information was used in a few aspects.

First of all, a crystal structure of vCNT, a bacterial homologous protein of hCNT3, was used to build via comparative modeling approach a major part of a subunit and also the whole hCNT3 homotrimer. Secondly, using statistical potentials derived from known crystal structures of membrane proteins which were implemented in Rosetta Broker [57] and one-dimensional prediction of transmembrane regions (TOPCONS [64]) I built a missing, N-terminal fragment of hCNT3. Consequently, I obtained a full and functional (in molecular dynamics simulations) proteins model of hCNT3 homotrimer. This transporter has been used by me as a test case for developing a procedure for *de novo* transmembrane protein structure prediction combined with evolutionary information.

IV. Detailed description of the achievement and its applications

GPCRM – a web application focused on a GPCR receptor structure

The homology modeling procedure for GPCR receptors, designed and described by me for the first time in **[H1]**, was automated in the form of the freely accessible web application GPCRM, described in **[H5]** and following publications: **[H6]** and **[H8]**. GPCRM can be used to generate models of not only class A GPCRs but also other classes (e.g., class B GPCRs **[H9-10]**). What is more, it can be used in 'hard' homology modeling, when target-template similarity is low (sequence identity below 30%), which was described in **[H8]**. Building of a homology model by GPCRM starts with a template selection step. Based on an amino acid sequence provided by User a multiple sequence alignment is generated with CLUSTALW2 [65]. The multiple sequence alignment including a target sequence and all GPCRM database template sequences is used to compute pairwise target-template sequence identity and similarity. The most similar templates are selected for modeling. In the case of 'Advanced Mode' for experienced Users, that was designed and described by me in **[H5]**, it is possible to select any number of templates for modeling, considering not only sequence similarity but also biological profiles of a target receptor and a template. In the case of 'Automatic mode', if target-template sequence identity exceeds 40%, only one template is selected for modeling. In other case, two, the most similar templates are selected. By building of a GPCR model using the multiple templates approach a mistaken template selection can be avoided, because the final GPCR model do not relay on a single template structure but is a weighted average of all template structures used for modeling. As a consequence, the final GPCR model of the lowest energy may be, e.g., a sum of fragments from various templates, which were selected as the best for a given target sequence fragment. This was described in detail in **[H5]** and **[H6]** (SMO receptor) and previously in [66].

In the next modeling step, using BLAST [59], a target sequence profile is generated and aligned with sequence profiles of templates (profile-profile alignment) using MUSCLE [67] (a progressive alignment method). Generated profile-profile alignment proved to be more accurate in prediction of gaps in GPCRs target-template sequence alignments than corresponding multiple sequence alignments or simple pairwise sequence alignments, that was confirmed in **[H5]**. Generated sequence alignment is then optimized to fit target and templates corresponding sequence motifs which are typical of GPCR receptors and are conserved during evolution **[H4]**. I described typical sequence motifs for various GPCRs classes in detail, e.g., in **[H4]**, where I focused on their role in GPCR receptor activation. Sequence alignments for each target-template pair are used to build a protein model with MODELLER [41] using the described above multiple templates approach. Loops between transmembrane helices are additionally refined using MODELLER-DOPE and an iterative CCD protocol in Rosetta (cyclic coordinate descent). The latter loop modeling method is a fragment assembly modeling method and is particularly useful for long loops (more than 20 amino acids), e.g., extracellular loop 2 (EC2), when computational resources of a web application is limited. In the case of short loops, a more precise conformational search with an *ab initio* method (e.g. MODELLER DOPE) in which a loop is *de novo* constructed may provide better results, that was described in **[H5]** and in a more detailed way in [68]. In many GPCRs, EC2 between transmembrane helices V and VI contains a conserved disulphide bridge, that splits EC2 loop in two parts which undergo the separate refinement in GPCRM **[H5]**. In addition, I had introduced a potential, that minimized cell membrane penetration by extracellular and intracellular loops and N and C-termini.

In the last step of the modeling a generated protein model is assessed with three different scoring functions: Rosetta Total Score, Rosetta-MP, BCL::Score, that was described in **[H8]**. Due to implementation of statistical potentials derived based on the Boltzmann distribution only for transmembrane protein structures deposited in PDB, the latter two scoring functions are more accurate in GPCR structure modeling than the former one **[H6]**.

Except for the computational layer GPCRM provides User an interface, that allows to manually modify sequence alignments and to visualize a generated protein model. The web application GPCRM has been used successfully in two international computational competitions: GPCR Dock 2010 **[B6]** and GPCR Dock 2013 **[B1]**. GPCRM has been also a basis for studying the mechanism of the GPCR receptors activation, that was described in **[H1]** and **[H3-4]**, and also

in other publications which I co-authored (see: **5. Description of the remaining scientific and research achievements**). Finally, GPCRM has been a basis for another web application which is focused on ligand-receptor interactions (GUT-DOCK) **[H9-10]**. The original publication describing the GPCRM algorithm was cited 46 times (according to Google Scholar), a number of visitors since the first release is 7788 from 64 countries (on 22/03/2019). The corresponding data for GUT-DOCK (published in January of 2019): 164 visitors from five countries. Among publications which cited **[H5]**, was a publication on new agonists of 5-HT_{2A} receptor by J. Selent [69]. The described above statistics illustrates the range of the GPCRM use, despite the fact that it had only three permanent contributors: a group leader S. Filipek, D. Latek, a programmer P. Pasznik. Only recently, in 2018 **[H8]**, a number of GPCRM contributors has significantly increased.

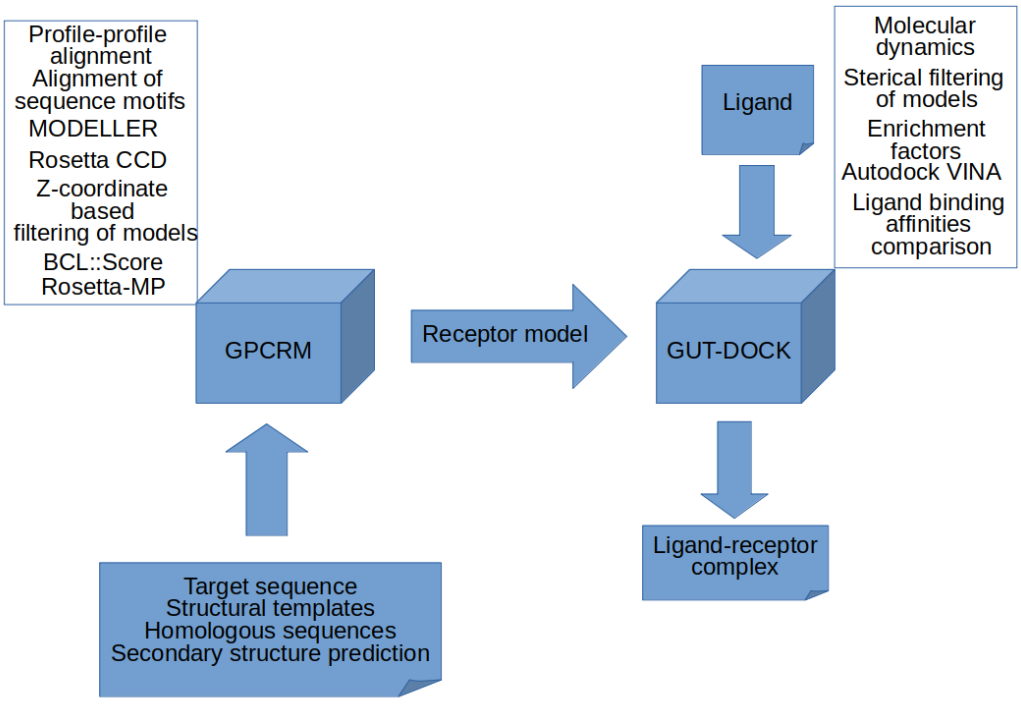

Figure 4. A simplified scheme of web applications GPCRM and GUT-DOCK, which are focused on a GPCR structure and function, respectively.

A comment is needed referring to other web services for GPCR structure modeling which were developed before and after GPCRM. GPCRM had been released in June 2012 as a beta version (GUT-DOCK was released as a beta version in December 2017). In 2012, there were few freely accessible web services for GPCR structure modeling, e.g., GPCRDB [70], GPCR-SSFE [71] and GPCR-ModSim [72]. For this reason, researchers used other services for general protein modeling (not only transmembrane), e.g., Robetta, I-TASSER [73] and SWISS-MODEL.

Results of the mentioned above methods were compared to results of GPCRM in Supplementary of **[H5]**. As for the methodology, all services for GPCR structure modeling existing before 2012 implemented MODELLER and a novel concept was only implemented in GPCR-SSFE [71]. It included GPCR model building by assembly of whole transmembrane helices derived from different template structures and subsequent energy minimization of such composed seven helices bundle. However, templates were not included in the weighted averaging manner, like it was done in GPCRM. A method similar to GPCR-SSFE was developed in 2012 in the field of *de novo* GPCR structure modeling (BiHelix [74]). Also, in this case a receptor model was built from seven ideal helices, which were rotated with respect to each other in order to minimize energy of long-range interactions. Therefore, the mentioned above modeling methods differ fundamentally from GPCRM, that is a major element of the here described scientific achievement.

The modeling methods for GPCRs developed after GPCRM (mainly next versions of GPCRDB [75] and GPCR-SSFE [76] and fitting I-TASSER to GPCRs – GPCR-I-TASSER [77]) started to include the multiple templates approach in their modeling pipelines, which was published in 2013 by me **[H5]**. Noteworthy, GPCRM was a novelty not only in terms of computational algorithms but also in terms of technology (Biopython and Django) **[H5]**. These technologies were implemented in GPCRDB after 2013. Recent developments in GPCRM referring to transmembrane proteins-fitted statistical potentials (BCL::Score, Rosetta-MP) for model quality assessment are present neither in GPCRDB nor in GPCR-SSFE. Only GPCR-I-TASSER implements such statistical potentials, but in a completely different way because in principle, it is a threading program.

GUT-DOCK – a web application focused on a GPCR receptor function

I applied the described above GPCR modeling procedure to another web application GUT-DOCK focused on gut hormone receptors, that has been recently published **[H9-10]**. The previous web application GPCRM was used for an input generation for the followinghj application GUT-DOCK. Namely, I generated GPCR models using GPCRM and used them in GUT-DOCK. In this case I selected only these GPCR models, that performed the best in virtual screening. The central part of GUT-DOCK is flexible ligand docking to rigid GPCRs structures with a parallel algorithm Autodock VINA [62]. The ligand binding mode of the lowest energy is presented as a GUT-DOCK output, in the form of PDB files and visualization of ligand-receptor interactions (Ligplot). Based on my additional study on the molecular basis of the drug-

induced glucose metabolism disruption described in **[H9-10]** I concluded that it would be useful to compare the theoretical ligand binding energy obtained in docking with respective precomputed energies obtained for other drugs (e.g., the beta-adrenolytic drug class). In this way User may design novel, pharmacologically active compounds with a potentially minor diabetogenic effect. Despite obvious limitations associated with the usage of open-sourced tools in the freely accessible web application (Autodock VINA, OpenBabel), GUT-DOCK results have turned out to be comparable with results of licensed docking programs, e.g., Glide (Schrodinger LLC), that was described in **[H9]**. This fact is another proof for effectiveness of my GPCR modeling procedure.

Structural characterization of selected transmembrane proteins

In 2012, when **[H2]** was published, only a crystal structure of bacterial homolog of V-ATPase, that was released in 2005 (PDB id: 2BL2), was known [78]. I built a homology model of yeast vacuolar V-ATPase using this crystal structure and study two functionally important amino acid residues: glutamic acid and tyrosine. These two residues repeated in each rotor subunit (e.g., Glu137, Tyr66, Glu108, Tyr 142 – see Fig. 1) are important also for binding of V-ATPase inhibitors – derivatives of archazolid A. The orientation of tyrosine rotamers in my homology model of V-ATPase was confirmed by two experimental structures released in 2015 [79] and in 2018 [80], that were determined by cryo-EM. In my homology model of V-ATPase the rotamer of glutamic acid (e.g. Glu137 or Glu108) was directed towards the cytoplasmic side, but after refinement in the all-atom molecular dynamics simulation it changed its direction towards the lumen side (see Fig. 1), a few Angstroms away from tyrosine (Tyr66 or Tyr142). An experimental structure of V-ATPase (cryo-EM) released in 2015, that described three conformational states of a rotor, contained that Glu rotamer only in few subunits of a transmembrane ring of one conformational state (PDB id: 3J9V). This conformational state was indeed associated with decreased protons stoichiometry with respect to a protein (3:1 instead 4:1) and thus decreased enzyme activity. However, an experimental structure of V-ATPase released in 2018 contained Glu rotamers in the same open orientation which I predicted in molecular dynamics simulations. This open orientation of Glu was associated with the enzyme autoinhibition [80]. This comparison to experimental results illustrates the accuracy of my computations described in **[H2]**.

Crystal structures of cannabinoid receptors CB_1 and CB_2 have been released in 2016 (CB_1 – an inactive receptor conformation), 2017 (CB_1 – an active receptor conformation) [81] and in 2019

CB_2 – an inactive receptor conformation) [82], but experimental studies from the beginning of the century reported existence of a particular amino acid microswitch [83]. In **[H1]** I described in detail the active site regions of CB_1 and CB_2, that included this Trp/Phe microswitch whose setting can be modulated by ligands. Homology models of CB_1 and CB_2, that were used in **[H1]**, were built using a A_{2A} receptor template and a standard modeling procedure in MODELLER because GPCRM had not been available at that time. Following studies on cannabinoid receptors (e.g., [84]), which cited **[H1]**, used the multiple template approach and site-directed mutagenesis data [85] to generate more precise homology models. Nevertheless, these studies confirmed the existence and settings of this particular microswitch in the receptor inactive state (Trp6.48 – gauche+, ca. -60°, and Phe3.36 – trans, ca. -180°) and in the receptor active state (Trp6.48 – trans and Phe3.36 – gauche+). Experimental structures of CB_1 released in 2016 and in 2017 (e.g., 5TGZ – inactive, 5XR8 – active) have illustrated the same settings of the Trp/Phe microswitch which I showed five years earlier in **[H1]**. Among other studies, Doerksen et al. published this year [86], that also cited **[H1]**, again confirmed the importance of the CB_2 microswitch presented earlier by me. It is worth to mentioned, that a crystal structure of CB_2 bound to antagonists, released in 2019 (PDB id: 5ZTY) [82], has shown the Trp/Phe microswitch in such setting that it has resembled the CB_1 receptor structure but bound to an agonists. In my study described in **[H1]**, that was mainly focused on ligand docking instead of molecular dynamics, I did not confirm clearly this difference between CB_1 and CB_2 microswitches settings. Nonetheless, we supplemented **[H1]** with results from short (ca. 1ns) simulations of molecular dynamics, that showed differences between CB_1 and CB_2 functional profiles. Namely, in Fig. S2 of Supplementary of **[H1]** we presented χ_1 plots for Trp and Phe in two best CB_1 and CB_2 models recorded during MD simulations. Values of χ_1 are defining rotamers of Trp and Phe. In the case of CB_1 values of χ_1 were constant in contrary to results for CB_1 (changes of χ_1 values for Trp and Phe of ca. 120°). This change in χ_1 values for CB_1 could indicate the spontaneous change of the microswitch setting. However, we performed too short simulations to formulate similar conclusions like in [82].

The example of cannabinoid receptors illustrates difficulties in GPCRs research. Every new GPCR crystal structure (on average a few per one year) allows for more and more accurate studying of mechanisms of GPCR action comparing former theoretical studies. I experienced this situation while studying glucagon GPCR receptors **[H9-10]**. I built a homology model of GIPR receptor using apo crystal structures of GCGR receptor released in 2013 and 2016. Latter holo crystal structures of GCGR and holo crystal structures of GLP1R (another close

homologous protein to GIPR), both released in 2017, slightly differed from the former apo structures of GCGR (1.2Å and 0.8Å, respectively). Nevertheless, binding of allosteric antagonists altered mainly the so-called 'stalk' region which joins extracellular and transmembrane domains. This region was not studied in **[H9-10]**. For this reason, I have not repeated homology modeling using the latter template structures. A homology model of GIPR and of two other class B GPCRs (PAC1R and VIPR1) together with molecular dynamics-refined crystal structures of GLP1R and GCGR have been implemented by me in the freely accessible web application GUT-DOCK.

Also, in the case of CNT3, my research described in **[H7]**, was conducted based on a template structure of vCNT released in 2012 (PDB id: 3TIJ), despite the latter crystal structures released in 2014 [88]. Crystal structures of vCNT released in 2014 were determined using the former 2012 crystal structure as an input for PHENIX – a program for crystallographic data analysis. Consequently, the former and the latter crystal structures of vCNT were very similar, differing slightly in the ligand binding site region due to interactions between transporter and different nucleoside ligands, e.g., cytidine analogs (2014) vs. uridine (2012). Nonetheless, **[H7]** refers mainly to the global structure of hCNT3 so the template structure released in 2012 could be used in this case. The modeling procedure used for hCNT3, described in **[H7]**, combined homology modeling with the presence of evolutionary conserved ligands (MODELLER) and *de novo* modeling of transmembrane helices (TOPCONS, Rosetta Broker). CNT3 is a member of the SLC28 protein family including much more complicated biological systems than G protein-coupled receptors in terms of diversity of topology and mechanisms of action. In the case of GPCR, its known mechanism of action involves a local conformational change in the active site resulting from the ligand binding, that induces a global conformational change (mainly helix VI) but without any global changes in topology. In the case of SLCs, ligand transferring from the cell exterior to the cell interior requires such significant conformational change of a transporter that both conformations (outward-open and inward-open) completely differ in terms of topology, but not in terms of secondary structure. Due to an insufficient number of crystal structures of transporters in various transportation cycle states it is not possible to derive specific statistical potentials for them, describing their long-rage interactions, which could be used, e.g., in *de novo* structure modeling. For many years, it was not possible to build a homology model of an outward-open transporter using its inward-open crystal structure. However, L. R. Forrest have described in, e.g., [89], that there is a solution to a complete change of transporter topology during its action cycle. This solution utilizes a concept

of pseudo-symmetry between both conformations of a transporter (so-called 'inverted topology'). Obviously, the solution proposed by L. R. Forrest cannot be used when there is no crystallographic data for any part of a given transporter, or there is no structural data for its fragment, like in the case of hCNT3, whose close homolog in PDB (vCNT) lacks the N-terminal sequence fragment.

In **[H7]** I proposed a solution for modeling of missing regions in SLC transporters structures which were not characterized by crystallographic data. This solution is based on de novo modeling by fragments assembly (Rosetta) and a particular Rosetta protocol for large protein systems (Broker [90]), that implements large protein subunits as rigid bodies. This procedure allowed me to obtain a structure of a more than 100 amino acids-long N-terminal fragment of hCNT3. That N-terminal fragment included, e.g., a functionally important single nucleotide polymorphism (SNP). Modeling of hCNT3 was conducted considering the whole homotrimer structure and interactions of its subunits. The homology model of hCNT3 homotrimer was obtained using vCNT3 with its evolutionary conserved ligands (sodium ion and uridine) as a template. The hCNT3 model validation was performed by a 100 ns-long molecular dynamics simulation **[H7]**. The molecular dynamics simulation allowed to observe: the homotrimer interface, the hCNT3 active site and adjacent helices (i.a. impact of Na^+ ion on a protein conformation during symport) and mobility of amino acids, that were associated with SNPs.

Study of mechanism of activation of GPCR receptors

I used homology models obtained with GPCRM for studying the mechanism of activation of GPCR receptors, that was described in: **[H1] [H3-4]** and in other publications, mentioned in: **5. Description of the remaining scientific and research achievements**. In **[H1]** I described the Trp/Phe amino acid microswitch, whose conformational change is considered as a beginning of local conformational changes of cannabinoid receptors causing the ensuing global conformational changes, that finally lead to functionally active conformations of these receptors interacting with G protein. My study supplemented the former experimental studies on these receptors [83, 91]. In the following years, an increase number of known GPCR structures in PDB allowed to build a more accurate models of cannabinoid receptors with GPCRM, that in turn allowed to study, e.g., the ligand entry pathway to the receptor interior [84].

In a highly cited review publication **[H4]** I described GPCR subfamilies in terms of their conserved sequence motifs and associated amino acid microswitches. A change in a

conformation of these microswitches, e.g., resulting from agonist-receptor interactions, causes decrease of a protein internal energy and thus increase in the active receptor states population [20]. In Table 1, I presented a selected fragment from **[H4]**, describing mechanisms of activation of rhodopsin-like receptors.

Table 1. Amino acid microswitches of class A GPCRs, described in **[H1]** and **[H4]**.

A stage of activation[1]	Microswitches and corresponding sequence motifs	An induced global conformational change of a receptor structure
1	3-7 lock microswitch Glu3.28 (Asp3.32), Lys7.43 (Tyr7.43)	Breaking of a helix III and helix VII connection caused by breaking of a salt bridge between Lys7.43 and a Schiff base (retinal and Glu3.28).
2	Transmission microswitch (former name: toggle microswitch, see: **[H4]**) Trp6.48, P5.50, P6.50 CwxP (helix VI)	Helix VI is moving towards helix V and is slightly rotated and then its cytoplasmic part is moved away.
3	Toggle microswitch Tyr7.53 nPxxy (helix VII)	Regrouping in a cytoplasmic part of helix VII is preceded with a rotation of tyrosine side chain in order to reorganize the hydrogen bonding network inside the receptor, that allows for interaction with G protein.
4	Ionic lock microswitch Glu3.49/Arg3.50 – Glu6.30/Thr6.34 (d/e)Ry (helix III)	Breaking of a salt bridge between arginine and glutamic acid, followed by closeup of helix III with helix V and helix VI with helix V.

[1]The order of activation stages of class A GPCRs was based on Wescott et al. (PNAS 2016, 113(35): 9928-9933). It starts with conformational changes in the extracellular part after the ligand binding, then in the central part of a receptor, and finally in the cytoplasmic part interacting with G protein.

Design of transmembrane protein inhibitors

As I mentioned above, GUT-DOCK **[H9-10]**, can also be used in drug design. It was confirmed by benchmark tests for Autodock VINA (7600 times cited, according to Google Scholar) comparing approximated theoretical binding energy with experimental results confirming the actual ligand binding. As I mentioned in **[H9-10]**, citing another publication, success rates of low-cost virtual screening are comparable and even higher than success rates of experimental screening. Nonetheless, a final confirmation of a biological activity of a given ligand requires a number of biochemical and clinical studies [30]. Another complication is associated with off-target interactions of a given drug with other biological targets in a cell, whose effects are observed only after many years in clinical trials. Some off-target interactions are beneficial, i.e. in the case of drug repositioning. Drug repositioning allows to avoid expensive studies on drug toxicity for novel active compounds. In **[H10]**, using a concept of drug repositioning, I proposed compounds which could be used in pharmacotherapy of diabetes. These compounds have been already used in pharmacotherapy of other diseases but may also exhibit a therapeutic effect on

the incretin hormones signaling pathway, that regulates glucose metabolism. A potential effect of these compounds on glucose metabolism can be associated either with inhibition of GCGR (decrease in hepatic glucose production) or with activation of GLP1R and GIPR receptors (increase in insulin secretion). Nevertheless, to distinguish between these two effects a number of theoretical and experimental studies should be conducted **[H10]**.

I observed an inhibition of a transmembrane protein also in the case of a complex of V-ATPase with archazolid A derivatives **[H2]**, that exhibit an anti-cancer activity [92]. Archazolid A, consisted of a macrocycle lactone ring, exhibits a cytotoxicity effect on cancer cells by stopping the rotation of transmembrane domain of V-ATPase, that is necessary for the H^+ gradient formation (see Fig. 1). Adhesion specificity of this inhibitor while binding the rotor surface is caused not only by the hydrogen binding but also an electrostatic complementary, that was described in Supplement of **[H2]**. My part in this study (*in silico*) was to obtain a homology model of a membrane domain of V-ATPase, to dock archazolid A derivatives and to select ligands of the highest affinity for biochemical studies. Noteworthy, *in silico* results of this study were compared with experimental results, described in **[H2]** (see Fig. 1). A study on anti-cancer activity of archazolids, described in **[H2]**, has been still in progress since 2011, in a laboratory of D. Menche [93].

Molecular bases of drug-induced reactions

One of the functionalities of GUT-DOCK refers to comparison of ligands binding affinities towards selected GPCR receptors **[H9-10]**. In both publications, I described a few common drug classes of a known diabetogenic effect in terms of their binding affinities towards glucagon receptors. The highest binding affinity was observed for statins, β-adrenolytic drugs and glucocorticosteroids, the average affinity for: mineralcorticosteroids and recently launched on the market - thiazides analogs. The lowest binding affinity was observed for: neurosteroids, diuretics and thiazides. In the case of statins and β-blockers I observed that, new-generation pharmaceuticals, of a decreased diabetogenic effect, at the same time exhibit relatively high binding affinities towards glucagon receptors. Glucagon receptors (GCGR, GLP1R, GIPR) regulate the glucose serum level. A similar correlation was observed for a small set of drugs from different drug classes, that was deposited in SIDER (Side Effect Resources) [36] **[H10]**. Based on these results, I proposed a new way to avoid disruption of glucose metabolism by such chemical modifications of lead compounds to increase the binding affinity towards glucagon receptors. As a result, except for the basic therapeutic effect (on-target), an active

compound could also enhance an incretin effect, that regulates glucose serum levels, as an off-target interaction. Such approach is currently implemented in the form of polytherapy, e.g., in the treatment of hypertension (a concurrent therapy with hydrochlorothiazides and sartans [94]). A theoretical study described in **[H9-10]** will be followed by biochemical and clinical studies.

Molecular basis of pharmacologically significant polymorphisms

A structural interpretation of rare, because of a strong negative selection observed for hCNT3 transporter, non-synonymous mutations allowed me to discover molecular basis of diverse response to pharmacotherapies of cancer and viral infections, that were observed in clinical trials. I described it in a publication, in which I was the only author **[H7]**. For example, a Gly↔Arg mutation (see Fig. 4) in the nucleoside binding active site may cause a loss of important interaction of a ligand with adjacent Gln due to a possible Arg-Gln hydrogen bonding. As a consequence, a concentrative capacity of nucleoside molecules by hCNT3 is decreased. Absorption of nucleoside pharmaceuticals (e.g. ribavirin) is decreased and thus its negative, cytotoxic effect on host cells is also decreased. For example, patients with observed Gly↔Arg mutations less often suffer from anemia during the ribavirin treatment of hepatitis C virus infections. A biological effect of the second, described in **[H7]**, polymorphism Tyr↔Cys, that is localized in *de novo* built hCNT3 fragment, remains still inadequately described, though some experimental studies have been started (Badagnani et al. 2005).

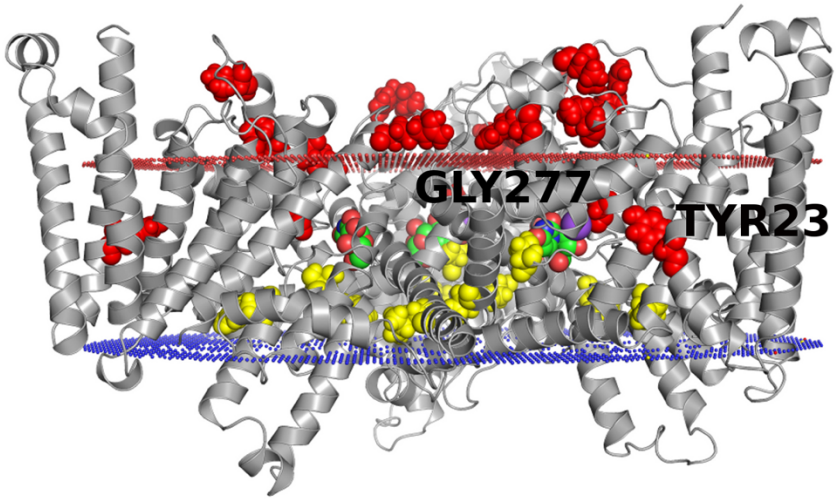

Figure 4. Single nucleotide polymorphisms observed so far for hCNT3 (red – non-synonymous mutations, yellow – synonymous mutations). Two of them were indicated (e.g., Gly277 in the active site region). Also, a sodium ion (violet) and urirdine were indicated.

In **[H7]** I also described a few polymorphisms localized in *de novo* rebuilt helix VI of a GPCR receptor CRF_1. Similarly, to described hCNT3 polymorphisms, described polymorphisms of CRF_1 were associated with decreased ligand binding, that causes diverse response to pharmacotherapy (e.g., treatment of asthma with inhaled corticosteroids). In this case, mutations in the helix VI region may cause decrease in ligand binding due to a loss of important interactions and also a loss of space accessible for an allosteric ligand. The latter effect may be caused by, e.g., with mutations-induced blocking of the helix VI deformation, that is necessary for the allosteric ligand binding **[H7]**.

Summary

The described above scientific achievement regarding determination of structures and mechanisms of action of selected transmembrane proteins considering their evolutionary diversity has allowed to explain in a complete way on the molecular level a number of observed biological and chemical phenomena. A scheme shown in Fig. 5 illustrates main areas of the conducted research.

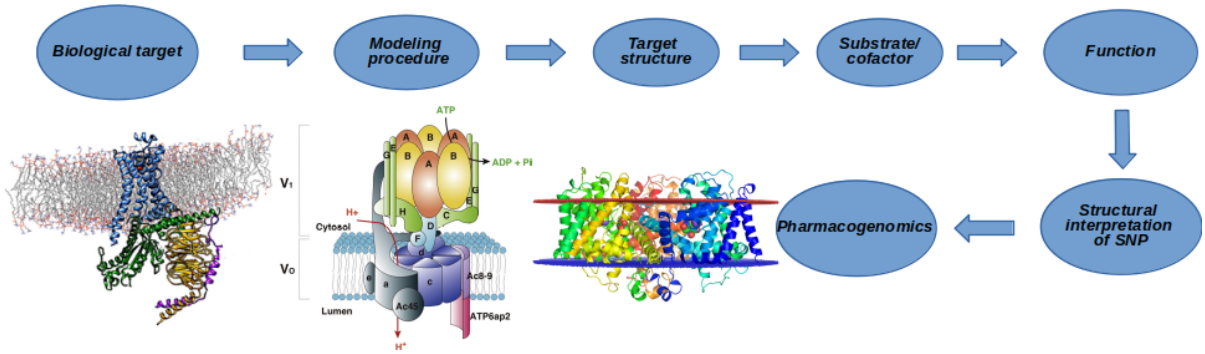

Figure 5. The scheme of the scientific achievement, that was described in the current summary. In my research I performed a complete description of structures and mechanisms of action of members of three selected transmembrane protein families (from left: GPCR receptors, V-ATPase, SLC28 transporters) considering their evolutionary diversity. Based on: Clark et al. Beilstein J. Org. Chem. 2017, 13, 1071–1078, Sun-Wada et al. Biochimica et Biophysica Acta (BBA) – Bioenergetic 2015, 1847(10):1166-72 oraz **[H7]**.

The described achievement allowed to develop a new computational methodology, that could be used to study structures of transmembrane proteins. Thus, it became possible to develop a web application, that allows to test various compounds similarly to experimental testing and allows to achieve the big data level. A particular result of my research was to ensure the repeatability standard, that is necessary for freely accessible web applications but is difficult to

obtain for heuristic methods which are a large part of computational methods. In my opinion, my research on structure and mechanisms of action of transmembrane proteins was inevitable to explain reasons for their evolutionary diversity. In my research I chose to be an observer of examined objects, despite the ongoing research, e.g., on directed evolution of GPCR receptors to modify their function [95].

VI. References

[1] Buller AR, van Roye P, Cahn JKB, Scheele RA, Herger M, Arnold FH (2018) Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. JACS 140 (23), 7256-7266

[2] Jones DT (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics 23(5): 538–544

[3] Kufareva I, Katritch V, Participants of GPCR Dock 2013, Stevens RC, Abagyan R (2014) Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. Structure 22(8):1120-1139

[4] César-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, Reithmeier RA, Hepworth D, Hediger MA, Edwards AM, Superti-Furga GA (2015) Call for Systematic Research on Solute Carriers. Cell 62(3):478-87

[5] Hinton A, Bond S, Forgac M (2009) V-ATPase functions in normal and disease processes. Pflugers Arch. 457(3):589-98

[6] Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291-325.

[7] Woetzel N, Karakaş M, Staritzbichler R, Müller R, Weiner BE, Meiler J (2012) BCL::Score—Knowledge Based Energy Potentials for Ranking Protein Models Represented by Idealized Secondary Structure Elements. PLOS ONE 7(11): e49242.

[8] Wang DG et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280(5366):1077-82.

[9] Evans WE, McLeod HL (2003) Pharmacogenomics--drug disposition, drug targets, and side effects. N Engl J Med 348(6):538-49.

[10] Lin L, Yee SW, Kim RB, Giacomini KM (2015) SLC transporters as therapeutic targets: emerging opportunities. Nat Rev Drug Discov 14(8):543-60

[11] Denisov IG, Sligar SG (2016) Nanodiscs for Structural and Functional Studies of Membrane Proteins. Nature Structural & Molecular Biology 23(6), 481-486

[12] Barth P, Schonbrun J, Baker D (2007) Toward high-resolution prediction and design of transmembrane helical protein structures. PNAS 104(40):15682-7

[13] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem 55(14):6582-94

[14] Fan H, Irwin JJ, Webb BM, Klebe G, Shoichet BK, Sali A (2009) Molecular docking screens using comparative models of proteins. J Chem Inf Model 49(11):2512-27

[15] Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE (2017) Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Discov 6(12):829-842.

[16] Hendrickson WA (2016) Atomic-level analysis of membrane-protein structure. Nat Struct Mol Biol 23(6):464-7

[17] Cheng Y (2015) Single-Particle Cryo-EM at Crystallographic Resolution. Cell 161(3):450-457

[18] Rossmann MG, Blow DM (1962) The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst. 15, 24-31.

[19] Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289(5480):739-45

[20] Manglik A, Kobilka BK, Steyaert J (2017) Nanobodies to Study G Protein-Coupled Receptor Structure and Function. Annu Rev Pharmacol Toxicol. 57:19-37

[21] Rasmussen SG et al. (2011) Crystal structure of the β2 adrenergic receptor-Gs protein complex. Nature 477(7366):549-55

[22] Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A, Palczewski K (2003) Atomic-force microscopy: Rhodopsin dimers in native disc membranes. Nature 421(6919):127-8.

[23] Robertson N, Jazayeri A, Errey J, Baig A, Hurrell E, Zhukov A, Langmead CJ, Weir M, Marshall FH (2011) The properties of thermostabilised G protein-coupled receptors (StaRs) and their use in drug discovery. Neuropharmacology 60(1):36-44

[24] Liang YL et al. (2018) Cryo-EM structure of the active, Gsprotein complexed, human CGRP receptor. Nature 561(7724):492-497

[25] Warne T, Moukhametzianov R, Baker JG, Nehmé R, Edwards PC, Leslie AG, Schertler GF, Tate CG (2011) The structural basis for agonist and partial agonist action on a β(1)-adrenergic receptor. Nature 469(7329):241-4.

[26] Zhao J, Benlekbir S, Rubinstein JL (2015) Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. Nature 521(7551):241-5.
[27] Roh SH, Stam NJ, Hryc CF, Couoh-Cardel S, Pintilie G, Chiu W, Wilkens S (2018) The 3.5-Å CryoEM Structure of Nanodisc-Reconstituted Yeast Vacuolar ATPase Vo Proton Channel. Mol Cell 69(6):993-1004.e3.
[28] Johnson, Zachary Lee et al. (2014) Structural basis of nucleoside and nucleoside drug selectivity by concentrative nucleoside transporters. eLife vol. 3 e03604.
[29] Hirschi M, Johnson ZL, Lee SY (2017) Visualizing multistep elevator-like transitions of a nucleoside transporter. Nature 545(7652):66-70.
[30] Jacobson KA (2015) New paradigms in GPCR drug discovery. Biochem Pharmacol 98(4):541-55.
[31] Merkel JS, Michaud GA, Salcius M, Schweitzer B, Predki PF (2005) Functional protein microarrays: just how functional are they? Curr Opin Biotechnol 16(4):447-52.
[32] Meyer UA (2000) Pharmacogenetics and adverse drug reactions. Lancet 356(9242):1667-71
[33] Yang M, Kiang M, Shang W (2015) Filtering big data from social media--Building an early warning system for adverse drug reactions. J Biomed Inform 54:230-40
[34] Zhou H, Gao M, Skolnick J (2015) Comprehensive prediction of drug-protein interactions and side effects for the human proteome. Sci Rep 5:11090.
[35] Obermeyer Z, Emanuel EJ (2016) Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 375(13):1216-9.
[36] Kuhn M, Letunic I, Jensen LJ, Bork P (2015) The SIDER database of drugs and side effects. Nucleic Acids Res. doi: 10.1093/nar/gkv1075
[37] Kolinski A (2004) Protein modeling and structure prediction with a reduced representation. Acta Biochim Pol 51(2):349-71
[38] K Ginalski, A Elofsson, D Fischer, L Rychlewski (2003) 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 19 (8), 1015-1018
[39] Kurowski MA, Bujnicki JM (2003) GeneSilico protein structure prediction meta-server. Nucleic Acids Research 31 (13), 3305-3307
[40] Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. Methods Enzymol 383:66-93
[41] Eswar N, Webb B, Marti-Renom MA, et al. (2006) Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics. Chapter 5:Unit-5.6.

[42] Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18(15):2714-23.
[43] Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS (2012) Three-dimensional structures of membrane proteins from genomic sequencing. Cell 22;149(7):1607-21
[44] Nugent T, Jones DT (2012) Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis. PNAS 109(24):E1540-7.
[45] Khafizov K, Staritzbichler R, Stamm M, Forrest LR (2010) A Study of the Evolution of Inverted-Topology Repeats from LeuT-Fold Transporters Using AlignMe. Biochemistry 49 (50), 10702–10713
[46] Woetzel N, Karakaş M, Staritzbichler R, Müller R, Weiner BE, Meiler J (2012) BCL::Score--knowledge based energy potentials for ranking protein models represented by idealized secondary structure elements. PLOS One 7(11):e49242
[47] Yarov Yarovoy V, Schonbrun J, Baker D (2006) Multipass membrane protein structure prediction using Rosetta. Proteins: Structure, Function, and Bioinformatics 62 (4), 1010-1025
[48] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem., 4: 187-217
[49] Case DA, Cheatham TE, Darden T, et al. (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16):1668-88
[50] Lazaridis T (2003) Effective energy function for proteins in lipid membranes. Proteins 52(2):176-92.
[51] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free J Comput Chem 26(16):1701-18.
[52] Phillips JC, Braun R, Wang W, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781-802
[53] Miao Y, McCammon JA (2016) G-protein coupled receptors: advances in simulation and drug discovery. Curr Opin Struct Biol 41:83-89
[54] Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267(3):727-48
[55] Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779-815

[56] Larsson P, Wallner B, Lindahl E, Elofsson A (2008) Using multiple templates to improve quality of homology models in automated homology modeling. Protein science: a publication of the Protein Society 17: 990–1002.
[57] Porter JR, Weitzner BD, Lange OF (2015) A framework to simplify combined sampling strategies in Rosetta. PLOS One 10(9):e0138220.
[58] Gribskov M, McLachlan AD, Eisenberg D (1987) Profile analysis: detection of distantly related proteins. PNAS 84(13):4355-8
[59] Rychlewski L, Jaroszewski L, Li W, Godzik A (2000) Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Sci 9(2):232-41
[60] SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman (1990) Basic local alignment search tool. Journal of Molecular Biology 215 (3), 403-410
[61] Kufareva I, Katritch V, Participants of GPCR Dock 2013, Stevens RCS, Abagyan R (2014) Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. Structure 22(8), 1120 – 1139
[62] Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455-61
[63] Krieger E, Vriend G (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics 30(20):2981-2.
[64] Bernsel A, Viklund H, Hennerdal A, Elofsson A (2009) TOPCONS: consensus prediction of membrane protein topology. NAR 37(Web Server issue):W465-8.
[65] Larkin MA et al. (2007) Clustal W and Clustal X version 2.0, Bioinformatics 23 (21) 2947–2948
[66] Srinivasan N, Blundell TL (1993) An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. Protein Engineering, Design and Selection 6(5), 501–512
[67] Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792-7
[68] Messih MA, Lepore R, Tramontano A (2015) LoopIng: a template-based tool for predicting the structure of protein loops. Bioinformatics 31(23):3767-72
[69] Martí-Solano M, Iglesias A, de Fabritiis G, Sanz F, Brea J, Loza MI, Pastor M, Selent J (2015) Detection of New Biased Agonists for the Serotonin 5-HT2A Receptor: Modeling and Experimental Validation. Mol Pharm 87(4), 740-746

[70] Vroling B, Sanders M, Baakman C, et al. (2010) GPCRDB: information system for G protein-coupled receptors. Nucleic Acids Res 39(Database issue):D309-19
[71] Worth CL, Kreuchwig A, Kleinau G, Krause G (2011) GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models. BMC Bioinformatics 12:185
[72] Rodríguez, D.; Bello, X; Gutiérrez-de-Terán H (2012) Modeling G Protein-Coupled Receptors Through the Web. Mol Inf 31:334–41
[73] Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40
[74] Abrol R, Bray JK, Goddard WA (2011) Bihelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. Proteins 80(2):505-18
[75] Pándy-Szekeres G, Munk C, Tsonkov TM, et al. (2017) GPCRdb in 2018: adding GPCR structure models and ligands. Nucleic Acids Res 46(D1):D440-D446
[76] Worth CL, Kreuchwig F, Tiemann JKS, et al. (2017) GPCR-SSFE 2.0-a fragment-based molecular modeling web tool for Class A G-protein coupled receptors. Nucleic Acids Res 45(W1):W408-W415
[77] Zhang J, Yang J, Jang R, Zhang Y (2015) GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome. Structure 23(8):1538-1549
[78] Murata T, Yamato I, Kakinuma Y, Leslie AG, Walker JE (2005) Structure of the rotor of the V-Type Na+-ATPase from Enterococcus hirae. Science 308(5722):654-9
[79] Zhao J, Benlekbir S, Rubinstein JL (2015) Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. Nature 14;521(7551):241-5
[80] Roh SH, Stam NJ, Hryc CF, Couoh-Cardel S, Pintilie G, Chiu W, Wilkens S (2018) The 3.5-Å CryoEM Structure of Nanodisc-Reconstituted Yeast Vacuolar ATPase Vo Proton Channel. Mol Cell 69(6):993-1004.e3
[81] Hua T, Vemuri K, Nikas SP, et al. (2017) Crystal structures of agonist-bound human cannabinoid receptor CB1. Nature 547(7664):468-471.
[82] Li X et al. (2019) Crystal Structure of the Human Cannabinoid Receptor CB2. Cell 176(3):459-467.e13
[83] McAllister SD et al. (2003) An aromatic microdomain at the cannabinoid CB(1) receptor constitutes an agonist/inverse agonist binding region. J Med Chem 46(24):5139-52
[84] Jakowiecki J, Filipek S (2016) Hydrophobic Ligand Entry and Exit Pathways of the CB1 Cannabinoid Receptor. J Chem Inf Model 56(12):2457-2466

[85] Shim JY, Bertalovitz AC, Kendall DA (2011) Identification of essential cannabinoid-binding domains: structural insights into early dynamic events in receptor activation. J Biol Chem 286(38):33422-35
[86] Pandey P, Roy KK, Doerksen RJ (2019) Negative allosteric modulators of cannabinoid receptor 2: protein modeling, binding site identification and molecular dynamics simulations in the presence of an orthosteric agonist. J Biomol Struct Dyn 5:1-16
[87] Zhang H, Qiao A, Yang D, et al. (2017) Structure of the full-length glucagon class B G-protein-coupled receptor. Nature 546(7657):259-264
[88] Johnson ZL, Lee JH, Lee K, et al. (2014) Structural basis of nucleoside and nucleoside drug selectivity by concentrative nucleoside transporters. Elife 3:e03604
[89] Forrest LR (2013) Structural biology. (Pseudo-)symmetrical transport. Science 339(6118):399-401
[90] DiMaio F, Leaver-Fay A, Bradley P, Baker D, André I (2011) Modeling Symmetric Macromolecular Structures in Rosetta3. PLoS ONE 6(6): e20450
[91] McAllister SD, Hurst DP, Barnett-Norris J, Lynch D, Reggio PH, Abood ME (2004) Structural mimicry in class A G protein-coupled receptor rotamer toggle switches: the importance of the F3.36(201)/W6.48(357) interaction in cannabinoid CB1 receptor activation. J Biol Chem 279(46):48024-37
[92] Pérez-Sayáns M1, Somoza-Martín JM, Barros-Angueira F, Rey JM, García-García A (2009) V-ATPase inhibitors and implication in cancer treatment. Cancer Treat Rev 35(8):707-13
[93] Scheeff S, Menche D (2017) Total syntheses of the archazolids: an emerging class of novel anticancer drugs. Beilstein J Org Chem 13:1085-1098
[94] Sowers JR, Raij L, Jialal I, et al. (2010) Angiotensin receptor blocker/diuretic combination preserves insulin responses in obese hypertensives. J Hypertens 28(8):1761-9
[95] Keri D, Barth P (2018) Reprogramming G protein coupled receptor structure and function. Curr Opin Struct Biol 51:187-194

5. Description of the remaining scientific and research achievements

a) Summary of professional accomplishments

A number of publications	**19**
Total impact factor according to Journal Citation Reports and the publishing year	**81.443**
A total number of citations according to Web of Science (on 08/04/2019)	**659**
A number of citations according to Web of Science without self-citations (on 08/04/2019)	**621**
Hirsch index for all 19 publications according to Web of Science (on 08/04/2019)	**11**

In my current research, that was not described in: **4. Description of the achievement**, I performed structure modeling of transmembrane proteins and their small-molecules and peptides complexes. It was described in publications: **[B1-4]** and **[B6]**, **[B8]**. Publications released during my PhD studies **[B7-9]** and shortly after **[B5]**, were dedicated only to globular proteins and were the methodology basis for my latter research on membrane proteins.

The main achievement, that was the central part of my PhD thesis, referred to the development of an algorithm for structure prediction of large (more than 150 residues) globular proteins using sparse NMR data **[B5]** and **[B8-9]**. What is more, I developed an algorithm for using evolutionary information, in the form of long-range inter-residue contacts, for modeling of globular proteins structures **[B7]**.

The most important achievements after the PhD studies, except those described in **4. Description of the achievement**, involved studying serotonin receptors and class F GPCR (SMO receptor) in 2013 and studying dopamine D3 and chemokine CXCR4 receptors in 2010, described in **[B1]** and **[B6]**, respectively. These achievements were focused on the determination of the interaction mode of the mentioned above receptors with small-molecule ligands: ergotamine (5-HT_{1B} and 5-HT_{2B}), SANT-1 and LY-2940680 (SMO), eticlopride (D3) and an isothiourea derivative IT1t and a cyclic peptide analog CVX15 (CXCR4). These studies involved phylogenetics analysis of GPCR receptors and their homologous sequences, structure modeling of receptors and their complexes with ligands and determination of un-bound and receptor-bound ligand conformations. Among others, studying of interactions of ergotamine with both serotonin receptors allowed to determine the molecular basis of their diverse

functional selectivity profile **[B1]**. The described above studies allowed to succeed in the international competition GPCR Dock 2013 **[B1]** (our group – Warsaw was ranked 1st in all 44 research groups) and GPCR Dock 2010 **[B6]** (our group – Warsaw was ranked 6th in all 34 research groups). In both competitions, researchers aimed to determine structures of selected class A and F GPCRs together with their small-molecule ligands binding modes. The only data provided was a receptor amino acid sequence and a chemical formula of its ligand. The mentioned above competitions, similarly to CASP (theoretical determination of various proteins, mostly globular ones) and CAPRI (theoretical determination of protein-protein complexes structures), assess critically the accuracy of the computational methods used by many research groups. In my opinion, both competitions GPCR Dock 2010 and GPCR Dock 2013, organized at the beginning and end of my research regarding GPCR receptors, finalized with two web applications, summarize the development of my computational methodology.

The remaining achievements, that were described in publications **[B2]** and **[B4]** refer to studying of the mechanisms of activation of two class A GPCRs: formyl receptor FPR1 and lipid receptor S1P1. In the case of FPR1 **[B4]** a described mechanism of activation involved a toggle microswitch (see Table 1), while in the case of S1P1 it involved a transmission switch, that redirected the water molecules flow involved in the hydrogen bonding inside the receptor.

The last but not the least achievement was described in **[B3]**. In **[B3]** and also in my former publication **[B5]** I described studying of globular proteins by nuclear resonance methods supplemented with computational methods. In both studies, I used *de novo* methods, that did not require template structures from PDB. Consequently, both publications supplemented my research methodology with other computational methods, that were not mentioned in **4. Description of the achievement**. In **[B3]**, that was published in Journal of American Chemical Society, I described my research conducted during a short five-months post-doc in EMBL Heidelberg. It involved usage of NMR data in determination of ligand binding mode (INPHARMA). My part of this study was focused on development of an algorithm prototype, that extensively performed conformational search of a ligand-protein complex. This algorithm was the central module in the data analysis of the INPHARMA method. INPHARMA allows to determine ligand binding modes with NMR without determination of the whole protein structure.

b) List of scientific publications that constitute all scientific achievements (excluding publications mentioned in: 4. Description of the achievement) and were published in journals indexed by Journal Citation Reports.

- after obtaining a PhD in chemistry degree

Item no.	Author / authors, publishing year, publication title, journal name, volume, pages	Impact Factor[1]	Times cited[2]
B1	Kufareva I, Katritch V, **Participants of GPCR Dock 2013**, Stevens RCS*, Abagyan R* (2014) *Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges.* **Structure**, 22(8), 1120 – 1139	5.618	92
B2	Yuan S*, Wu R, **Latek D**, Trzaskowski B, Filipek S* (2013) *Lipid Receptor S1P1 Activation Scheme Concluded from Microsecond All-Atom Molecular Dynamics Simulations.* **PLOS Comput Biol**, 9(10), e1003261	4.829	17
B3	Codutti L, Skjaerven L, Angelini A, Grimaldi M, **Latek D**, Monecke P, Dreyer M & Carlomagno T* (2013) *Accounting for conformational variability in protein-ligand docking with NMR-guided rescoring.* **J Am Chem Soc**, 135(15), 5819-27	11.444	17
B4	Yuan S, Ghoshdastider U, Trzaskowski B, **Latek D**, Debinski A, Pulawski W, Wu R, Gerke V & Filipek S* (2012) *The role of water in activation mechanism of human N-formyl peptide receptor 1 (FPR1) based on molecular dynamics simulations.* **PLOS ONE**, 7, e47114	3.730	15
B5	**Latek D*** & Kolinski A (2011) *CABS-NMR--De novo tool for rapid global fold determination from chemical shifts, residual dipolar couplings and sparse methyl-methyl NOEs.* **J Comput Chem**, 32, 536-544	4.583	7
B6	Kufareva I, Rueda M, Katritch V, **GPCR Dock 2010 participants**, Stevens RCS*, Abagyan R* (2011) *Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment.* **Structure**, 19, 1108-1126	6.347	195

* - corresponding author

[1]according to the publishing year

[2]according to Web of Science, date: 08/04/2019

- before obtaining a PhD in chemistry degree

Item no.	Author / authors, publishing year, publication title, journal name, volume, pages	Impact Factor[1]	Times cited[2]
B7	**Latek D*** & Kolinski A (2008) *Contact prediction in protein modeling: scoring, folding and refinement of coarse-grained models.* **BMC Struct Biol**, 8, 36	2.000	10
B8	**Latek D***, Ekonomiuk D & Kolinski A (2007) *Protein structure prediction: combining de novo modeling with sparse experimental data.* **J Comput Chem**, 28, 1668-1676	4.297	20
B9	**Plewczynska D** & Kolinski A* (2005) *Protein Folding with a Reduced Model and Inaccurate Short-Range Restraints.* **Macromol Theory Sim**, 14, 444-451	1.544	8

* - corresponding author

[1]according to the publishing year

[2]according to Web of Science, date: 08/04/2019

c) Scientific monographs

Item no.	Author / authors, publishing year, publication title, publication series title, editor, publisher, volume, pages
1	**Latek D**, Trzaskowski B, Niewieczerzal S, Miszta P, Mlynarczyk K, Debinski A, Pulawski W, Yuan S, Sztyler A, Orzeł U, Jakowiecki J, Filipek S (2019) *„Modeling of Membrane Proteins: From Bioinformatics to Molecular Quantum Mechanics."* w *„Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes From Bioinformatics to Molecular Quantum Mechanics"*, Liwo A, Springer, vol. 8, 357-431.
2	Miszta P, Jakowiecki J, Rutkowska E, Turant M, **Latek D**, Filipek S (2018) *„Approaches for Differentiation and Interconverting GPCR Agonists and Antagonists"* w *"Computational Methods for GPCR Drug Discovery. Methods in Molecular Biology"*, Heifetz A, Humana Press, New York, NY, vol. 1705, 265-296
3	Rutkowska E, Miszta P, Mlynarczyk K, Jakowiecki J, Pasznik P, Filipek S, **Latek D** (2017) *„Application of a Membrane Protein Structure Prediction Web Service GPCRM to a Gastric Inhibitory Polypeptide Receptor Model"* w *„Bioinformatics and Biomedical*

	Engineering. IWBBIO 2017. Lecture Notes in Computer Science", Rojas I, Ortuño F, Springer, Cham, vol. 10209, 151-162.
4	**Latek D**, Trzaskowski B, Niewieczerzal S, Miszta P, Mlynarczyk K, Debinski A, Pulawski W, Yuan S, Filipek S (2014) *„Modeling of Membrane Proteins"* w *„Computational methods to study the structure and dynamics of biomolecules and biomolecular processes – from bioinformatics to molecular quantum mechanics"*, Liwo A, Springer Series in Bio-/Neuroinformatics, vol. 1, 357–431.
5	**Latek D** & Filipek S (2012) *„Automatic template-based model generation of G-protein coupled receptors"* w *„Workshop 2012: From Computational Biophysics to Systems Biology"*, Carloni P et al., Publication Series of the Institute for Advanced Simulation, IAS Series, vol. 8, 111-114.
6	Gront D, **Latek D**, Kurcinski M, Kolinski A (2008) *„Template-Free Predictions of Three-Dimensional Protein Structures: From First Principles to Knowledge-Based Potentials"* w *„Prediction of Protein Structures, Functions, and Interactions"*, Bujnicki J, John Wiley & Sons, Ltd., 117-141.
7	Gront D, **Latek D**, Kurcinski M, Kolinski A (2008) *„Wieloskalowe modelowanie bialek"* w *„Na pograniczu chemii i biologii"*, Koroniak H, Barciszewski J, Wydawnictwo Naukowe Uniwersytetu Adama Mickiewicza, vol. 17, 239-256.
8	Kolinski A, Gront D, Kmiecik K, Kurciński M & **Latek D** (2006) *„Modeling protein structure, dynamics and thermodynamics with reduced representation of conformational space"* w *„NIC Workshop 2006: From Computational Biophysics to System Biology"*, Meinke J et al., John von Neumann Institute for Computing, Juelich, NIC Series, vol. 34, 21-28.