Summary of the PhD dissertation

Knowledge about a molecular mechanism of living cells does not satisfy a curiosity of scientists, only, but also has influence on a development of new therapies and drugs against various pathological processes. Discovery of the three dimensional structure of a protein involved in such process is a necessary prerequisite. Experimental methods are expensive, timeconsuming, and depend on the experience of specialists. The computational methods can be helpful in the study of protein structures. However, correct results are not always returned by the computational tools used for modelling of protein structures. Therefore, a protein structure model has to be properly assessed. For this purpose data about known protein structures may be used, for example: to validate, if the modelled structure is similar to any of known proteins. Also the interactions between atoms can be calculated to estimate if the model is energetically stable in nature. Unconventional idea is to look at the protein modeling process backward and verify, if the model functions correctly. The main goal of the PhD study was to optimize methods that choose the best protein model out of its options with emphasis on the transmembrane proteins.

In the first part of the studies the first thesis, proposing that "usage of the computational functional methods in quality assessment of the transmembrane protein rises the efficiency of the optimal model selection, characterized with appropriated functionality" was verified. For this purpose 7,500 of protein models were studied. Their structural features were analyzed in comparison to the current-voltage characteristics, calculated with Poisson-Nerst-Planck model (PNP). Additionally, the electrostatic profiles inside the channels were used. For structural assessment the Root Mean Square Deviation (RMSD) was used. In paper Konopka et al., 2014 I showed that using electrostatic profiles inside the channel (Root Mean Square Error, RMSE) as additional metric, allows for better distinguishing between models. RMSE highlighted especially the structural errors inside the pore which were not seen in global assessment, while they are really important in ion transport through the protein channel. In paper Dyrka et al., 2016 I analyzed the relation between structure, electrostatic and function of the model. The usage of functional parameters significantly narrowed the set of the models which were assessed as correct. Computed conductance and selectivity of the protein models indicated the globally optimal models (with low RMSD) with correct electrostatic potential inside the channel (with low RMSE). On the other hand, the incorrect values of the functional parameters indicated the subset of the models with low RMSD, but with small changes in the structures. The most important achievement of this study was a separation of the models with small structural changes with significant impact on the correct functionality of the protein. In that way the first thesis was proved.

The second thesis said that "there are energy terms of the total energy, which enable differentiating between native and mirror protein models obtained from contact maps". Modeling of unknown protein structures is improved by a development of the modelling methods based on residue-residue contacts. The residue-residue contact is usually defined as a pair of amino-acids with the distance between Cα or Cβ atoms less than 8 Å,

and 2-dimensional representation of the residue-residue contacts in protein is called a contact map. However, the contact maps do not include information about chirality of the protein, therefore the reconstruction procedure may prepare two type of the results: native and mirror models. Nevertheless, the mirror models may be stable and competitive conformations of the native proteins in nature depending on the external conditions. To solve the problem of distinguishing between native and mirror models, first the models of the α-helices proteins were studied. Their chirality is easier distinguished in visual assessment than β-sheets or more complicated shape of the protein. In paper Kurczyńska et al., 2016 I proposed the systematic analyses of more than 1000 protein models. In the study I showed that the structural features without the knowledge about related, collected structures are not sufficient to differentiate between native and mirror models. I also showed, that the total energy of native and mirror models are not always statistically different. The number of the energy terms that were significantly different for native and mirror models depended on a quality of the reconstructed models. Finally, the most reliable indicators for distinguishing between native and mirror models were energy terms described a probability of the occurrence of amino acid at defined values of dihedral angles and their Ramachandran preferences. In the next step the study has been extended on the protein domains from different structural classes. The calculations were conducted for more than 130,000 of models. In paper Kurczyńska and Kotulska, 2018 I showed which energy terms are most differentiating for native and mirror models for all types of the protein structures. I also proposed the automated method for differentiation between native and mirror models based on their energy terms. Therefore, with k-means clustering and differentiating energy terms, the second thesis was verified.