Paweł Krupa

Summary of the PhD thesis

Research performed during my PhD studies is described in my PhD thesis "Extension of the UNRES force field, with local potentials and information from comparative modeling methods, to enhance protein-structure predictive power" and consists of four stages:

- Estimation of the prediction power of the UNRES force field in CASP10 experiment
- Implementation of the local side-chain – backbone correlation potentials into UNRES force field based on:
  - Statistical analysis of the loop fragments in the PDB database
  - Physicochemical calculations by semiempirical AM1 method
- Implementation of the constraints based on template-based models into UNRES force field

My contribution to the wfCPUNK group within the WeFold project was to perform a series of the multiplexed replica exchange molecular dynamics simulations (MREMD) using UNRES force field supplemented with the residue-contact and secondary structure information provided by other participating groups. Obtained models of the structures, in particular for target T0740, which was the best prediction for this protein among all predictions done in CASP10, revealed the advantage of using in simulations information from other bioinformatics methods.

To improve the quality of the predicted models of proteins we decided to implement to the UNRES force field new local side-chain - backbone correlating potentials. For this purpose we used the statistical analysis of the loops and other weekly formed fragments from 4585 proteins from the PDB database. Because the used number of the loops did not provide necessary number of the combination of each of the amino-acid pairs, we adopted the simplified amino-acid alphabet, dividing twenty standard amino-acid residues into 5 groups, based on their physicochemical properties. Obtained potentials of the mean force (PMFs) were implemented into the UNRES force field using the Fourier series. Then, I performed the optimization of the weights of the new potential on the set of 8 test proteins. The analysis of the performed MREMD simulations with the optimal value of the weight of the new potential, equal to 0.57, shows the significant improvement of the obtained structures. The mean RMSD value decreased by 0.5 Å in comparison to the simulations without new potentials. The biggest improvement was observed for the loops and other weakly defined elements of the proteins.

We decided to derive the analogical potentials with the use of the physicochemical methods instead of the statistical analysis. For this purpose we adopted previously performed

AM1 calculations of the blocked amino-acid residues by Kozłowska et al., which we extended for the proline. Then we calculated the PMFs and implemented the potentials into the UNRES force field. During the optimization of the weights of the new potentials we observed that they are overlapping with the existing torsional potential. For that reason I have to optimize these weight together. I used the set of the 8 proteins, previously used to test the statistical potentials, as the training set. Within 12 tested sets of the weights, one of them shows the best results – the new potentials with the weight equal to 0.25 and the old torsional potential weight decreased by 0.50. The observed RMSD improvement for the training set was equal to 0.41 Å. Then I performed the additional tests on the 22 peptides and proteins with the range of the 12 to 126 amino-acid residues. The analysis of the obtained results showed the improvement of the lowest RMSD value by 0.33 Å (from 4.19 to 3.86 Å) and the improvement of the RMSD of the representative structure of the best from five clusters, obtained using Ward's minimum variance clustering method, by 0.86 Å (from 7.80 to 6.94 Å).

After the new correlation potentials were implemented into the UNRES force field, I also implemented the new constraints based on the MODELLER energy function, enabling to use in UNRES simulations information of the local structure of the protein from i.e. models obtained by the template-based modeling methods. To assess the performance of the method, I performed tests using five proteins from CASP9 experiment, modeled by the MODELLER software, and four two-domain proteins from CASP10 experiment, for which the structure of the domains was predicted correctly, but their orientation was improper. Performed tests on the single-domain proteins showed that obtained structures are comparable quality to the best of used models. However, tests of the multi-domain proteins, on which constraints were placed only within domains, not between them, revealed the biggest advantage of the UNRES force field, which is the ability to correctly pack the secondary and tertiary structure elements with each other. For three out of four tested two-domain proteins, domains were oriented better than in the used models and better than in any of the models submitted during CASP10 experiment.