Abstract

Proteins are essential for proper physiological functioning of organisms, and their functions are largely determined by the amino acid sequence, which governs not only proteins three-dimensional structure but also their conformational flexibility and dynamic behavior in specific environments. Protein flexibility is a crucial factor that influences the way protein structure is correlated with function. Therefore, identifying flexible regions within these macromolecules is essential to understand their biological functions. Experimental investigation of protein flexibility is often difficult or even impossible, which makes computational approaches particularly valuable. Computations using classical atomic resolution modeling tools allow for mapping all the atoms of the studied system, which requires large computational cost. In the 1970s, coarse-grained models have been developed that allow for reducing this time by grouping several atoms as one interacting center, allowing for computational speed-up.

In this doctoral thesis, computer simulations I carried out for a test set of 100 proteins using selected coarse-grained models (UNRES-flex, UNRES-DSSP-flex, CABS-flex and NOLB) to evaluate their accuracy in predicting the flexibility of amino acid chain fragments based on experimental structures (obtained from nuclear magnetic resonance spectroscopy and X-ray crystallography). Based on the computational results, I calculated standard deviations of fluctuations, followed by Pearson and Spearman correlation coefficients with respect to the experimental fluctuation values. Using average correlation coefficients, I identified which coarse-grained method most accuratly predicts protein flexibility depending on the type of secondary structure (α, β, α + β) and the structure determination method (NMR ensembles or X-ray crystallography). I analyzed statistical significance using the Student's t-test and two-way ANOVA. In addition, the correlation coefficients I analyzed with respect to the numbers of count, the skewness, and the cumulative distribution functions, as well as the relationship between the size of the protein and the correlation values.

In the next part, I carried out simulations using the semi-empirical PM7 method and the ff14SB all-atom force field within the AMBER package obtaining a hypersurface of potential energy, based on which I determined the potentials of mean force and then fitted them with analytical functions for the studied systems between molecules from the UNRES model and molecules from the

MARTINI model. Analytical functions I fitted to the resulting PMF, comparing interactions between two hydrophobic, two polar, and two charged molecules in the UNRES and MARTINI models to determine which method more accurately represents specific interaction types. Furthermore, protein flexibility simulations in a lipid membrane environment were performed using the noew UNRES-MARTINI force field and compared to the existing UNRES model.

In the last part, I was using all-atom force fields ff14SB/GLYCAM06j (AMBER) and CHARMM36m simulations to assess which method better captures the influence of glycosaminoglycans (GAG) on the flexibility of GAG-bound in complex proteins and compared with unbound proteins, based on fluctuation profiles.

The obtained results allowed me to identify which coarse-grained method most effectively reproduce protein flexibility depending on the type of secondary structure and the structural determination method. I also comparisoned PM7 method with ff14SB in terms of simulating various types of interaction centers in UNRES-MARTINI. I also helped to determine which all-atom force field best reflects the influence of GAGs on protein flexibility. These insights can significantly enhance the integration of protein flexibility into drug design workflows using computationally efficient theoretical methods.