ABSTRACT

In computer-aided drug discovery, methods that are approximate, but computationally inexpensive play an essential role during the initial phase of the discovery process. Although often inaccurate, they enable the screening of vast drug libraries to identify potential inhibitors with favorable activities, before large amounts of computational resources could be dedicated to studying these individual molecules. This thesis presents such an approach, based on the concept of hydrogen bond wrapping, to study protein-ligand interactions in the context of drug discovery. The 'wrapping' refers to the tendency of hydrophobic groups to surround a hydrogen bond in water, leading to its desolvation, thereby stabilizing it.

Herein, a molecular descriptor was employed, which quantifies the extent of hydrophobic wrapping around a protein's backbone hydrogen bonds (BHBs) and could help speed up the discovery process by providing cues for the design or optimization of inhibitors. Additionally, these insights could help tailor not just the binding affinity of inhibitors, but also their specificity toward an intended target protein. The human transmembrane protease serine 2 (TMPRSS2) was used as an illustrative target protein due to the pressing need for COVID-19 therapeutics, and since the current understanding of the binding mechanisms of known TMPRSS2 inhibitors is limited.

Molecular docking with a Generalized Born - surface area (GBSA) scoring function was first performed to virtually screen for TMPRSS2 inhibitors. The molecular descriptor was then used to analyze the change in wrapping groups of TMPRSS2 BHBs due to docked ligands, with the aim of identifying BHBs with a high propensity for desolvation. The BHBs involving residues Cys437, Gln438, Asp440, and Ser441 of TMPRSS2 were seen to have some of the largest average increases in wrapping. These general results were also compared to results from docking of the known TMPRSS2 inhibitors, camostat, and nafamostat.

The data generated from docking were then used to examine potential applications of the wrapping molecular descriptor using machine learning techniques: (i) for prediction of the solvent-accessible surface area term ΔG_{sa} of the GBSA score using regression and (ii) for classifying the solvent interactions of a TMPRSS2-inhibitor complex as favorable or unfavorable. The descriptor was seen to be only weakly related to ΔG_{sa}; the best-performing regression model had a Pearson

correlation coefficient of 0.76 between the predictions and the actual values. The ability of the descriptor to classify solvent interactions was more satisfactory, with a highest value for area under the receiver operating characteristic curve of 0.75.

The descriptor was then used to analyze the effect of inhibitor binding on the dynamics of TMPRSS2 BHBs. For this, molecular dynamics simulation was carried out for the uncomplexed TMPRSS2, as well as its complex with known inhibitors and hit molecules from docking. The binding of these ligands was seen to improve the stability of TMPRSS2; certain BHBs which were unstable or not formed in the uncomplexed case, showed increased stability. These prominently included a couple of BHBs identified from docking as having gained a large increase in wrapping. The improved stability coincided with an increase in wrapping groups in several cases. The descriptor also successfully rationalized the desolvation of a few BHBs due to inhibitor binding.

This work demonstrates the potential application of the concept of hydrogen bond wrapping in understanding the mechanism of inhibitor binding and the resultant desolvation effects on a protein's BHBs, without computationally expensive calculations. While the analysis methods require further improvement, the wrapping descriptor shows promising results and could be developed into a simple, yet powerful tool for drug discovery.