Discovering Druggable Binding Sites on Intrinsically Disordered Proteins

Amol Punjabi
Massachusetts Academy of Math and Science
Worcester, MA

Under the direction of
Dr. Gil Alterovitz
Informatics Program
Boston Children’s Hospital

Abstract
Intrinsically disordered regions of proteins take part in protein-protein interactions and have been directly implicated in many life-threatening diseases, including cancer, cardiovascular disease, and Alzheimer’s disease. However, since their structural flexibility makes them difficult to study, there is no consistent approach to designing drugs that bind to them. In this study, we hypothesize that there exist stable, ordered binding sites immediately adjacent to intrinsically disordered regions where a drug molecule could bind and, by steric hindrance, inhibit a disease-causing protein-protein interaction. To test this hypothesis, a dataset of near-complete protein structures containing intrinsically disordered regions was assembled by parsing the Uniprot and Protein Data Bank databases. Then, a parallelized pipeline using the SiteMap algorithm was developed to search for potential drug binding sites on these proteins. Binding sites immediately adjacent to the intrinsically disordered regions were extracted via proximity filter. In order to evaluate their druggability, a supervised Random forest machine learning classifier was trained on known drug binding sites. This was then applied to the sites found next to the intrinsically disordered regions to evaluate their viability as drug binding sites. Out of the 1663 binding sites found near the intrinsically disordered regions, fourteen were viable as drug targets. Five of these sites are on proteins which are implicated in cancer and heart disease, the leading causes of death worldwide.

Summary
Drugs work by binding to a protein target in your cells, thereby blocking them from binding to something else. If a protein target is ordered and relatively rigid, designing a drug to “fit” onto it is not so difficult. However, recently, scientists have discovered a class of proteins called intrinsically disordered proteins. Such proteins have regions on them that are not rigid, and unfortunately, these disordered regions have been shown to drive many life-threatening diseases, including cancer, heart disease, and Alzheimer’s. Few researchers have been able to design drug molecules that bind to these intrinsically disordered proteins because the disordered regions do not have a fixed shape. In this study, we are trying to develop a method that can be used to discover drug molecules in the future that bind to intrinsically disordered proteins and block them. We hypothesize that right next to the disordered regions on these proteins’ surfaces, there are rigid sites where a drug could bind and maybe take up enough space to interfere with the disordered region’s disease-causing activity. To test this hypothesis, we assembled a database of intrinsically disordered protein structures and ran software to find pockets on the surfaces of these protein structures. Then, we used machine learning to predict which sites would be the easiest to target with drugs. Five novel sites were found on proteins involved in cancer and heart disease which can be targeted with drugs in the future.