Jacob Paul
12/13/02
Bioinformatics
TA Haiyuan Yu
Prof. Mark Gerstein

Targeting Antigenic Variation with Molecular Docking

The technique of molecular docking has emerged as an important method in elucidating the interactions between a target protein and a potential ligand. Early molecular docking algorithms supposed a rigid protein and a rigid ligand in order to render the docking problem tractable. Then flexible ligands became an option in molecular docking methods. Recent techniques such as FlexE even allow for some flexibility of target proteins. A promising avenue apparently absent in the current literature for this flexible protein-flexible ligand docking is modeling antigenic variation and screening for inhibitors and possible therapeutics to counter antigenically variable protein ensembles.

Taylor *et al* explain that a molecular docking method essentially seeks to model the conformational interactions between a small molecule ligand and a target protein. To this end, the docking algorithm has a search strategy and a scoring function. The search focuses the computational power upon a particular site on the target protein, since simply reading the "open surface" of the protein would render the docking problem intractable. Therefore, knowing the active site or at least the domains that would interact with potential small molecules helps. The scoring function assesses the interaction between the ligand and protein, usually by calculating electrostatic potentials and hydrogen bonding between interacting atoms, for instance. Maintaining a rigid ligand and protein reduced the amount of computational resources necessary, but it obviously presented a grossly unrealistic picture of the interaction. Docking with flexible ligands helped, but only until both the ligand and protein could be flexible in their structures could molecular docking be of real use to meeting the problem of antigenic variation.

Antigenic variation is a challenge from a therapeutic and pharmaceutical perspective. Parts of the surface proteins of influenza, neuraminidase (NA) and hemagluttinin, are highly variable and are thereby able to avoid immune detection. Only with periodic vaccination against the current antigenic forms can the body generate effective antibodies. Instead of having to devise elaborate means of generating large amounts of virus every year (Hoffman *et al* 2002), it would be much more advantageous if one could perform a virtual screen of a library of potential small molecule inhibitors to find an antagonist of multiple versions of NA and HA (see Freymann *et al* 2000).

Claussen *et al*'s work allows for a family of highly similar proteins that yet have important, functional differences to be considered the target as an ensemble in a molecular docking program. FlexE superimposes the family of structures in a way that avoids the blurring that a simple averaging of coordinates would manifest (Claussen *et al* 2001; Osterberg *et al* 2002). Since the locus of NA's interaction with the immune response is known (Lee and Air 2002), one can limit the area of focus on the protein, adding to the computational efficiency of the molecular docking algorithm. FlexE requires the inputs of known crystal structures with coordinates, so as many NA's as have ever been crystallized from various influenza strains could be entered into FlexE. Since many influenza NA variants may not have high resolution crystal structures, but may have determined protein sequences, homology modeling with the known NA structures could be used to predict the structures of the variant protein sequences. FlexE is prepared to include the perhaps ambiguous structures that are the result of homology modeling in allowing for flexibility in the receptor ensemble (Claussen *et al* 2001). In fact, the high degree of similarity characteristic (yet with functionally significant differences) among antigenically variable proteins, such as the surface proteins of the influenza virus, makes homology modeling and flexible protein superposition all the more apt. Furthermore, FlexE deals much more efficiently with the ensemble than does cross docking, or determining ligand binding one by one for each protein (Claussen *et al* 2001).

Other workers have made efforts to including the important solvation effects so often ignored in molecular docking programs (Shoichet *et al* 1999 for instance). Osterberg *et al*'s docking method allows for protein motion upon ligand binding and for the reconfiguration of water molecules integral to the protein structure (2002). It involves building a combined, representative interaction grid from all of the protein structural conformations available. Values from different grids are given weights that correspond to the interaction energies, such that the weighted averages that comprise the ensemble structure are not simplistic means but instead reflect the physical properties of the protein conformations. Osterberg *et al* draw from a set of possible inhibitors, but a virtual screening approach would scan through an entire comprehensive library of possible small molecule inhibitors, such as the Available Chemicals Directory, to find a broadly inhibitive molecule against multiple strains of antigenically variant proteins.

Osterberg *et al* test their model on an ensemble of variable HIV proteins. This kind of flexible docking and screening system could be used to identify antagonists to malarial switching as well, for instance. Structure based design of therapeutics could begin instead of having to rely upon incessant epidemiological measures to counter a disease. One drawback of the FlexE model, is however that "in order to evaluate FlexE we need to know the correct binding mode" (Claussen *et al* 2001). It is quite often the case that the particular binding site and mode is known, but this limitation means that experimental chemistry will need to be conjoined with virtual screening.

Also, while Osterberg *et al* take steps to correct the mistake of simple averaging, I find Claussen's criticism of Osterberg *et al*'s methodological parent, Knetgel *et al*, convincing and applicable to Osterberg in a limited sense, although Osterberg *et al*'s model is a fine one. Claussen's main criticism is that the sort of averaging performed by Knetgel *et al*, or in a restricted sense, Osterberg *et al*, neglects the "combinatorial nature of the problem resulting from explicitly distinguished alternative conformations" (Claussen *et al* 2001). Claussen *et al* average similar regions but keep dissimilar parts

of the ensemble proteins as discrete alternatives. This, I believe, is the convincing strength of the FlexE approach, although Osterberg *et al* do account for the structural contribution of water molecules. In the end, using both approaches to determine a set of candidate inhibitor molecules followed by *in vitro* and *in vivo* confirmation of the set should prove a powerful method in devising therapeutics against antigenically variable proteins implicated in disease.

Works Cited

Claussen H, Buning C, Rarey M, Lengauer T. 2001. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol. 308:377-395.

Freymann, Douglas M, Wenck, Mary Anne, Engel, Juan C, Feng, Jun, Focia, Pamela J, Eakin, Ann E and Sydney P Craig III. 2000. Efficient identification of inhibitors targeting the closed active site conformation of the HPRT from Trypanosoma cruzi. Chemistry & Biology. 7:957-968.

Hoffmann E, Krauss S, Perez D, Webby R, Webster R. 2002. Eight-plasmid system for rapid generation of influenza virus vaccines. Vaccine. 20: 3165–3170

Knegtel RM, Kuntz ID, Oshiro CM. 1997. Molecular docking to ensembles of protein structures. J Mol Biol. 266:424-440.

Lee JT, Air GM. 2002. Contacts between Influenza Virus N9 Neuraminidase and Monoclonal Antibody NC10. Virology. 300: 255–268.

Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. 2002. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins. 46: 34-40.

Shoichet, Brian K, McGovern, Susan L, Wei, Binqing and John J Irwin. 2002. Lead discovery using molecular docking. Current Opinion in Chemical Biology. 6: 439–446.

Shoichet, Brian K., Leach, Andrew R., and Irwin D. Kuntz. 1999. Ligand Solvation in Molecular Docking. Proteins: Structure, Function, and Genetics. 34: 4–16.

Taylor, R. D., Jewsbury, P. J. and J. W. Essex. 2002. A review of protein-small molecule docking methods. Journal of Computer-Aided Molecular Design. 16: 151–166.