

Algorithmic design of ligand binding pockets on protein surfaces

S. Eyrisch¹

V. Helms¹

¹ Center for Bioinformatics, Saarland University, Saarbruecken, Germany

In the last few years, the modulation of protein-protein interactions and, in particular, the discovery of so-called small molecule protein-protein interaction inhibitors (SMPPPIs) has become a very active field of research. So far, SMPPPIs have been identified for several protein complexes [1]. However, structure-based drug design of such inhibitors is still in its infancy. In contrast to the well-defined binding pockets in enzymes, most unbound structures of proteins involved in protein-protein interactions lack deep clefts or clearly shaped binding pockets. We therefore developed a pocket detection protocol that provides a starting point for *in-silico* drug design for such cases. This method was validated on three protein-protein interaction systems for which small molecule inhibitors are known, namely MDM2:p53, BCL-X_L:Bak, and IL-2:IL-2R α . We found that large pockets not detectable in the unbound structure opened frequently on the protein surface during a 10 ns molecular dynamics (MD) simulation in explicit water at room temperature. These transient pockets represent potential binding sites of new inhibitors. At the native binding site, pockets of similar size as with a known inhibitor bound could be observed for all three systems, although these pockets were not - or only partly - present in the starting structure. Docking known inhibitors into these transient pockets resulted in docking poses with less than 2 Å RMS deviation from the crystal structures [2]. Unfortunately, the underlying MD simulations make this protocol quite time-consuming. However, if the potential binding site of a SMPPPI (e. g. the protein-protein interaction interface) is known, conformational sampling focused on this region appears more promising than scanning the whole protein surface for transient pockets. Therefore, we present here an efficient method for generating putative binding pockets on protein surfaces algorithmically. After defining the starting structure(s), the approximate location and volume of the desired binding pockets, and a radius dictating which surrounding residues are treated as flexible, the algorithm calculates a pre-defined number of energetically favorable conformations containing these pockets. Internally, the algorithm represents the pockets by dummy atoms and scores their volume via van der Waals energies with the flexible part of the protein. The actual generation of conformations consists of two stages: In the first stage, all flexible residues are mutated to glycine and all rotameric states of their real side chains are pre-calculated. In the second stage, a tree is built up in which each node represents a (partial) conformation of the flexible part of the protein. Each node is scored according to the energy of this conformation and the pocket contribution. The top-scoring leaf nodes then represent the energetically most favorable conformations containing putative ligand binding pockets. For the three proteins mentioned before, the algorithm could generate promising low-energy pockets with realistic predetermined volumes within a few CPU minutes on a standard desktop PC.

1. Wells, J. A.; McClendon, C. L. Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces. *Nature* **2007**, 450, 1001-1009.
2. Eyrisch, S.; Helms, V. Transient Pockets on Protein Surfaces Involved in Protein-Protein Interaction. *J. Med. Chem.* **2007**, 50, 3457-3464.