Optimization of post-docking strategies for hit identification of HCV NS5B polymerase inhibitors

Tanaporn Uengwetwanit and Wolfgang Sippl

Department of Pharmaceutical Chemistry, Martin-Luther University Halle-Wittenberg, 06120, Halle (Saale), Germany

NS5B is an RNA-dependent RNA polymerase (RdRp), a key enzyme in HCV replication process, and a well validated drug target[1]. In an effort to establish an efficient for computational screening of HCV polymerase inhibitors, we tried a systematic combination of docking and post-docking strategies. Glide standard precision (SP) docking which allow flexible hydroxyl groups was applied to a set of known inhibitors [2]. We present an evaluation of 3 post-docking strategies including random forest (RF) classification, structural interaction fingerprint (SIFT)[3], and incorporation of docking to dummy binding sites. Random forest, an ensemble leaning method, was trained by 397 known inhibitors and used to build two models. RF model-1 predicted the compounds to bind or unbind and model-2 classified the compounds into potent or weakly actives. Structural interaction fingerprint was used to compare the interaction similarity of a given compound to the known inhibitors. But instead of one to one comparison between two molecules, we derived conserved interaction patterns from 29 crystal structures and used those as references. The last strategy called "two sites docking", compared the docking to a target site with docking to a dummy binding site. Both binding sites show different binding site structures. The compounds which scored well in both binding sites were discarded. This strategy was based on an idea that a good candidate compound should specifically bind only to one target binding site. All procedures were validated by enrichment studies of a collection of 99 known HCV polymerase inhibitors and 1693 decoys. The results show that combining Glide SP with RF models provides a substantially better discrimination than the other methods in case of the validation subset-1. RF model-1 could obtain 18 known inhibitors among the 20 top ranked compounds (about 1% top ranking). However, the RF model has the limitation that it predicts well what has used for training. In the validation subset-2, only 5 known inhibitors were found among the 1% top ranked hits. Thus without prior knowledge of inhibitors, two sites docking should be considered as most suitable strategy. Within all validation sets two sites docking accurately predicts 16 known inhibitors among the 20 top ranked compounds.

[1] C. M. Lange, C. Sarrazin, S. Zeuzem, *Aliment Pharmacol Ther* **2010**, *32*, 14-28.

[2] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J Med Chem* **2004**, *47*, 1739-1749.

[3] Z. Deng, C. Chuaqui, J. Singh, *J Med Chem* **2004**, *47*, 337-344.