Using large numbers of protein-ligand complex structures to develop targeted scoring functions

Marcel L. Verdonk, Suzanne C. Brewerton, Richard J. Hall, Wijnand, T.M. Mooij, Christopher W. Murray

Astex Therapeutics Ltd., 436 Cambridge Science Park, Milton Road, Cambridge, CB4 0QA, UK

BACKGROUND

Protein-ligand X-ray structures

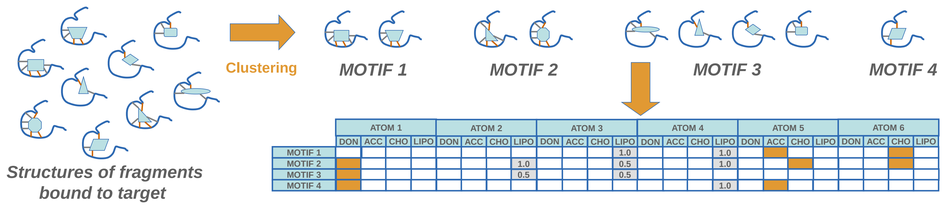
Over 15,000 structures of protein-ligand complexes are available in the Protein Data Bank. In addition, the Astex in-house database contains over 2,500 structures of protein-ligand complexes for >20 drug targets. In theory these databases form a rich source of data to aid structure-based drug design programs. However, for this data to be useful it needs to be curated and organised. Related structures need to be grouped together, aligned and superimposed, ligands need to be atom typed and stored in a searchable form, etc. The curated data also needs to be made available to the project teams in a form that allows them to quickly distil out the information that is relevant for the project. And finally, as often there is a large amount of structural data to consider, methodologies for automating the use of experimental structural data on structure-based design projects are required.

Scoring functions

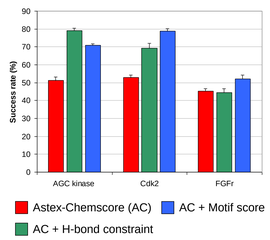
A multitude of scoring functions have been reported in the literature that all aim to predict the binding mode of compounds bound to a protein active site and/or attempt to predict the relative binding affinities of different compounds. The types of scoring functions range from simple knowledge-based and empirical scoring functions to sophisticated methods like free energy perturbation and QM/MM. State-of-the-art docking programs can successfully predict the binding modes of compounds in ~80% of the cases when the compounds are docked into their native conformation of the protein. However, when compounds are docked into non-native conformations of the protein, success rates are significantly reduced. In addition, predicting the relative activities of different compounds can be a real challenge, even with the most sophisticated scoring protocols. In our experience, incorporating structural data on relevant targets into scoring functions can improve both the docking performance and the ranking of compounds.

ASTEX INTERACTION MOTIFS

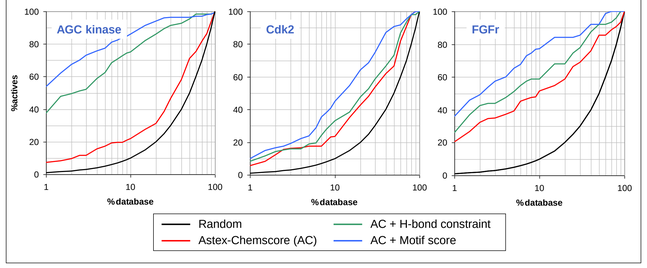
Protein-ligand interactions often involve 'motifs' consisting of two or more H-bonds and/or CH…O bonds. Traditional fingerprint scores are unaware of motifs; for example, A H-bond that is always formed as part of a motif is still scored if formed in isolation. This is addressed by grouping complexes into interaction motifs, which are then scored as follows: (i) The motif score is only added if all interactions in motif are formed; (ii) Only one motif can to contribute to score; (iii) A traditional fingerprint score is used for lipophilic interactions. We have incorporated this methodology in GOLD [1], and hence it can be used to drive the docking and/or to score/rank compounds.



Binding mode predictions

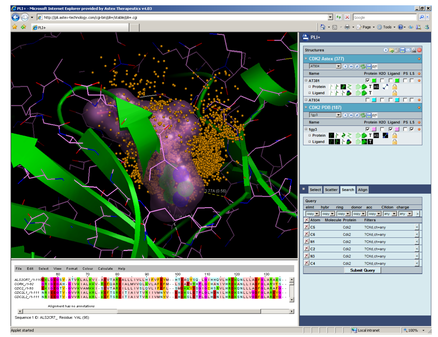


Database enrichments



PLI+

PLI+ is the latest version of Astex's Protein-Ligand Informatics platform. The ORACLE database behind PLI+ contains all Astex in-house structures and all PDB structures of targets of interest to Astex. Ligands have been atom-typed using proprietary software and are stored as SMILES strings. Proprietary software KFIT was used to pre-superimpose structures of related targets. PLI+ has a web-based user interface that uses AstexViewer [3] for 3D-structure visualisation and Jalview [4] for viewing sequence alignments.

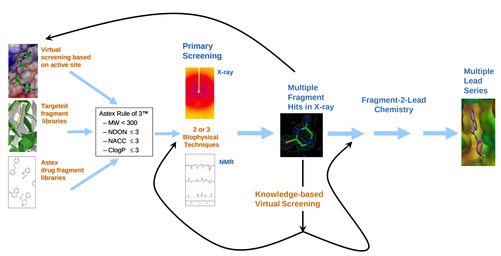


Functionality available in PLI+ includes: (i) Quick look-up of Astex and PDB structures for selected target(s); (ii) Searching for targets similar (globally or locally) to selected target(s); (iii) 2D (chemical substructure) ligand-based searching; (iv) 3D ligand-based searching; (v) Protein-ligand contact analysis; (vi) 3D ligand atom scatter plots; Etc. The structural and alignment data in PLI+ is directly linked to interfaces to construct pharmacophores, targeted scoring functions etc.

PYRAMID™

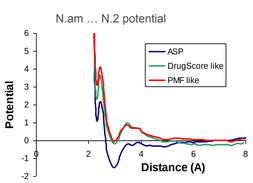
Astex Therapeutics uses a fragment-based screening approach (Pyramid™) which employs a range of biophysical and computational techniques, including X-ray crystallography, to discover new leads for drug discovery. The high-throughput X-ray crystallography capability at Astex allows low affinity, but ligand efficient small molecules to be identified, structurally characterised and transformed through the structure-based design process into compact and highly tractable drug leads.

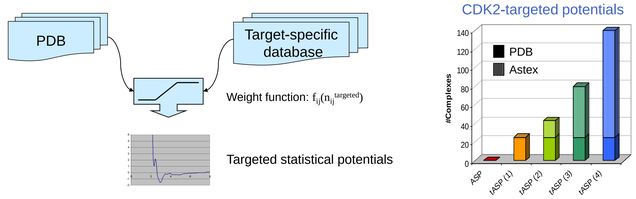
Astex has an in-house database of >2,500 protein-ligand complex X-ray structures. The unique fragment-based drug discovery approach means X-ray structures for a significant number of compounds (10s-100s) bound to the target are available at the very start of the project.



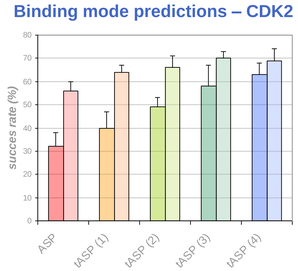
ASTEX STATISTICAL POTENTIAL (ASP)

The main difference between the Astex Statistical Potential (ASP) and other knowledge-based potentials like DrugScore and PMF is its reference state [2]. As a result, the ASP atom-atom potentials are often more physically meaningful, particularly for protein atom types that tend to be buried like backbone amines. Knowledge-based scoring functions are very suitable for generating targeted scoring functions, as the potentials are simply derived from target-specific data, i.e. complexes of a particular target or target class. We have developed methodologies for generating targeted forms of ASP [2], which can then be used in GOLD to either drive the docking or to score/rank compounds.

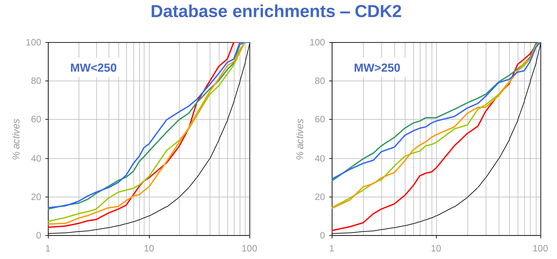




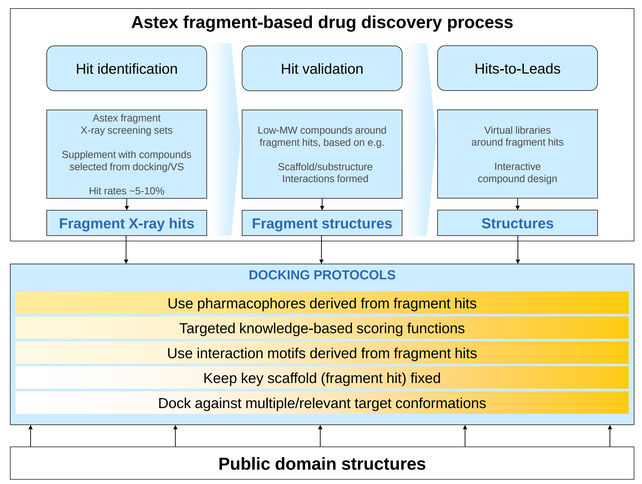
Binding mode predictions – CDK2



Database enrichments – CDK2



USING STRUCTURAL DATA FOR DOCKING/SCORING



1. Jones, G., Willet, P. Glen, R.C., Leach, A.R., Taylor, R., J.Mol.Biol. 267(3) 1997, 727-748.
2. Mooij, W.T.M, Verdonk, M.L., Proteins 61(2) 2005, 272-287
3. Hartshorn, M.J., J.Comput. Aided Mol. Des. 16(12) 2002, 871-881.
4. Clamp, M., Cuff, J., Searle, S.M., Barton, G.J., Bioinformatics 20 (2004), 426-427.

astex™ therapeutics
www.astex-therapeutics.com