Electronic Supplementary Material (ESI) for Medicinal Chemistry Communications


**Supporting Information**

**Title:** Targeting the inactive pocket of protein kinases: computational screening based on ligand conformation

**Authors:** Pascal Bonnet[*,1], Daniel Mucs[2] and Richard A. Bryce[2]

## Methods

*Datasets:* To validate the virtual screening approach, we have used two large diverse datasets. The first one contains a dataset of unique 683 Type I and 59 Type II kinase inhibitors extracted from the MOE Kinase database provided in MOE 2010.10 software. Amongst a total of 2328 protein kinase crystal structures available, the original MOE Kinase database contains 1180 and 135 protein kinases with DFG-in and DFG-out conformations respectively. A small number of kinase inhibitor scaffolds were not used in this study, such as allosteric binders or when the DFG motif was not determined by crystallography. Type I and Type II kinase inhibitors are classified based on the conformation of the phenylalanine or tyrosine from the DFG motif[l] located at the N-terminus of the activation segment.[2,3] Importantly, we have removed the four Type II kinase inhibitors (doramapimod from 1KV2, imatinib from 1IEP, sorafenib from 1UWH, nilotinib from 3CS9) used to build initially the pharmacophore query. The second dataset used in this study contains a database of the Directory of Useful Decoys[4] modified by Hinselmann and co-workers.[5] This subset which contains over 104000 diverse compounds is tailored for assessment of virtual screening.

*Shape-based screening* of these ligand sets was performed using the ROCS method[6] as implemented in OpenEye.[7] The ROCS method determines the maximum volume overlap of two molecules, modelled via Gaussians, using rigid-body superposition. In this study we used

the ComboScore scoring function to rank hits which combined the ShapeTanimoto and ColorScore scores with equal weight.[8]

*Pharmacophore generation* was achieved using MOE.[9] The optimal query Q7 identified previously[10] was used in this study. Q7 was generated via four reference ligands, binding to three different kinases: 1IEP, 1KV2, 1UWH and 3CS9. The four protein/ligand complexes were superimposed and pharmacophoric features added from common annotation points such as hydrogen bond donor and acceptor, hydrophobic, aromatic and excluded volume. Additionally, constraints were included such as the presence of essential pharmacophoric feature in the query.

*Macromodel* (version v95207) was used applying the Monte Carlo Multiple Minima method with a usage-directed search on the molecular structures. The default method is random walk; we find that the usage-directed search gives improved search performance. The structure selection allows the program to select starting geometries for Monte Carlo search steps in different ways. The MMFF94s force field was used, the RMSD tolerance was set to 0.25 Å, with an energy window to 20 kcal/mol and maximum number of conformers per molecule of 1000. In all cases, structures must be within 100 kcal/mol of the current global minimum to be candidates for starting geometry selection. It is a usage-directed because the least used structures will be used as starting geometries if their energies are allowed by a specific argument. All screening processes were implemented into Pipeline Pilot[11] to easily automate and process the analysis of the data.

*Analysis of screening.* The receiver operating characteristic[12,13] (*roc*) curve was used; a *roc* curve is a plot of the true positive rate versus the false positive rate. Such plots can be used to study the ability of screening protocols to distinguish active from inactive compounds. To further assess the performance of the screening, we also define the predicted positive rate

(PR), also known as the precision, as the number of known Type II inhibitors retrieved in the top *x*% of the ranked database, divided by the total number of predicted Type II structures. This quantity also corresponds to the enrichment factor (EF) of the model divided by EF of the perfect model at *x*% of the sorted database.

Ranking of retrieved molecules from pharmacophore search is performed based on the root-mean-square deviation in distance between the query features and their matching ligand annotation points implemented in MOE. This is not an established scoring function, and this can explain the inferior *roc* score obtained for pharmacophore screening compared to ROCS consensus screening which uses ComboScore as its scoring function.

*Similarity Searching*. A 2D-similarity search of the best DUD hits was performed on the X-ray Type II molecule database and on the Prous external database[14] using FCFP6 (functional class fingerprint description 6) molecular descriptors as implemented in PipelinePilot.[11] The Tanimoto index was used to calculate the similarity between two molecules with 0 being maximally dissimilar, 1 being maximally similar. Once again, all screening processes were implemented into Pipeline Pilot for ease of data analysis.

## References

[1] Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2*, 358–364.

[2] Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. Protein Kinase Inhibitors: Insights into Drug Design from Structure *Science*, **2004**, *303*, 1800-1805.

[3] Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science*, **2002**, *298*, 1912-1934.

[4] Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem*. **2006**, *49*, 6789-6801.

[5] Jahn, A.; Hinselmann, G.; Fechner, N.; Zell, A. Optimal Assignment Methods for Ligand-Based Virtual Screening. *J. Cheminf*. **2009**, *1*, 14-37.

[6] Rush, T. S., 3rd; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem*. **2005**, *48*, 1489–1495.

[7] Omega; OpenEye Scientific Software, http://www.eyesopen.com, 1997.

[8] Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How To Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information. *J. Chem. Inf. Model.* **2009**, *49*, 678-692.

[9] MOE 2009.10 and 2010.10; Chemical Computing Group Inc.: Montreal, Canada, http://www.chemcomp.com, 2010.

[10] Mucs, D.; Bryce, R. A.; Bonnet, P. Application of shape-based and pharmacophore-based in silico screens for identification of Type II protein kinase inhibitors. *J. Comput. Aided Mol. Des.* **2011**, *25*, 569-581.

[11] Pipeline Pilot version 8; Accelrys, Inc., San Diego, USA, http://www.accelrys.com, 2010.

[12] Gribskov, M.; Robinson, N. L. Use of receiver operative characteristic (ROC) analysis to evaluate sequence maching. *Comput. Chem*. **1996**, *20*, 25-33.

[13] Kirchmair, J.; Ristic, S.; Eder, K.; Markt, P.; Wolber, G.; Laggner, C.; Langer, T. Fast and efficient in silico 3D screening: Toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *J. Chem. Inf. Model*. **2007**, *47*, 2182–2196.

[14] Thomson Reuters Integrity database, http://thomsonreuters.com.