FLAP: Fingerprints for Ligands And Proteins. Latest Improvements and Applications.

[§]Simone Sciabola, [§]Gabriele Cruciani, [†]Loretta Aureli, [†]Ilaria Cerbara, [†]Alessandra Topai and [‡]Massimo Baroni

[§]Laboratory of Chemometrics, Department of Chemistry, **University of Perugia**, Via Elce di Sotto, 10 I-60123 Perugia, Italy; [†]**C4T** S.c.a.r.l., Via della Ricerca Scientifica, 00133 Rome, Italy; [‡]**Molecular Discovery** Ltd., 215 Marsh Road, HA5 5NE, Pinner, Middlesex, UK.

Introduction

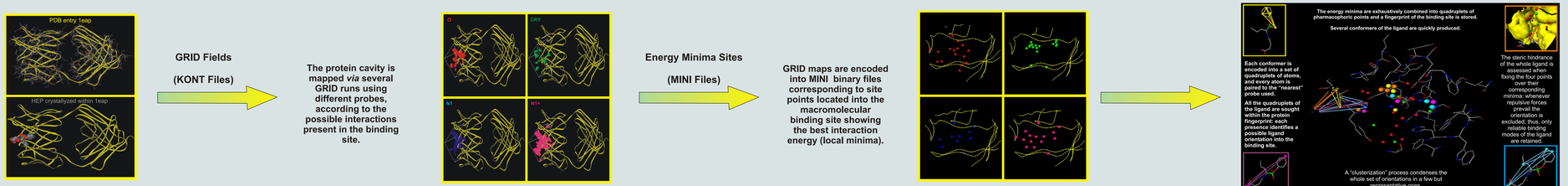
FLAP (Fingerprints for Ligands And Proteins) is a piece of software developed at University of Perugia in a collaboration between Pfizer and Molecular Discovery, able to describe small molecules and protein structures in terms of 3- or 4-point pharmacophore fingerprints using all the capability of the program GRID.

Molecular Interaction Fields (MIF) calculated by GRID [1], representing the interactions between probes and small molecules or defined regions of protein structures, contain relevant information on which kind of critical interactions a ligand may have with a receptor, or, in the case of proteins, which possible sites of interaction are present in a selected area of the macromolecular structure.

The information given by MIF is used by FLAP to identify key site points describing energetically favorable interactions between a given probe and a Target. Site points correspond to pharmacophoric features in the small molecule, such as the presence of a hydrophobic group, a HB donor and/or acceptor group. Site points so calculated can be used within FLAP to build all the possible 3- or 4-point pharmacophores, storing this information in a fingerprint.

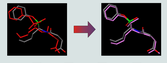
FLAP can be used as a docking tool, for Ligand Based Virtual Screening (LBVS), Structure Based Virtual Screening (SBVS), to investigate selectivity in proteins or receptors, to generate pharmacophore hypothesis of active compounds and for a fast generation of lattice independent molecular descriptors for 3D QSAR & QSPR studies.

Pharmacophore Perception using Molecular Interaction Fields



GLUE

All the reliable binding modes retained in the previous step are now the starting points for the intermolecular interaction energy evaluation. Each binding mode is optimized within the cavity by means of successive torsions and roto-translations driven by the GRID force field. Each little movement is followed by an energy reassessment according to the GRID standard equation applied over the whole ligand and active site.



$$E_{GRID}=E_{LJ}+E_{EL}+E_{HB}+E_{ENTROPY}$$

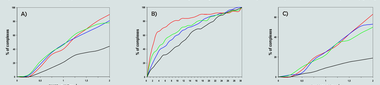
For the optimized orientations, the interaction energies between the ligand and the protein are calculated by using the GLUE equation, which provides an energy scoring function (E_{GLUE}).

$$E_{GLUE}=E_{SR}+E_{ES}+E_{RHB}+E_{DRY}$$

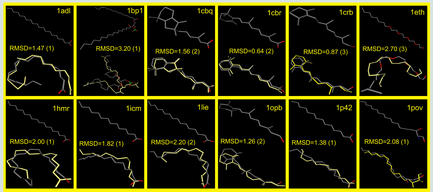
E_{SR} = steric repulsion term; E_{ES} = electrostatic term; E_{DRY} = hydrophobic term; E_{RHB} = hydrogen bonding charge reinforcement

Results

100 protein/ligand complexes from the RCSB [2] were used to generate a separate set of coordinates for the whole protein and its ligand. 3D geometry conformation of the ligands were obtained from smiles notation using CORINA [3]. Cumulative percentages of complexes as a function of the RMSD from the X-ray pose are reported. A) Docking accuracy: RMSD of the best pose (nearest to the experimental binding mode). B) Best pose ranking (RMSD cut-off → 2 Å). C) Scoring accuracy: RMSD of the top pose (best scored solution). GLIE results (red) were compared to those obtained by using GLIDE (green), GOLD (blue) and DOCK (black) [4,5].



The Glue docking procedure was also used to study 12 interesting targets for which other docking software usually seem to fail, such as fatty-acid binding proteins. For the whole set, the calculations started from random 3D ligand-coordinates: the best pose (rmsd<2.0A) was among the first three solutions proposed by Glue, and for 6 out of 12 the best pose was the first one.



Virtual Screening

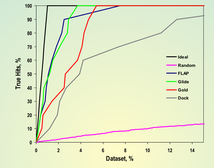
Virtual screening studies have been performed with FLAP in order to test both the Receptor and the Ligand Based approach.

In Ligand Based Virtual Screening, ligands are compared with each other using a similar method to that used for comparing ligands and a protein structure. Flap computes the ligand pharmacophores that are common between a ligand template and the other ligands under investigation.

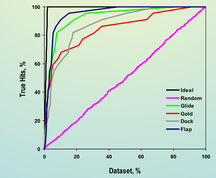
Receptor Based Virtual Screening

Thymidine Kinase (TK) is a challenging test case due to its flexible active site, and presence of several water molecules that participate in ligand binding.

Starting from the coordinates of the protein complexed with deoxythymidine (PDB code: 1KIM), a 1000 compound database containing 10 known actives was generated. The 990 other compounds were randomly extracted from the MDDR database with the constraint that they had to span the same Molecular Weight range shown by the 10 true active compounds. Virtual Screening performance was tested and compared using DOCK, GOLD and GLIDE docking methods [17].



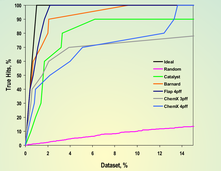
Factor Xa (Fxa) is challenging for multiple reasons. Its binding site is accessible to the solvent, several hydrophobic regions play a key role in the ligand binding event, and a salt bridges between region S1 in the protein and the P1 group in possible binders is thought to give high affinity.
PDB entry 1nfu was used as target and the decoy data set consisted of 22 known active compounds (12 in house + 10 public from RCSB PDB) and 1138 randomly selected compounds from MDDR.



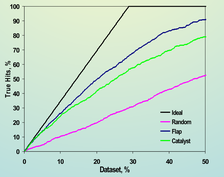
Ligand Based Virtual Screening

Estrogen Receptor-alpha ERα has already been successfully used as virtual screening target and represents a useful case study in order to compare performances of different Virtual Screening tools.

Starting from the coordinates of the ERα receptor complexed with 4-hydroxy-tamoxifen (PDB code: 3ERT), a 1000 compound database containing 10 known ERα antagonists was generated. The other 990 were randomly extracted from the MDDR database with the constraint that they had to span the same Molecular Weight range shown by the 10 true active compounds. Comparison with Catalyst [6], ChemX [7] and Barnard [8] fingerprints is reported.

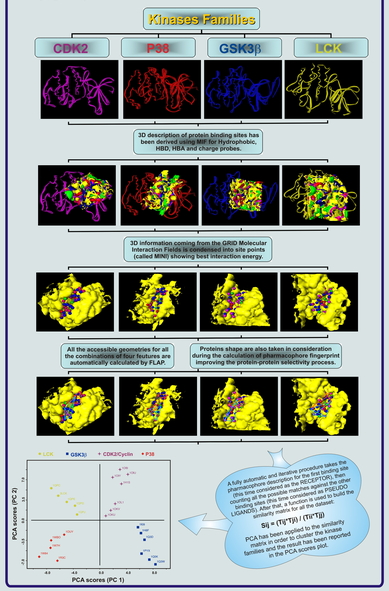


Ligand Based Virtual Screening has been performed with FLAP on a Pfizer in house project. Both active ligands and receptor structures were available.
7 different pharmacophores were built representing different series of ligands active against the same target. Virtual Screening has been carried out with the seven pharmacophores, one at a time, to rank a given library. The results from the seven runs have then been merged. The shape of the Target has been used as a constraint. A similar procedure has been applied in Catalyst for a comparison. See Enrichment plot below.



Protein Similarity

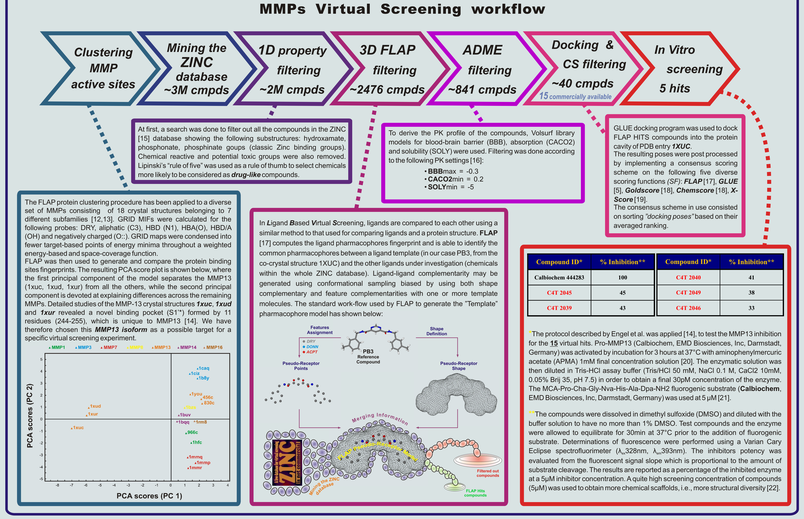
FLAP can also be used for protein similarity comparison using 3D information present in the binding site. The same work-flow described before has been applied in a preliminary study for a dataset made of 23 different protein structures belonging to 4 specific kinases sub-families.



Matrix MetalloProteinases: FLAP identification of potential MMP13 inhibitors

MMPs are a family of zinc endopeptidases which degrade proteins and many components of the extracellular matrix. The approximately 27 current known human MMPs can be grouped into subfamilies based on their substrate specificity [9]. MMPs up-regulation has been related to different pathological disorders, involving metastasis, angiogenesis, cardiovascular diseases, osteoarthritis, and rheumathoid arthritis. The development of potent and selective inhibitors with a wide variety of zinc binding motifs has been documented [10,11]. Nevertheless only few classes of binding modes showing no interaction between the inhibitor and the catalytic zinc have been so far investigated.

MMPs Virtual Screening workflow



Compound ID*	% Inhibition**	Compound ID*	% Inhibition**
Calbiochem 444283	100	C4T 2040	41
C4T 2045	45	C4T 2049	38
C4T 2039	43	C4T 2046	33

The protocol described by Engel et al. was applied [14], to test the MMP13 inhibition for the 15 virtual hits. Pro-MMP13 (Calbiochem, EMD Biosciences, Inc, Darmstadt, Germany) was activated by incubation for 3 hours at 37°C with aminophenylmercuric acetate (APMA) 1mM final concentration solution [20]. The enzymatic solution was then diluted in Tris-HCl assay buffer (Tris/HCl 50 mM, NaCl 0.1 M, CaCl2 10mM, 0.05% Brij 35, pH 7.5) in order to obtain a final 30pM concentration of the enzyme. The MCA-Pro-Cha-Gly-Nva-His-Ala-Dpa-NH2 fluorogenic substrate (**Calbiochem**, EMD Biosciences, Inc, Darmstadt, Germany) was used at 5 μM [21].

The compounds were dissolved in dimethyl sulfoxide (DMSO) and diluted with the buffer solution to have no more than 1% DMSO. Test compounds and the enzyme were allowed to equilibrate for 30min at 37°C prior to the addition of fluorogenic substrate. Determinations of fluorescence were performed using a Varian Cary Eclipse spectrofluorimeter (λ_{ex}328nm, λ_{em}393nm). The inhibitors potency was evaluated from the fluorescent signal slope which is proportional to the amount of substrate cleavage. The results are reported as a percentage of the inhibited enzyme at a 5μM inhibitor concentration. A quite high screening concentration of compounds (5μM) was used to obtain more chemical scaffolds, i.e., more structural diversity [22].

Conclusions

The FLAP (Fingerprint for Ligands and Proteins) program represents a promising approach to gain information from the Molecular Interaction Fields calculated by the GRID software within a region of a protein structure and from the atom classification in GRID probes for atoms in a ligand molecule.

FLAP is able to perform a comparison between protein and ligand pharmacophore fingerprints, between ligands pharmacophore fingerprints and between proteins pharmacophore fingerprints. This approach can be exploited very straightforwardly in Structure Based Drug Design and docking, Ligand Based Virtual Screening and protein similarity.

Flexibility and shape of the ligand and/or of the active site of the protein are taken into consideration. Constraints can be set by the user as well as other keywords able to describe particular features of the protein active site or within the ligand molecules. The calculation of the pharmacophore fingerprints is fast and a reasonably large number of molecules can be handled.

Finally, the possibility to apply Principal Component Analysis (PCA) and Partial Least Square (PLS) as tools for the statistical analysis of pharmacophore fingerprints represent an interesting and novel approach in the 3D QSAR field.

Finally, the MMP13 case study showed the potential of using a combination of Structure/Ligand Based Virtual Screening tools as implemented in the FLAP software, for mining commercially available databases. As a result, we were able to identify novel HITs, as promising MMP-13 inhibitors. Further *in vitro* selectivity studies still have to be performed, but due to their drug-likeness, they may already be used in the next step of Hit-to-Lead.

References

[1] Goodford, P.J., *J. Med. Chem. 1985, 28, 849-857*
[2] RCSB Protein Data Bank, http://www.rcsb.org/pdb/
[3] CORINA, http://www.mol-net.de/software/category/gen3dcorord.html
[4] Paul, N. and Rognan, D., *Proteins 2002, 47, 521-533*
[5] Sciabola et al., *12th Symp. QSAR & Molecular Modeling 2004, 47-49*
[6] Catalyst, version 4.9.1, *MSI, San Diego, CA, USA*
[7] Chem-X software, *Oxford Molecular, Medawar Centre, Oxford Science Park, UK*
[8] Barnard, *J.M., Chem. Inf. Comput. Sci. 37, 1997, 141-142*
[9] Borkakoti N., *J. Mol. Med. 2000, 78, 261-269*
[10] Whittaker, M. et al., *Chem. Rev. 1999, 99, 2735-2776*
[11] Fingleton, B., *Curr. Pharm. Design 2007, 13, 333-346*
[12] Pirard, B. et al., *J. Med. Chem. 2000, 49, 61-69*
[13] Terp, G.E. et al., *J. Med. Chem. 2002, 45, 2676-2684*
[14] Engel, J. et al., *Chem. Biol. Interact. 2005, 151, 151-160*
[15] Irwin, J.J. et al., *J. Chem. Inf. Model. 2005, 45, 177-82*
[16] Carosati, E. et al., *J. Med Chem. 2007, 50, 2117-2126*
[17] Baroni, M. et al., *J. Chem. Inf. Model 2007, 47, 279-294*
[18] Jones, G. et al., *J. Mol. Bio. 1995, 245, 43-53*
[19] Wang, R. et al., *J.Comput.Aid.Mol.Des. 2002, 16, 11-26*
[20] Knight C. G. et al., *FEBS Lett. 1992, 296, 263266*
[21] Knauper, V. et al., *J.Biol.Chem. 1996, 271, 1544-1550*
[22] Walters, W. P. et al., *Nat. Rev. Drug Discov. 2003, 2, 259-266*