A fragment-based computational protocol at PDB scale – Application to lead-optimization of DFG-out kinase inhibitors

MORIAUD F.[1], HENRY T. [1], ADCOCK S.A. [1], VOROTYNTSEV A.M. [1], MARTIN L. [1], DOPPELT O. [1,2], de BREVERN A.G. [2], DELFAUD F. [1]
[1] MEDIT SA, 2 rue du belvédère, 91120 Palaiseau France, [2] INSERM U726 EBGM, Université Paris 7 case 7113, 2 place Jussieu, 75251 Paris Cedex 05 France.

Abstract

Fragment-based drug design has emerged in the last decade. This exciting field has been recently reviewed [1]. Obtaining structural information on the fragment complexed to the protein target is a key factor and also a major limitation. Therefore, computational methods are needed to mine efficiently all the available 3D structures of ligands complexed to proteins, both treated as a whole or best as smaller fragments to increase the likelihood of fragment hopping from one target to another.

In this work, we've used MED-SuMo [2] to query and mine the Protein's Surface Chemical Functions surrounding fragments of PDB ligands, seeking similarities with the kinase of interest (i.e. Vegfr DFGout, pdb code 2oh4, ligand code GIG) and collecting a library of 1129 unique* fragments positioned in the vegfr's active site and annotated with the counts of contacts and h-bonds.

Fragments can be used to design novel ligand scaffolds (lead generation) or to optimize attachments on a fixed scaffold (lead optimization). Here we present the optimization of a substructure (i.e. phenylamide) of the GIG ligand to find others DFGout ligands.

The 3D hybridisation in 5 iterations of the phenylamide moiety with 1129 fragments suggested by our MED-Hybridiser protocol leads to 22824 molecules. In this list, we identified 3585 different scaffolds, 298 are in PubChem, 46 in the PDB attesting of the diversity and quality of those generated molecules. 25 are marked as active on protein kinase in PubChem bioassay.

MED-SuMo technology for Drug Design at PDB scale

Overview: MED-SuMo, a very powerful target-based drug design software, makes possible to compare any interaction surface against the full Protein Data Bank in few minutes. Independent from the notion of protein sequence or fold, MED-SuMo detects and compares biochemical functions on protein surfaces such as H-Bonds, Charges, Hydrophobic and Aromatic groups. In order to rapidly browse hundred of hits, the MED-SuMo scoring function takes into account biochemical functions and shape overlaps.

The core MED-SuMo algorithm is based on the representation of macromolecular structures using a set of Surface Chemical Functional groups (SCF). The functional groups are grouped into triplets, which are considered as the minimal unit for a biological function. The triplets form a graph which can be treated powerfully and quickly thanks to the graph theory. The result of comparisons consists of several matching sites. Each site is a set of pairs of matching functional groups (MED-SuMo Surface Chemical Features).

Advantages:

- The Whole PDB is browsed: hundreds of protein kinases are hit in this work.
- Others ATP binding proteins are hit, e.g. a synthetase/amidase (*Fig.1*) and are therefore amenable to drug design within MED-SuMo.
- One or more 3D alignments of a hit protein to the query are taken into account for a precise local alignment (*Fig.2*)

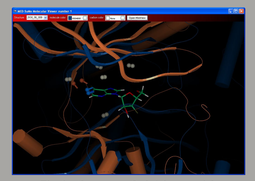


Fig.1 : the query 2oh4 (orange) is overlaid with an interfamily hit, 2oh8, found in the MED-SuMo results. 2oh8 is a bifunctional glutathionyl spermidine synthetase/amidase).

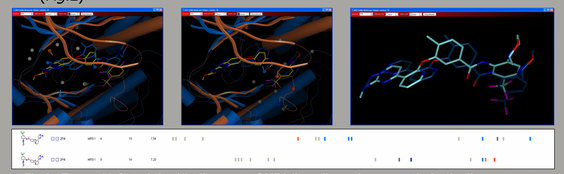


Fig.2 : Top and left to right: (1) alignment of MR9 ligand's environement in 2p4i pdb structure to 2oh4 (orange) in the hinge region in light blue (2) same but in the allosteric pocket region and shown as dark blue (4) same 2 ligands shown alone. Bottom: results table with only those 2 hits: for the same PDB file, two completely different SCFs signature are obtained, alignment in 3D are significantly different, e.g. non 3D duplicates. * defined by a rmsd > 0.25Å.*

Introduction to lead optimization with MED-SuMo "Case study of DFG-out VEGFR-2 kinase inhibitors"

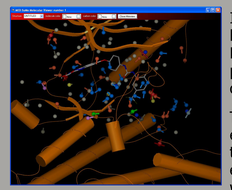


Fig.3 : 2oh4 is shown in orange. The co-crystallized ligand GIG is shown in grey. Others colored features are the MED-SuMo SCFs which describe the ligand environment at a maximum distance of 9Å

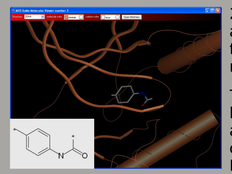


Fig.4 : 2oh4 is shown in orange. The phenylamide substructure of GIG, phenylamide is shown in grey in 3D. The 2D depiction of the phenylamide shows the attachment points for hybridisation with a star.

In this work, we've applied the protocol to generate kinase's DFG-out ligand. We've selected a VEGFR-2 DFG-out structure (2oh4) and we've kept the phenylamide moiety of its GIG ligand as a scaffold to optimize with attachments.

The database of PDB ligands, is queried with the 9 Å environment of GIG, shown in *Fig.3*. The database is the Med-SuMo PDB site database where the environment is defined as the environment at 4.5 Å where is potentially interacting with the ligand. Triplets of SCFs are set to the highest quality available in MED-SuMo (maximum edge = 20 Å, maximum sum of edges = 60 Å)

The scaffold phenylamide is shown in *Fig.4*, both in 2D and 3D. The atomic 3D coordinates are the same as in 2oh4. We've kept the only 2 possible positions for attachment (as found in GIG): para of the phenyl ring and to the C of the amide group.

The results of the query towards the whole PDB are hundreds of hits aligned to the query. Hits are analyzed, sorted and clean of false positives by a clustering according to the signature (Figure below) Duplicate Ligands in 2D are kept only if they differ by a rmsd > 0.25Å.

Hybridisation of such a large number of aligned ligand from the PDB had been forecasted recently [3] and we did an attempt with MED-SuMo in this work.

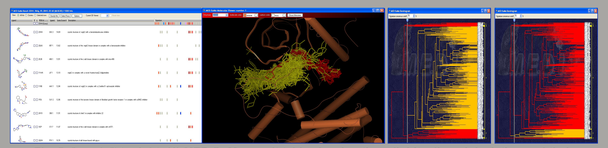


Fig. 5: from left to right: (1) MED-SuMo result table showing the pdb hits: ligand depiction and their environment with a signature. (2) ligands associated to pdb hits are shown in 3D in the reference frame of 2oh4 which is shown as a cartoon view in orange. (3) and (4) dendrogram which represents the classification of the ligands according to their environment, i.e. their signature of Surface Chemical Features. Two clusters contain only protein kinases with a correct fold alignment: cluster 4 is the largest one and contains 309 ligands which are aligned in the hinge region, and cluster 5 contains 25 ligands

Fragment-based application on VEGFR-2 with *MED-Hybridiser* protocol

The hybridisation of the phenylamide scaffold had been done with the MED-Hybridiser protocol, i.e. a Med-SuMo protocol with a strong focus on the portions of ligands (MED-portions): two ligands can have a rather different environment as a whole but share common portions with a highly conserved protein's environment (like the hinge neighborhood in *Fig.1*.).

We've used a phenylamide moiety as a starting scaffold to bias the the lead optimisation towards DFG-out ligands. It is an interesting case study as inhibitors targeting the DFG-out conformation may exhibit a higher selectivity and less competition with ATP, as exemplified by the drug imatinib. We've obtained a set of 1129 MED-portions positioned in sub-parts of the VEGFR-2 binding site (*Fig.6*).

We've characterized the hybrids (duplicates are filtered) by Molecular weight (average and standard deviation) at each iteration (step). The hybrids are growing at each step, lead-likeness (rule of three) is decreasing quickly as the optimization proceeds and the drug-likeness (rule of 5) is higher than 60% within the hybrids (*Fig.7*).

To assess the diversity and potential affinity to protein kinases of the hybrids, we've analyzed the results in terms of unique scaffolds. We've found 3585 unique scaffold (average MW =394). 50 scaffolds are found in the PDB and 298 in PubChem (*Fig.8*.). 25 scaffolds are found in the active molecules of the protein kinases biossays, 3 of them are not found in the PDB.

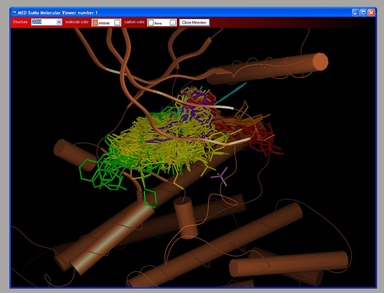


Fig.6. 2oh4 is show in orange cartoon view. Portions of ligands (MED-portions) are shown in 3D in the reference frame of 2oh4 which is shown as a cartoon view in orange. Coloring is done according to a MED-portions environment clustering.

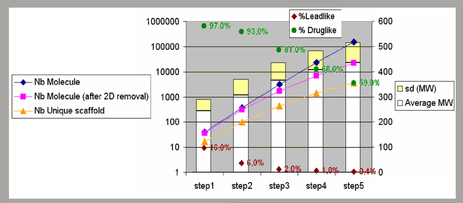


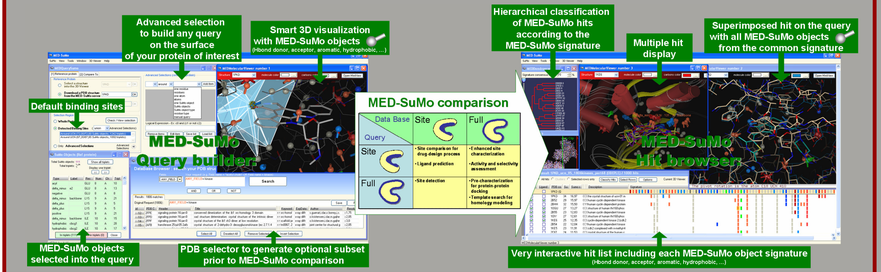
Fig 7 : Characterization of the hybrids in terms of Molecular Weight (average and standard deviation histogram). Left log. scale is related to the scaffold count and right scale is related to MW.

Fig 8: (left) Unique hybrids' scaffold found in PDB and Pubchem. (right) Unique hybrids' scaffold found in protein kinases bioassay and among them thos ewhich are not found in the PDB.

Conclusion

Based on the detection of similar 3D interactions between proteins and fragments into the PDB, our MED-Hybridiser protocol is able to generate hybrids which are likely to be DFG-out ligands. Hybrids are diverse: 1% (50/3585) of the generated scaffolds are found within the PDB. 50 scaffolds are found within the active molecules from the protein kinase Pubchem bioassays, 3 of them are new compare to the PDB.

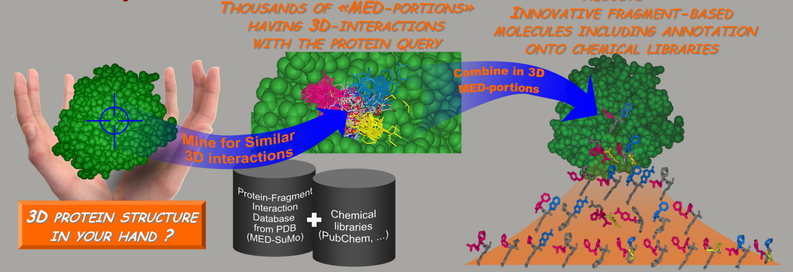
Annexe on MED-SuMo user interfaces:



Color coding of MED-SuMo SCF (Surface Chemical Features):

imidazole	positive	delta_plus	glycine
amide	struct_water	aromatic	proline
guanidinium	acyl	hydrophobic	thioether
hydroxyl	other	negative	delta_minus
thiol			

Annexe on MED-Hybridiser :



References:

[1] Hajduk PJ, Greer J. "A decade of fragment-based drug design: strategic advances and lessons learned" Nat Rev Drug Discov. 2007 Mar;6(3):211-9.

[2] Jambon M, Andrieu O, Combet C, Deléage G, Delfaud F, Geourjon C « The SuMo server : 3D search for protein functional sites" Bioinformatics Vol 21, n°20, 3929-3930 2005.

[3] Pierce AC, Bemis GW, "BREED: generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease" J. Med. Chem. 2004 May 20;47(11):2768-75

[4] Jambon M., Imberty A., Deleage G., Geourjon C. (2003) A new bioinformatic approach to detect common 3D sites in protein structures, Proteins, 52:137-134.