BRIEFINGS IN BIOINFORMATICS. VOL 12. NO 4. 336–340
Advance Access published on 21 April 2011
doi:10.1093/bib/bbr017

Identify drug repurposing candidates by mining the Protein Data Bank

Fabrice Moriaud, Stéphane B. Richard, Stewart A. Adcock, Laetitia Chanas-Martin, Jean-Sébastien Surgand, Marouane Ben Jelloul and François Delfaud

Submitted: 2nd December 2010; Received (in revised form): 6th March 2011

Abstract

Predicting off-targets by computational methods is gaining increasing interest in early-stage drug discovery. Here, we present a computational method based on full 3D comparisons of 3D structures. When a similar binding site is detected in the Protein Data Bank (PDB) (or any protein structure database), it is possible that the corresponding ligand also binds to that similar site. On one hand, this target hopping case is probably rare because it requires a high similarity between the binding sites. On the other hand, it could be a strong rational evidence to highlight possible off-target reactions and possibly a potential undesired side effect. This target-based drug repurposing can be extended a significant step further with the capability of searching the full surface of all proteins in the PDB, and therefore not relying on pocket detection. Using this approach, we describe how MED-SuMo reproduces the repurposing of tadalafil from PDE5A to PDE4A and a structure of PDE4A with tadalafil. Searching for local protein similarities generates more hits than for whole binding site similarities and therefore fragment repurposing is more likely to occur than for drug-sized compounds. In this work, we illustrate that by mining the PDB for proteins sharing similarities with the hinge region of protein kinases. The experimentally validated examples, biotin carboxylase and synapsin, are retrieved. Further to fragment repurposing, this approach can be applied to the detection of druggable sites from 3D structures. This is illustrated with detection of the protein kinase hinge motif in the HIV-RT non-nucleosidic allosteric site.

Keywords: *drug repurposing; PDB; pocket mining; drug design; protein similarities; target based*

INTRODUCTION

Drug repurposing is supported by the core observation that a single drug often interacts with multiple targets. It offers an appealing strategy; it enables reuse of existing therapeutic compounds and all the more those that were considered 'dirty' [1] because of the side effects they induce.

Corresponding author. Fabrice Moriaud, MEDIT SA, 2 rue du Belvedere, 91120 Palaiseau, France. Tel: +33 (0)160148743; Fax: +33 (0)183636746; E-mail: fmoriaud@medit.fr

Fabrice Moriaud, Chief Scientific Officer, jointly leads MEDIT SA with responsibility for the company's R&D and drug design services. Fabrice was a postdoctoral fellow at Sanofi-Synthelabo (Strasbourg, France) and holds a PhD in chemistry from the Joseph Fourier University (Grenoble, France).

Stéphane B. Richard graduated his PhD in X-ray crystallography from the University Joseph Fourier (Grenoble, France). He conducted 10 years of Chemical Biology Research at the Salk Institute for Biological Sciences (La Jolla, CA) as a Post-Doc and Staff Scientist. He is now MEDIT SA's Vice President of Business Development.

Stewart A. Adcock has overall responsibility for development and implementation of MEDIT's world-class product portfolio. He holds a DPhil from the Department of Physical and Theoretical Chemistry at Oxford University.

Laetitia Chanas-Martin, Application Scientist at MEDIT, manages MEDIT's customer support tasks, collaborative projects and research contracts. She holds a PhD in Molecular Modelling from the University of Montpellier II, France. She was hosted at the Centre de Biochimie Structurale.

Jean-Sébastien Surgand, R&D software developer at MEDIT. He holds a PhD in chemistry from the University of Strasbourg.

Marouane Ben Jelloul, R&D software developer at MEDIT. He was R&D engineer at Structural Biology Group at the Netherlands Cancer Institute (Amsterdam, The Netherlands) and holds Master Degree in physics from Paris 11 University.

François Delfaud, CEO MEDIT, founded MEDIT SA in 2003. He holds a master of Theoretical Chemistry and Informatics from Orsay University (France). He prepared a thesis at Synthelabo-Recherche in Molecular Dynamics, and worked for 5 year at Accelrys France.

Resolved 3D protein structures are a major source of information for understanding protein functional properties. The current explosive growth of publicly available protein structures in the Protein Data Bank (PDB) [2] is producing massive volumes of data for computational modelling and drug design methods. Target-based *in silico* drug design tools aid in the design and optimization of compounds to bind to specific targets. MED-SuMo is a technology for comparing protein surfaces, allowing structural similarities to be discovered and explored. The ligands are positioned using the assumption that identical ligands are likely to bind in the same protein environment with the same relative pose. Despite the numerous methods described for measuring 3D similarities between protein–ligand binding sites [3, 4], there are still very few reports of predictive target identifications by systematic binding site comparisons [4].

In a preliminary work, we described the application of MED-SuMo to target-based drug repurposing within the protein kinase superfamily using the complex of B-RAF with sorafenib as a query [5]. In this work, we describe this same application with another example using the complex of phosphodiesterase (PDE) with tadalafil. The full surface of every protein in the PDB was explored. The results were annotated with Uniprot and PFAM information, allowing the user to decide whether the candidate target is either a new potential application for this drug or a potential side effect to take into account. This approach is successful when a highly similar binding site is found in the PDB (or any macromolecular structure database). This high similarity is virtually only found within the same protein family or proteins known to share high binding site similarities like the HSP90 fold [6] or the aspartic proteases. Repurposing from one target to another of a different superfamily has met more success using ligand- and activity-based methods using sets of active ligands on various targets. The pioneering work of Vieth *et al*. [7] defined a metric between protein kinase binding sites based on common ligands and their IC_{50}. This innovative chemical view of biology has been applied successfully to drug repurposing with a different approach by Keiser *et al*. [8], by computing a pairwise similarity metric between ligand sets of various targets.

Target-based approaches are regaining interest when applied to fragment repurposing rather than to typical drug-like ligands. Local similarity between two proteins from different superfamilies is more likely to be found than a site similarity. Therefore, target-based repurposing is more likely to identify target candidates to repurpose fragments interacting with a local protein region. In this work, we describe how MED-SuMo can go beyond the limitation of searching for similar binding sites among a set of known binding sites by searching for local surface similarities in full protein structures. The potential binding of the drug and its pose in the target is implied by mining the PDB for fragments of the drug from all PDB ligands in a fragment-based approach [2, 9]. In the second part of this work, we searched exhaustively the full surface of all proteins in the PDB for regions similar to a given protein kinase hinge using MED-SuMo software, independently of any pocket detection. We identified tens of potential targets (data not provided) including two experimentally validated examples (see Target-based drug fragment repurposing).

In general, results of fragment repurposing can also be used to evaluate binding site druggability: for example, the detection of regions similar to the hinge of a validated druggable protein kinase is a guide for the detection of other druggable sites even if these were reported with no co-crystallized ligands. This is illustrated here by the detection of a local protein kinase hinge motif in the non-nucleosidic allosteric site of HIV-RT.

The repurposed fragment can be combined by hybridization to design drug-like compounds. Known active molecules, drugs or close analogues were retrieved using protein kinase [2], G protein-coupled receptor (GPCR) [2] and kinesin [9] as queries. Interestingly, repurposed fragments originate from the same protein superfamily as the query and from other superfamilies.

TARGET-BASED DRUG REPURPOSING

The target-based drug repurposing in MED-SuMo is implemented by building a query with the protein environment of a ligand and submitting it to a database containing the full surface of all PDB structures. The similarity between cavities is quantified by the MED-SuMo score, which is a metric for protein surface similarity. The higher the score, the higher is the protein similarity. This is an absolute value closely related to the number of matching surface chemical features (SCFs) between the proteins. SCFs represent

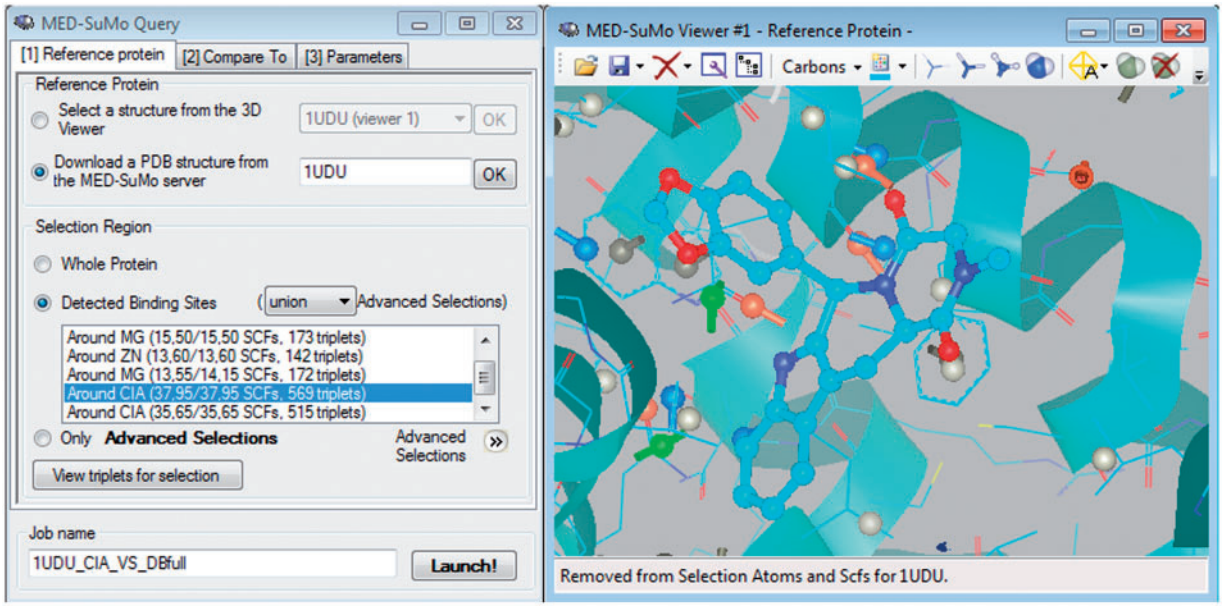

Figure 1: Building of the MED-SuMo query using the structure containing the drug tadalafil in complex with human PDE5A (1udu). In the MED-SuMo Query window, the list of bound ligands is provided and the first occurrence of tadalafil (CIA) is selected. The corresponding MED-SuMo query is a set of SCFs, which are shown in the right panel as balls and ball and stick features within a 6 Å distance from tadalafil.

possible chemical interactions that might be formed at the protein surface, analogous to a pharmacophoric representation of small molecules.

To illustrate this approach, we considered the drug tadalafil. This compound targets two proteins: PDE4A and PDE5A. We used the structure of tadalafil bound to human PDE5A (PDB code 1udu) to build the query (Figure 1). The first human PDE4a hit (3i8v) is found in the hit list at Rank 20 after others: PDE4, PDE2 and PDE10. The size of the ligand and the relative flexibility of the pocket imply steric clashes between tadalafil and PDE4A, mostly between the 1,3-benzodioxole moiety and M549. The geometry of the complex was therefore optimized (ligand and neighbouring protein sidechains) using the MMFF94 forcefield. In this way, steric clashes were eliminated while keeping the same binding mode (Figure 2).

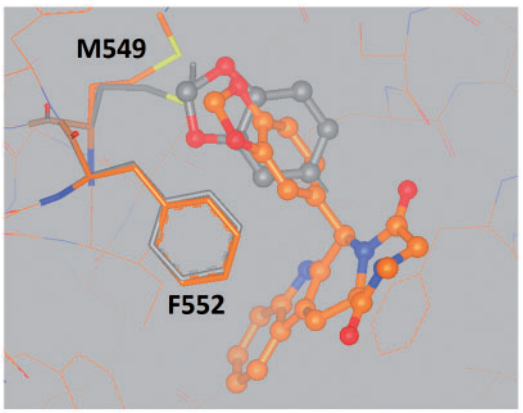

Figure 2: Structure of the complex between PDE4A and tadalafil (carbon atoms in orange) resulting from the repurposing with MED-SuMo of tadalafil from PDE5A. Tadalafil is rendered with ball and sticks. Coordinates of some atoms before geometry optimization are rendered with darker carbons.

TARGET-BASED DRUG FRAGMENT REPURPOSING

Repurposing ligands as described above is limited to similar binding sites. Therefore, repurposing fragments of ligands is more likely to occur because similar sub-pockets are more often found than similar binding sites. To rank the repurposed fragments, we use the piecewise linear potential $(PLP)_{inter}$ scoring function [10], which was identified in a previous work as a simple and successful function to rank repurposed fragments and hybrids of them [9]. PLP_{inter} consists of the intermolecular terms only of the PLP protein-ligand scoring function.

In this work, we illustrate this fragment repurposing process using as a query the hinge region of a protein kinase. This region is interesting because it is a shared motif between protein kinase, a validated druggable target and other proteins belonging to

different superfamilies. The MED-SuMo query is the 6 Å protein environment of the PDB ligand CIG (2-amino-6-chloropyrazine) of a cdk2 protein kinase structure (1WCC). This ligand is a small compound bound to the hinge with a chloro substituent pointing towards the gatekeeper (Phe80). The aim is to search the full surface of all proteins of the PDB and look for a similar motif. The advantage of searching a database describing the full surface of proteins is that the results are independent of pocket detection and also of the size and binding mode of the hit PDB ligand.

The results consist of a list of hits ordered by decreasing MED-SuMo score. In this particular case, all hits with a MED-SuMo score >6.1 (795 hits) are seen to match the hinge backbone. Those hits belong to the protein kinase family along with eight other protein superfamilies. Among those eight (data not provided), the one with the highest score is a biotin carboxylase (2w71), ranked 271 with a MED-SuMo score of 8.6. Interestingly, this similarity between protein kinase and biotin carboxylase is in agreement with experimental results recently reported in a publication where biotin carboxylase inhibitors were derived from a protein kinase inhibitor pharmacophore [11].

Another experimental validation of the similarity between protein kinase hinge and a protein belonging to a different family is the case of Synapsin I [12]. This superposition is identified by MED-SuMo with a MED-SuMo score of 5.0 (rank 1199) and a $PLP_{inter} = -21$. Sorting the hits by PLP_{inter} put Synapsin III (2p0a) at Rank 1086 with a $PLP_{inter} = -21$, which is a similar rank to the one obtained with the MED-SuMo score. Among those 1086 hits above synapsin, there are 60 unique PFAM codes that we are currently analysing, including biotin carboxylase, which are therefore identified as potential drug targets for hinge protein kinase fragment repurposing. Among those hits, some have no ligand bound, highlighting the advantage of searching the full surface of proteins rather than only the known binding sites.

Beyond repurposing of protein kinase fragments, this approach could be used to identify druggable sites by detecting known druggable protein motifs across structural databases. The hit with the best $PLP_{inter} = -43$ score is the non-nucleoside HIV reverse transcriptase binding site with the PDB ligand SPP bound (1KLM). The MED-SuMo score is 3.6 and the corresponding rank is 4251, in that case the PLP_{inter} is therefore of a great help to identify this local similarity. In Figure 3, the 3D superposition of cdk2 ATP site with the HIV-RT hit with the highest MED-SuMo score and the the HIV-RT non-nucleoside binding site is shown. The local similarity is high, based on the structural alignment with MED-SuMo of the hinge, together with the matching of the hydrophobic side chains Val106 with L134 and L100 with F80. The similarity is also observed on the ligand side, which explains the high scoring of CIG in HIV-RT. The ligands of the kinase and HIV-RT have a similar H-bond acceptor interacting with the NH H-bond donor from the backbone.

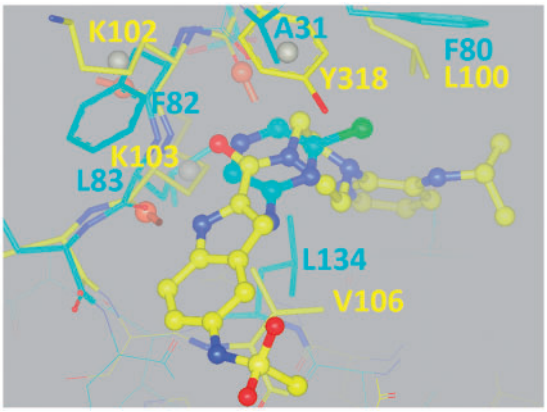

Figure 3: Superimposition of the protein kinase cdk2 (pdb code 1wcc) CIG ligand environment to the full surface of HIV-RT (pdb code 1klm). A region similar to the protein kinase hinge is detected in the non-nucleoside HIV reverse transcriptase binding site occupied by the PDB ligand SPP. The kinase hinge region (residue 82–84) is matching HIV-RT (102–104), both are depicted in sticks. Hydrophobic residues are matching between 1wcc and 1klm: Leu134 and Val106; Phe80 and Leu100; Ala31 and Tyr 318. The ligands CIG and SPP are rendered in ball and sticks.

Key Points

- Our target-based software can identify cases of drug repurposing on a rational structural basis.
- Repurposing relies on the detection of 3D similarities between proteins and not on docking or scoring functions.
- The entire PDB can be mined exhaustively in 3D for ligand repurposing. Example of PDE4/PDE5 is provided.
- The ligand repurposing protocol can be extended as a fragment-based approach. Example of a protein kinase ligand fragment repurposed in biotin carboxylase, synapsin is given.
- The protein kinase hinge region is detected mostly, but not exclusively, in purine binding sites. Its detection in the HIV-RT non-nucleosidic allosteric site indicates that target-based fragment repurposing could serve also as a druggable site detection method.

References

1. Grau D, Phil M, Serbedzija G. Innovative strategies for drug repurposing. *Drug Discov Dev* 2005. http://www.dddmag.com/innovative-strategies-for-drug.aspx (18 May 2005, date last accessed).
2. Moriaud F, Doppelt-Azeroual O, Martin L, *et al*. Computational fragment-based approach at PDB scale by protein local similarity. *J Chem Inf Model* 2009;**49**:280–94.
3. Berman HM, Westbrook J, Feng Z, *et al*. The Protein Data Bank. *Nucleic Acids Res* 2000;**28**:235–42.
4. Rognan D. Structure-based approaches to target fishing and ligand profiling. *Mol Inf* 2010;**29**:176–87.
5. Doppelt-Azeroual O, Moriaud F, Adcock SA, *et al*. A review of MED-SuMo applications. *Infect Disord Drug Targets* 2009;**9**:344–57.
6. Doppelt-Azeroual O, Moriaud F, Delfaud F, *et al*. Analysis of HSP90-related folds with MED-SuMo classification approach. *Drug Des Devel Ther* 2009;**3**:59–72.
7. Vieth M, Higgs RE, Robertson DH, *et al*. Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochim Biophys Acta* 2004;**1697**:243–57.
8. Keiser MJ, Setola V, Irwin JJ, *et al*. Predicting new molecular targets for known drugs. *Nature* 2009;**462**:175–81.
9. Oguievetskaia K, Martin-Chanas L, Vorotyntsev A, *et al*. Computational fragment-based drug design to explore the hydrophobic sub-pocket of the mitotic kinesin Eg5 allosteric binding site. *J Comput Aided Mol Des* 2009;**23**:571–82.
10. Gehlhaar DK, Verkhivker GM, Rejto PA, *et al*. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* 1995;**2**:317–24.
11. Miller JR, Dunham S, Mochalkin I, *et al*. A class of selective antibacterials derived from a protein kinase inhibitor pharmacophore. *Proc Natl Acad Sci USA* 2009;**106**:1737–42.
12. Defranchi E, Schalon C, Messa M, *et al*. Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements. *PLoS One* 2010;**5**. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0012214.