Computational fragment-based drug design to explore the hydrophobic sub-pocket of the mitotic kinesin Eg5 allosteric binding site

Ksenia Oguievetskaia [1 * #], Laetitia Martin-Chanas [1 #], Artem Vorotyntsev [1], Olivia Doppelt-Azeroual [1 2], Xavier Brotel [1], Stewart Adcock [1], Alexandre G. De Brevern [2], Francois Delfaud [1], Fabrice Moriaud [1]

[1] *Molecular Extended Distribution in Information Technology MEDIT SA , 2 rue du Belvédère, 91120 Palaiseau,FR*

[2] *Protéines de la membrane érythrocytaire et homologues non-érythroides INSERM : U665 , Université Paris VII - Paris Diderot , INTS , INTS 6, Rue Alexandre Cabanel 75739 PARIS CEDEX 15,FR*

* Correspondence should be addressed to: Ksenia Oguievetskaia <ksenia@medit.fr >

# These authors contributed equally to this work

Abstract

Eg5, a mitotic kinesin exclusively involved in the formation and function of the mitotic spindle has attracted interest as an anticancer drug target. Eg5 is co-crystallized with several inhibitors bound to its allosteric binding pocket. Each of these occupies a pocket formed by loop 5/helix α2 (L5/α2). Recently designed inhibitors additionally occupy a hydrophobic pocket of this site. The goal of the present study was to explore this hydrophobic pocket with our MED-SuMo fragment based protocol, and thus discover novel chemical structures that might bind as inhibitors. The MED-SuMo software is able to compare and superimpose similar interaction surfaces upon the whole PDB. In a fragment based protocol, MED-SuMo retrieves MED-Portions that encode protein-fragment binding sites and are derived from cross-mining protein-ligand structures with libraries of small molecules. We have excluded intra-family MED-Portions derived from ligands that occupy the hydrophobic pocket and predicted new potential ligands that would fill simultaneously both pockets. Some of the latter are identified in libraries of synthetically accessible molecules by the MED-Search software.

Author Keywords fragment-based drug design ; FBDD ; anti-mitotic ; Eg5, KSP ; mitotic kinesins.

Introduction

During metaphase, the mitotic spindle maintains a constant shape and size though their principal constituent, the microtubules are continuously being polymerized, depolymerized and transported towards the two spindle poles [1 –3]. Motor proteins from the kinesin family are involved in the mitotic spindle assembly, as well as an important number of cellular processes such as intracellular vesicles transport, chromosome segregation, cell division and motility in a tight collaboration with proteins of the cytoskeleton tubulin and actin [4 , 5].

Inhibition of the mitotic spindle formation is an interesting target in cancer chemotherapy. Anti-mitotic agents used to date for cancer treatment, target microtubule stability (*e.* g., Vinca alkaloids and taxanes). They cause serious side-effects, such as neurotoxicity (reviewed in [6]). Furthermore the development of resistance against these anti-mitotic agents restricts their application.

Another approach to inhibit the mitotic spindle formation has emerged. This alternative acts by inhibiting the mitotic motors that interact with microtubules [7]. Eg5 (also known as Kinesin Spindle Protein, KSP, or kinesin-5s) has emerged from recent studies as a leading candidate for targeted anti-mitotic therapies. This kinesin is exclusively involved in the formation and function of the mitotic spindle, by driving a relative sliding of microtubules in the mitotic spindle (for review see [8]). The inhibition of Eg5 leads to cells with monopolar spindles (monaster) during mitosis, resulting in a cell cycle arrest and apoptosis, without interfering with other microtubule dependent processes. Several small molecules are now known that inhibit human Eg5 by binding to its catalytic motor domain. These discoveries are attracting a potential interest in this protein as an anticancer drug target.

The first inhibitor discovered by the groups of Stuart Schreiber and Tim Mitchison is a small cell permeable molecule, monastrol [7]. One year after the resolution of the structure of the unbound Eg5 [9], the structure of the Eg5 bound to monastrol was also solved [10]. Monastrol binds specifically to the Eg5 and allosterically inhibits the ATPase activity of the kinesin. Further studies brought several inhibitors into clinical trials and several structures of the Eg5 with diverse types of inhibitors are now solved (Figure 1) [11 –16]. Most of these inhibitors interact with an hydrophobic (HYD) pocket that is inoccupied by the monastrol (Figure 2). This difference in the binding mode results in a ten fold increase of the affinity in the case of the monastrol and its homologue mon-97 (PDB code 2ieh): *(S)* -monastrol inhibits the basal ATPase activity with an IC_{50} value of 1.7 μM, whereas the *(R)* -mon-97 has an IC_{50} value of 110 nM [17]. The current knowledge of the available inhibitors for this target enlarges the scope for rational drug design approaches to discover more potent inhibitors for the Eg5.

The SuMo software was originally developed to detect structural similarities between proteins, in particular protein binding sites [18]. MED-SuMo, the commercially available version for drug design applications, is able to detect potential binding sites, characterize an unknown protein binding site and probe active-site selectivity within a protein family. In MED-SuMo, all chemical groups that can make interactions with other molecules are represented by Surface Chemical Features (SCFs) such as hydrogen-bond donor and acceptor, aromatic group, hydrophobic group, and so forth. Once detected, all neighboring SCFs are grouped in triplets which represent nodes of a graph encoding the protein interacting surface. The graph is then compared to a precompiled database of similarly constructed graphs (Figure 3 , [18 , 19]). MED-SuMo is extremely efficient, allowing application to the entire PDB, and therefore maximizing the probability of finding relevant local similarities on the protein binding site [18 –22]. The number of macromolecular structures publicly available from the Protein Data Bank (PDB) is raising quadratically, with 7 065 structures deposited in 2008 and more than 500 new entries currently released every month. In total, more than 55 000 macromolecule structures are accessible as of January 27th, 2009 (http://www.rcsb.org/pdb/statistics).

Recently, we developed a fragment based drug design approach based on MED-SuMo/MED-Fragmentor/MED-Hybridise technologies [21]. For this new approach, the MED-Fragmentor software component was designed to generate a MED-SuMo database of MED-Portions describing local protein-fragment interactions by cross-mining the PDB and a chemical library of small molecules. The MED-Portions are defined by several criteria including chemical moiety represented in 3D and the corresponding SCF environment. Local graph similarities are found using the MED-SuMo algorithm thus retrieving from the database MED-Portions aligned in 3D to the query structure. These MED-Portions can be combined using their 3D coordinates with the MED-Hybridise software.

The goal of this study was to design inhibitors of Eg5 by populating the HYD pocket of the allosteric binding site. Our fragment-based protocol is applied to populate the HYD pocket with chemical moieties of inter-family MED-Portions that MED-SuMo retrieves by local similarity search upon the whole PDB. Intra-family MED-Portions are discarded to bias towards the generation of new ligands compared to the known kinesin PDB ligands. The retrieved MED-Portions were combined using the MED-Hybridise software and 3D substructure filtering rules to focus on the most promising hybrids having similar binding mode to known inhibitors. This protocol allowed us to generate hybrid molecules that can be new potential ligands, among which, a certain number can be retrieved by MED-Search software in publicly available chemical libraries, ChemBridge, Maybridge, and PubChem database [23 –25].

Results and discussion

Definition of the custom SCF dictionary

Several water molecules are observed in the allosteric binding site of Eg5, especially inside the HYD pocket. The default MED-SuMo dictionary does not consider these molecules for the graph generation of the binding site. They could however bring important structural information, when they are involved in hydrogen bond bridging the ligand and the protein target. We therefore decided to add H-bond donors and H-bond acceptors on the water molecules defined in the PDB file. As a result these can be superposed with the H-bond donors and acceptors on either the water molecules or the protein SCFs. Since the graph generation of the database is done at 4.5 Å around the ligand, only the water molecules in the near neigbourhood of the ligand are taken into account, those mostly involved in H-bond interactions.

With this modified dictionary each binding site is defined by a larger number of SCFs, since water molecules can be considered as H-bond donors or acceptors. In addition the SCFs that are already engaged in intra-protein H-bond are not eliminated, resulting in formation of graphs containing a greater number of triplets. For example, for the query defined by any protein atom at 10 Å around monastrol (NAT ligand) in the 1x88 structure [26], 125 SCFs are considered in the default database and form 3 918 triplets, compared to the modified dictionary database where 187 SCFs are considered and form 6 090 triplets. This increasing number of SCF in query signature results generally in hits having larger number of SCFs and by consequence with greater MED-SuMo score [27]. In this work, all H-bond donors and acceptors SCFs are chemical features without any specified orientation.

This new SCF dictionary is able to better detect conserved secondary structures surrounding binding site, and to better populate the binding sites or surfaces where water molecules are present. An inconvenient consequence is that some of the SCFs considered are already engaged in an H-bond and could not interact with MED-Portions chemical moieties. This approximation seems to be better than removing them because backbone carbonyl and NH can be involved in two H-bonds, one intra-protein and one with a water molecule. Nonetheless, we are able to retrieve with this modified dictionary database with additional pertinent MED-Portions (data not shown).

Choice of the Query structure and selection of collected MED-Portions

In order to predict new categories of inhibitors that could fill the HYD pocket, we have chosen to use a structure containing an inhibitor that does not fill the HYD pocket. We constructed the query at 10 Å around monastrol of the 1x88 entry in the PDB. The query is then submitted to MED-SuMo to be compared to the MED-Portions database built with the new dictionary described above (Figure 4). During MED-SuMo comparison an automatic calculation of steric clashes between query protein and retrieved MED-Portions chemical moieties has been done (as described in Background and methods). We kept only the MED-Portions chemical moieties that possessed

fewer than 10 bumps with the query once superimposed inside its binding pocket. To show that the protocol is capable of exploring the HYD pocket with inter-family MED-Portion chemical moieties, we have eliminated all the MED-Portion chemical moieties that are directly derived from co-crystallized ligands of Eg5 that are targeting this pocket (Figure 2 , see Background and Methods for detailed description). The retained MED-Portions (1092 MED-Portions chemical moieties) constitute the starting point for the next steps.

Hybridization and filtering of the MED-Portions into potential ligands

The retained MED-Portions chemical moieties were subjected to five hybridization and filtering steps with MED-Hybridise software [21]. Hybridization of all the MED-portions would produce a very large number of hybrids. Here we focused the hybrid generation by targeting the HYD pocket. At the same time the hybrids have been selected to keep interactions with the L5/α2 pocket to mimic the known inhibitors binding.

To do so a sublist of rings populating one or the other pockets, called 'seeds', has been selected from the MED-Portions (Figure 5 , 83 in the HYD pocket, 153 in the L5/α2 pocket, see Background and methods for detailed description). The first step consisted of the hybridisation of these 'seeds' with the whole list of MED-Portions chemical moieties using our Chain Combine algorithm of MED-Hybridise (described in [21]). At each following step the 'seeds' were hybridised to the original MED-Portions and the hybrids generated during the previous step. During each hybridisation step exact 3D duplicates (same chemical structure and same 3D coordinates) are discarded.

In addition we applied a 3D sub-structure filter keeping only the hybrids that explored either the HYD or the L5/α2 subpocket with at least one ring (as described in Background and methods). This rigorous filtering resulted in a limited number of hybrids (4115) that possessed the required characteristics (Table 1).

Description of generated hybrids derived from exclusively inter-family MED-Portions chemical moieties

The resulting hybrids have then been filtered for bad geometry and then minimized in the query binding site using CHARMm forcefield implemented in Discovery Studio 2.0 [28]. Several hybrids having a low energy after minimization are shown on Figure 6 . They all result from the hybridization of inter-family MED-Portions chemical moieties deriving from various ligands (Figure 6). All hybrids have been chosen to have an aromatic ring that binds in the L5/α2 pocket. Indeed, except for the bulky ligand (pyrrolo-triazine-4-one analogue) in 2gm1 [16], all the other Eg5 co-crystallized ligands from the PDB have an aromatic ring of five or six atoms in that pocket. In L5/α2 pocket several conserved residues are involved in the binding of the inhibitors in the solved crystal structures of Eg5 from the PDB: Glu116, Glu118, Arg119, Pro137 and Leu214 [17]. All those residues are also involved in the predicted binding of the selected hybrids (Figure 6). Given the restriction during the hybridisation protocol to keep hybrids targeting the HYD pocket, most of the hybrids are interacting with the hydrophobic residues Arg221, Leu160, Ile136 and several are forming T-stacking interactions with Phe239 (Figure 6(a)). Some hybrids are also predicted to be involved in H-bond interactions. The hybrid represented in Figure 6(a) (Hybrid 1215) has a third group that points toward the solvent like the *(S)*- monastrol ethyl ester group. The nitrogen of the amide is predicted to form a H-bond with the Tyr211. The hybrids represented in Figure 6(b) (Hybrid 969) and (c) (Hybrid 1317) are both predicted to form a hydrogen bond with Glu118 backbone as does the *(S)*- monastrol.

One of the MED-Portions chemical moieties is present in the three selected hybrids. It derives from a ligand co-crystallized in the 1q0z PDB structure [29] and is very similar to a sub-structure of the Eg5 inhibitor co-crystallized in 2fme [15]. However it is derived from a very different ligand (10-decarboxymethylaclacinomycin A) and is co-crystallized with a protein of the aclacinomycin methylesterase (RdmC) family.

The hybrids selected at that step are hit-like and can be considered as new leads compared to the known PDB ligands that could potentially bind to the allosteric site of Eg5.

Comparison of the obtained hybrids to compound libraries with MED-Search

Having generated new chemistry for potential kinesin inhibitors, we also checked if molecules synthetically accessible were found to be potential ligands targeting the HYD pocket. Hybrids generated at the last step were searched in diverse chemical libraries including known kinesin inhibitors, commercial libraries like the ChemBridge diverset, Maybridge hitfinder and the entire PubChem database [23 – 25]. A new module called MED-Search has been developed for that purpose. As hybrids can still contains Du atoms that were considered as wildcard atoms during hybridization, MED-Search needs to deal with that special type of atom.

To accomplish this task we needed an accurate search algorithm capable to treat Du atoms as wildcard atoms that could match any atom type or even nothing. Due to a huge amount of comparisons it was crucial to be able to pre-filter crudely and very quickly the possible matches. For that purpose, the SCINS representation [30 ,Jenkins, 2004 #27] of the molecules were compared. Only the retained matches are submitted to an exact comparison using the SMART representation [21]. During the search, if a match is found, the Du atoms of the hybrid are replaced by the real atoms of the matched molecule. All the molecules coming out of MED-Search do not have Du atoms,

and are known, chemically accessible molecules. Their 3D coordinates are kept and they are thus aligned in 3D in the reference frame of the query binding site.

We have also generated and searched the Murcko scaffolds computed according Bemis and Murcko [31]. For the generation of the scaffolds exocyclic double bonds were kept. The results of the comparison are listed in Table 1 .

Amongst the hybrids found exactly in the PubChem library several are predicted to bind both in the HYD and in the L5/α2 pockets (Figure 7). Some contain the substructure highlighted in the hybrid description coming from the MED-Portion derived from the 1q0z ligand. One of them has been minimized in the query binding site, while keeping he protein side chains fixed (Figure 8). This molecule does not make hyrdrophobic interaction in the back of the hydrophobic pocket but it interacts with all the important residues involved in the binding of monastrol analogs and has a different chemistry. It is predicted to be involved in two hydrogen bonds with Glu118 and Leu 214.

Summary and conclusion

Development of resistance in cancer therapies to drugs that target directly the microtubules, as well as their serious side-effects such as neurotoxicity have forced the research to chase for other targets that would more specifically inhibit the mitotic spindle function. Eg5, a mitotic kinesin exclusively involved in the formation and function of the mitotic spindle has attracted interest as an anticancer drug target.

Eg5 is co-crystallized with several inhibitors bound to its allosteric binding site. All of them occupy a pocket of this site formed by loop 5 and helix α 2 (L5/α2), some occupy as well the hydrophobic pocket (HYD) (Figure 2). The inhibitors that are binding to both pockets simultaneously have shown sometimes a ten fold higher affinity then the ones occupying only the L5/α2 pocket.

The goal of the present study was to obtain potential new ligands that explore this HYD pocket with our MED-SuMo fragment based protocol. We have excluded intra-family MED-Portions derived from ligands that occupy the HYD pocket and predicted new potential hybrids that would fill simultaneously both pockets. The MED-Portions chemical moieties have then been combined with the MED-Hybridise software. Some of the latter were found in chemical libraries with the MED-Search software and could therefore be tested experimentally. Based on structural information, the obtained hybrids bring new ideas about the possible scaffolds and chemical functions that could be binding the allosteric site of Eg5 and occupy both pockets.

Background and methods

The results presented in this paper were generated using software developed by MEDIT SA, except where stated otherwise.

MED-SuMo technology

The MED-SuMo technology is described briefly hereafter. For further details as well as for the manner of mining the PDB to extract MED-Portions, to populate protein surfaces with MED-Portions and to hybridize the MED-Portions refer to the papers and the thesis work of Jambon *et al.* [18 , 19 , 27] and a recent paper of Moriaud *et al* . describing fragment based drug design protocol derived from our core technology [21].

MED-SuMo is derived from the SuMo software [19]. MED-SuMo is able to locate similar regions on macromolecular surfaces associated with similar chemical groups. It applies a heuristic based on a 3D representation of macromolecular surfaces using Surface Chemical Features (SCFs) like, for example, H-bond donor, H-bond acceptor, formal positive and negative charges, hydrophobic, aromatic, or more specific features like amide and guanidinium. Each feature describes a putative physical interaction including its precise geometrical characteristics. The MED-SuMo comparison methodology can be divided into two major steps: **(1) The Graph Formation:** SCFs are displayed on the protein structure through a lexicographic analysis of the atoms in the PDB files, *i.e.*, a residue is represented by a set of representative SCFs. The correspondence between the SCFs and the chemical groups is stored in the MED-SuMo dictionary. SCFs are assembled into triplets with specific geometric characteristics. Such geometric characteristics include edge length, perimeter length, and angles. The triplet network is then stored as a graph data structure. The triplets are vertices and the edges are connecting adjacent triplets in the MED-SuMo database. **(2) The Graph Comparison:** To compare two graphs, MED-SuMo looks for compatible triplets; composed of compatible SCF. These triplets are called comparison 'seeds'. When a seed is detected, MED-SuMo extends the comparisons to the vertices of the neighbourhood, until no more similarities are found. These comparisons generate the formation of similar patches (common groups of SCFs) between two graphs, weighted by the MED-SuMo score. The MED-SuMo search heuristics is based also on two criteria that consider the SCFs 3D positions flexibility and so improve the search for compatible graphs. These comparisons are usually performed between a query and a database of precompiled graphs.

Steric clash filtering during the MED-SuMo comparison

During MED-SuMo comparison, a filter of steric clashes is applied. This is particularly important for our fragment based approach where we need to find relevant MED-Portions chemical moieties that could potentially bind the target. This filter is based on the count contacts between the query protein and MED-Portion chemical moieties.

Using SCF superimposition of the query protein and a retrieved hit, atoms of MED-Portions are placed in the query binding site. Distances between each atom of the MED-Portion chemical moiety, except the Du atoms, and the atoms of the query protein are calculated. If the distance between two atoms is shorter than the sum of their Van der Waals radii weighted by a coefficient a contact is registered. This coefficient was set to 0.8. This value was was empirically determined to allow flexibility in the clash detection. The filter can be set on the number of atoms clashing or on the total number of clashes for the given MED-Portions. Finally, if the number of clashes is greater than a user defined value, the MED-Portion is excluded.

In this study, only the MED-Portions chemical moieties that exhibit more than ten steric clashes with the query protein are discarded.

MED-Portion definition

MED-Portions are the MED-SuMo representation of protein-fragment patterns obtained by fragmentation of the protein-ligand structures using the MED-Fragmentor [21]. MED-Portions are defined by several criteria: (1) a chemical moiety where atoms are topologically matching a molecule of less then 250 Da from molecular libraries. Therefore, these molecules are expected to be synthetically accessible molecules or building blocks, (2) open valences filled by Du atoms that indicate where it was connected in the original ligand, and (3) the protein interaction surface surrounding that chemical moiety described by the SCFs. The pre-calculated MED-SuMo fragment data base contains 340 888 “binding-sites” and thus 340 888 MED-Portions. MED-Portion chemical moieties are collected in the reference frame of the query and form a pool which is used to generate new 3D hybrid compounds by hybridization.

Custom dictionary database

To use the structural information provided by water molecules crystallized into the binding sites, the MED-SuMo dictionary was modified. Indeed, MED-SuMo software allows the advanced user to modify the dictionary of SCF definitions by editing a text file.

Given the potential importance of the water molecules for the considered target, we have taken them into account. This allows them to match with either water molecules, or H-bond donors and acceptors on the hit protein structure.

In the classical MED-SuMo dictionary, H-bond donor and acceptor SCFs have a defined orientation. The H-bond donor SCF is also projected to the location of a potential H-bond acceptor. For water molecules it is not possible to derive any orientation or projection for the H-bond donor or acceptor groups. Therefore we added two new ponctual geometrical variant of H-bond acceptor and H-bond donor in the dictionary syntax. Both are neither orientated nor projected. In this new version of the MED-SuMo dictionary, the water molecules are considered as part of the receptor and are described with one H-bond acceptor and H-bond donor.

In MED-SuMo only chemical groups of the same type can be compared and possibly considered as equivalent. One aim of the new dictionary is to be able to match the H-bond donor and/or acceptor of the water molecules with similar groups on the proteins. Therefore former orientated H-bond donors and acceptors on the protein structure are replaced by the new punctual H-bond donor and acceptor. Once the SCFs are positioned on the macromolecular structure their positions and orientations are filtered. In the default MED-SuMo configuration a filtering step discards any SCFs likely to be involved in intra-molecular hydrogen bonds. With the new dictionary this filter is skipped. However the SCFs too buried to interact with a potential ligand are still discarded.

Selection of collected MED-Portions

From the collected MED-Portions we eliminated the ones derived from the Eg5 ligands that are located in the targeted hydrophobic pocket (2gm1, 2g1q, 2q2y, 2uyi, 2uym, 2pg2, 2q2z, 1yrs, 2fl6, 2ieh, 2fky, 2fl2, 3cj0). MED-Portions derived from three structures of Eg5 were kept, since their ligands are not positioned in the HYD pocket: 1x88, 2fme and 1q0b. The selected MED-Portions' atoms were typed with OpenBabel [32].

Selection of seeds

To create new ligands that would bind simultaneously the targeted HYD and the L5/α2 pocket, we selected a sub-list of rings derived from the retrieved MED-Portions chemical moieties that explored these pockets as the starting points, the seeds. The seeds were selected as follows. We fragmented the retrieved MED-Portions in rings and linker parts. Then we filtered the rings, to keep only those that had at least one carbon atom inside one of the pockets. The emplacement of this atom was selected at 3 Å around the deepest one buried in the L5/α2 pocket of the 1x88 ligand and deepest one buried in the HYD pocket of the mon-97 ligand from the 2ieh PDB code structure as it was placed inside the 1x88 PDB code structure binding pocket after the superposition of the two structures by MED-SuMo (Figure 9). The selected rings were decorated with Du atoms at every possible position defined by implicit hydrogen atoms.

Hybridization parameters and substructure filtering

We applied 5 hybridization steps with the Chain Combine algorithm of MED-Hybridise [21]. The basic principle of Chain Combine is to look for overlapping bonds between a pair of aligned MED-Portions. The Du atoms of the MED-Portions are considered as wildcard atoms during the hybridization steps.

We have hybridized the seeds to the whole list of selected MED-Portions during the first step. Then we hybridized the seeds to the molecules obtained in the previous step. During steps 2 to 5 we used the seeds, without Du atoms for the substructure 3D filtering, by keeping only the hybrids that were superstructures of the one of those seeds. After each step we filtered the obtained hybrids for 3D duplicates (RMSD 2.0 during the first step, RMSD 4.0 during the 2nd to the 4th steps, and RMSD 6.0 afterwards). These 4841 molecules have been submitted to the MED-Search module as explained later.

In parallel, several filtering steps were applied, molecules with intra-molecular clashes and elevated strain energy, the ones containing a phosphate atom or a ribose are discarded. The remaining hybrids were minimized in the binding site of the query structure with the CHARMm forcefield included in Discovery Studio 2.0 [28]. The query structure is 1x88 used for the MED-SuMo run after removing all ligands and water molecules. During the minimization the receptor is considered rigid.

MED-Search module

The aim of the MED-Search software is to search the designed hybrid molecules that contain in many cases dummy atoms, versus a large compound library rapidly and accurately. It uses a two-step algorithm.

The first step rapidly skips molecules that can be trivially seen to not match without discarding false negatives. For reasons of efficiency, this step is performed via comparison of a novel mini-fingerprint that encode molecules' scaffold structures. The design of this mini fingerprint was inspired by the Novartis SCINS classifier [30 , 33]. It consists of an array of thirteen integer values, corresponding to the thirteen values in the SCINS classifier. The queries are read from the input file and SCINS fingerprints for these molecules are generated during the first loop of comparison against the first target molecule. Molecules are further compared with an exact comparison only when their Tanimoto similarity metric exceeds a 0.99 threshold value.

The second step is the exact structure search that analyses the connectivity of the compared molecules and then compares corresponding atom pairs. Two atoms are treated as matching if they have the same atom type (Sybyl-style atom type), if they are or not in a ring and if they are or not aromatic. These rules apply only to real atoms. Du atoms are considered as wildcard atoms and can match any atom or even nothing.

At this stage, Du atoms can be replaced with real atom types to complete the match. During the comparison, the algorithm tries first to superimpose the non-Du atoms of the query. If the exact matching is successful, then the mapping of Du atoms is done and they are replaced according to the corresponding atom in the hit. Where there is no possible match, the Du atom can be deleted.

The exact structure of 3077 hybrids were compared to the chemical libraries (20 million compounds) with MED-Search. This comparison took forty hours on eight CPUs (2 Xeon Quad Core 5335, 16 GB RAM).

Acknowledgements:

We thank Raphaël Guerois for providing us useful comments and suggestions for the initiation of this study (Commissariat à l'Energie Atomique (CEA), Institut de Biologie et Technologies de Saclay, and Centre National de la Recherche Scientifique (CNRS), Gif-sur-Yvette, F-91191, France). Olivia Doppelt-Azeroual is a Ph.D. student funded by the ANRT. This work was supported by the Carriocas collaborative project (http://www.carriocas.org/) and funded by the French office "Direction Générale des Entreprises".

Abbreviations

SCF : Surface Chemical Feature
PDB : Protein Data Bank
KSP : Kinesin Spindle Protein
HYD : Hydrophobic
L5/α2 : loop 5/helix α2
Å : Angström
Du : MED-Portion dummy atom

References:

- 1 . Mitchison TJ , Salmon ED . 2001 ; Mitosis: a history of division . Nat Cell Biol . 3 : (1) E17 - 21
- 2 . Mitchison T , Kirschner M . 1984 ; Dynamic instability of microtubule growth . Nature . 312 : (5991) 237 - 42
- 3 . Mitchison TJ . 1989 ; Polewards microtubule flux in the mitotic spindle: evidence from photoactivation of fluorescence . J Cell Biol . 109 : (2) 637 - 52

- 4 . Vale RD , Fletterick RJ . 1997 ; The design plan of kinesin motors . Annu Rev Cell Dev Biol . 13 : 745 - 77
- 5 . Amos LA , Cross RA . 1997 ; Structure and dynamics of molecular motors . Curr Opin Struct Biol . 7 : (2) 239 - 46
- 6 . Wood KW , Cornwell WD , Jackson JR . 2001 ; Past and future of the mitotic spindle as an oncology target . Curr Opin Pharmacol . 1 : (4) 370 - 7
- 7 . Mayer TU , Kapoor TM , Haggarty SJ , King RW , Schreiber SL , Mitchison TJ . 1999 ; Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen . Science . 286 : (5441) 971 - 4
- 8 . Kwok BH , Kapoor TM . 2007 ; Microtubule flux: drivers wanted . Curr Opin Cell Biol . 19 : (1) 36 - 42
- 9 . Turner J , Anderson R , Guo J , Beraud C , Fletterick R , Sakowicz R . 2001 ; Crystal structure of the mitotic spindle kinesin Eg5 reveals a novel conformation of the neck-linker . J Biol Chem . 276 : (27) 25496 - 502
- 10 . Maliga Z , Kapoor TM , Mitchison TJ . 2002 ; Evidence that monastrol is an allosteric inhibitor of the mitotic kinesin Eg5 . Chem Biol . 9 : (9) 989 - 96
- 11 . Yan Y , Sardana V , Xu B , Homnick C , Halczenko W , Buser CA , Schaber M , Hartman GD , Huber HE , Kuo LC . 2004 ; Inhibition of a mitotic motor protein: where, how, and conformational consequences . J Mol Biol . 335 : (2) 547 - 54
- 12 . Cox CD , Breslin MJ , Mariano BJ , Coleman PJ , Buser CA , Walsh ES , Hamilton K , Huber HE , Kohl NE , Torrent M , Yan Y , Kuo LC , Hartman GD . 2005 ; Kinesin spindle protein (KSP) inhibitors. Part 1: The discovery of 3,5-diaryl-4,5-dihydropyrazoles as potent and selective inhibitors of the mitotic kinesin KSP . Bioorg Med Chem Lett . 15 : (8) 2041 - 5
- 13 . Cox CD , Torrent M , Breslin MJ , Mariano BJ , Whitman DB , Coleman PJ , Buser CA , Walsh ES , Hamilton K , Schaber MD , Lobell RB , Tao W , South VJ , Kohl NE , Yan Y , Kuo LC , Prueksaritanont T , Slaughter DE , Li C , Mahan E , Lu B , Hartman GD . 2006 ; Kinesin spindle protein (KSP) inhibitors. Part 4: Structure-based design of 5-alkylamino-3,5-diaryl-4,5-dihydropyrazoles as potent, water-soluble inhibitors of the mitotic kinesin KSP . Bioorg Med Chem Lett . 16 : (12) 3175 - 9
- 14 . Fraley ME , Steen JT , Brnardic EJ , Arrington KL , Spencer KL , Hanney BA , Kim Y , Hartman GD , Stirdivant SM , Drakas BA , Rickert K , Walsh ES , Hamilton K , Buser CA , Hardwick J , Tao W , Beck SC , Mao X , Lobell RB , Sepp-Lorenzino L , Yan Y , Ikuta M , Munshi SK , Kuo LC , Kreatsoulas C . 2006 ; 3-(Indol-2-yl)indazoles as Chek1 kinase inhibitors: Optimization of potency and selectivity via substitution at C6 . Bioorg Med Chem Lett . 16 : (23) 6049 - 53
- 15 . Tarby CM , Kaltenbach RF 3rd , Huynh T , Pudzianowski A , Shen H , Ortega-Nanos M , Sheriff S , Newitt JA , McDonnell PA , Burford N , Fairchild CR , Vaccaro W , Chen Z , Borzilleri RM , Naglich J , Lombardo LJ , Gottardis M , Trainor GL , Roussell DL . 2006 ; Inhibitors of human mitotic kinesin Eg5: characterization of the 4-phenyl-tetrahydroisoquinoline lead series . Bioorg Med Chem Lett . 16 : (8) 2095 - 100
- 16 . Kim KS , Lu S , Cornelius LA , Lombardo LJ , Borzilleri RM , Schroeder GM , Sheng C , Rovnyak G , Crews D , Schmidt RJ , Williams DK , Bhide RS , Traeger SC , McDonnell PA , Mueller L , Sheriff S , Newitt JA , Pudzianowski AT , Yang Z , Wild R , Lee FY , Batorsky R , Ryder JS , Ortega-Nanos M , Shen H , Gottardis M , Roussell DL . 2006 ; Synthesis and SAR of pyrrolotriazine-4-one based Eg5 inhibitors . Bioorg Med Chem Lett . 16 : (15) 3937 - 42
- 17 . Garcia-Saez I , DeBonis S , Lopez R , Trucco F , Rousseau B , Thuery P , Kozielski F . 2007 ; Structure of human Eg5 in complex with a new monastrol-based inhibitor bound in the R configuration . J Biol Chem . 282 : (13) 9740 - 7
- 18 . Jambon M , Imberty A , Deleage G , Geourjon C . 2003 ; A new bioinformatic approach to detect common 3D sites in protein structures . Proteins . 52 : (2) 137 - 45
- 19 . Jambon M , Andrieu O , Combet C , Deleage G , Delfaud F , Geourjon C . 2005 ; The SuMo server: 3D search for protein functional sites . Bioinformatics . 21 : (20) 3929 - 30
- 20 . Doppelt O , Moriaud F , Bornot A , de Brevern AG . 2007 ; Functional annotation strategy for protein structures . Bioinformation . 1 : (9) 357 - 9
- 21 . Moriaud F , Doppelt-Azeroual O , Martin L , Oguievetskaia K , Koch K , Vorotyntsev A , Adcock SA , Delfaud F . 2009 ; Computational Fragment-Based Approach at PDB Scale by Protein Local Similarity . J Chem Inf Model .
- 22 . Doppelt-Azeroual O , Moriaud F , Delfaud F . 2009 ; MED-SuMo Applications . Infectious Disorders-Drug Targets .
- 23 . The PubChem Project . [cited June 9, 2008]; Available from: http://pubchem.ncbi.nlm.nih.gov
- 24 . ChemBridge . [cited; Available from: http://www.chembridge.com/
- 25 . Maybridge . [cited; Available from: http://www.maybridge.com/
- 26 . Maliga Z , Mitchison TJ . 2005 ; Structural Basis of Eg5 Inhibition by Monastrol .
- 27 . Jambon M . A bioinformatic system for searching functional similarities in 3D structures of proteins . 2003 ; Université Claude Bernard Lyon 1 ; Lyon
- 28 . Discovery Studio . Accelrys Software Inc ; 10188 Telesis Court, Suite 100 San Diego, CA 92121, USA: San Diego
- 29 . Jansson A , Niemi J , Mantsala P , Schneider G . 2003 ; Crystal structure of aclacinomycin methylesterase with bound product analogues: implications for anthracycline recognition and mechanism . J Biol Chem . 278 : (40) 39006 - 13
- 30 . Czerminski R . 2005 ; Using ROCS, EON, FEPOPS and 2D methods for prospective and retrospective similarity searching: experiences and results .
- 31 . Bemis GW , Murcko MA . 1996 ; The properties of known drugs. 1. Molecular frameworks . J Med Chem . 39 : (15) 2887 - 93
- 32 . The Open Babel Package . [cited July 5, 2008]; Available from: http://openbabel.sourceforge.net/
- 33 . Jenkins JL , Glick M , Davies JW . 2004 ; A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes . J Med Chem . 47 : (25) 6144 - 59

Fig. 1

2D representation of known Eg5 inhibitors. The 3,5-diaryl-4,5-dihydropyrazole (**a**) is a potent and selective inhibitor of Eg5 that is active in cells at low nanomolar concentrations. It was identified in a high throughput screening by Merck (Cox, et al. 2005). X-ray crystallographic evidence is presented which demonstrates that this inhibitor bind the same allosteric pocket as the monastrol (**b**) and mon-97 (**c**) of Eg5 that is distant from the nucleotide and microtubule binding sites. Further studies showed that by appending a propylamine substituent at the C5 carbon of a dihydropyrazole core, lead to compounds with increased potency and aqueous solubility (**d** and **e**)

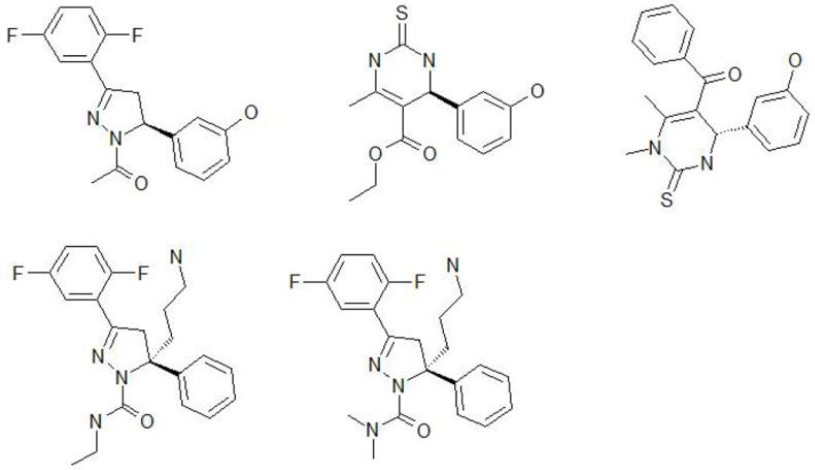

Fig. 2

(**a**) View of the Eg5 binding site surface from the 1x88 PDB structure showing the co-cristallized monastrol (green) and the mon-97 (clear blue) from the 2ieh PDB structure, retrieved from the superposition of the two protein structures with MED-SuMo. Both ligands occupy the sub-pocket formed by loop 5 (L5) and the helix α2 (α2). The mon-97 occupies as well the hydrophobic pocket HB shown in (**b**) Grey: hydrophobic, blue: hydrophilic

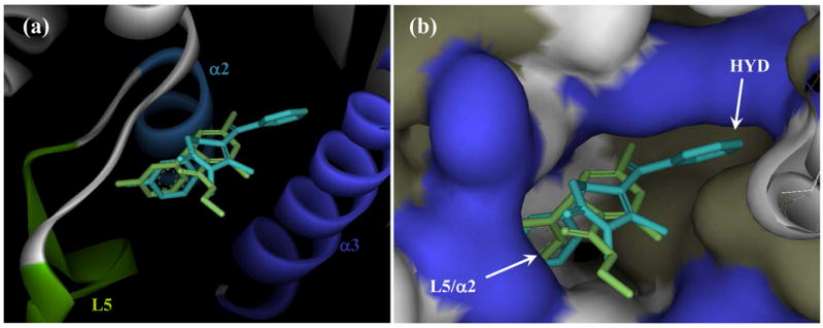

Fig. 3

MED-SuMo core algorithm: graph creation-comparison. 1. The protein molecular structure is scanned and mapped to a dictionary of physicochemical/geometric constructors that detect the MED-SuMo Surface Chemical Features (SCFs, represented by colored circles). 2. The geometric rules are applied to build a network of triplets (green triangle). 3. This network is transposed into a graph data structure. 4. The resulting graph is compared to the precompiled database of similarly constructed graphs

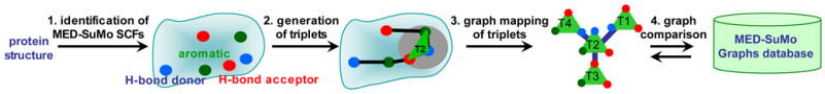

Fig. 4

Query definition. The binding site is defined at 10 Å around every atom of the (S)-monastrol (NAT) of the 1x88 structure, MED-SuMo detects 187 Surface Chemical Futures (SCF, described in Background and methods section) (**a**). By building the query with these SCFs, the graph contains 6090 triplets (**b**) which are then submitted to the database

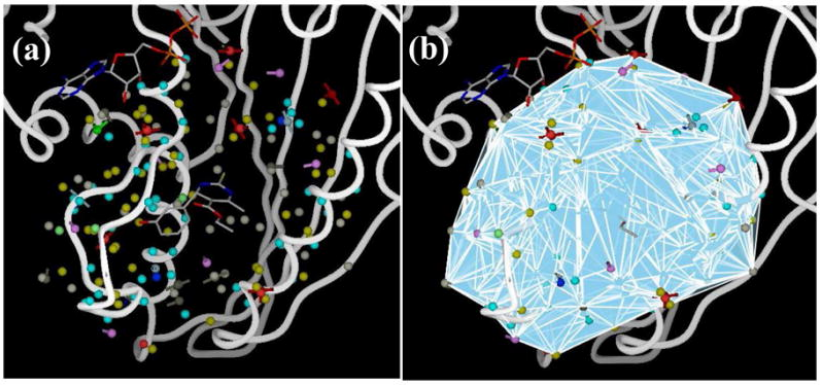

Fig. 5

Hybridisation seeds from the HYD pocket (salmon). The seeds used in the hybridization steps have been selected as described in Background and methods. Green: monastrol ligand from the 1x88 structure. Blue: mon-97 ligand from the 2ieh structure

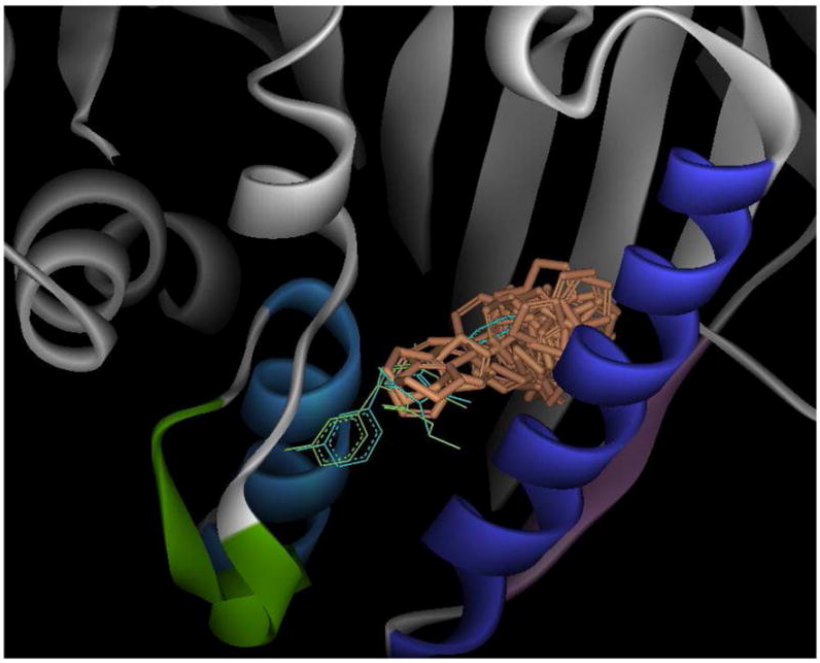

Fig. 6

Examples of the selected hybrids and the MED-SuMo chemical moieties that have been used for their formation presented in 3D coordinates, inside the 1x88 allosteric binding site. In L5/α2 pocket several conserved residues are involved in the binding to the inhibitor in all the crystal structures of Eg5 from the PDB: Glu116, Glu118, Arg119, Pro137 and Leu214. All those residues are also involved in the predicted binding of the selected hybrids. In the HYD pocket, all the hybrids are interacting with the hydrophobic residues Arg221, Leu160, Ile136 and especially are forming T-stacking interacting with Phe239. Hybrid 1215 represented in (**a**) has a third group that points toward the solvent like the (S)-monastrol ethyl ester group. The nitrogen of the amide is predicted to form a hydrogen bond with the Tyr211. Hybrid 1317 represented in (**b**) and hybrid 969 (**c**) are both predicted to form a hydrogen bond with Glu118 background as does the (S)-monastrol. For each of the hybrids above MED-Portions were derived from the following PDB structures : (**d**) 1fo4 in blue, 1n78 in orange, 1q0z in green and 2ox0 in red; (**e**) 1n78 in dark blue, 1r3w in pink, 1q0z in purple and 1oo5 in clear blue (**f**) 1r3w in clear blue, 1q0z in orange and 1eqb in green 1fo4: xanthine dehydrogenase, ligand 2-hydroxybenzoic acid 1n78: glutamyl-trna synthetase, ligand glutamol-amp 1q0z: aclacinomycin methylesterase, ligand 10-decarboxymethylaclacinomycin a 2ox0: histone demethylase, ligand n-oxalyolglycine 1r3w: uroporphyrinogen decarboxylase, ligand coproporphyrin iii 1oo5: nitroreductase prodrug-activating system, ligand flavin mononucleotide 1eqb: serine hydroxymethyltransferase, ligand 5-formyl-6-hydrofolic acid

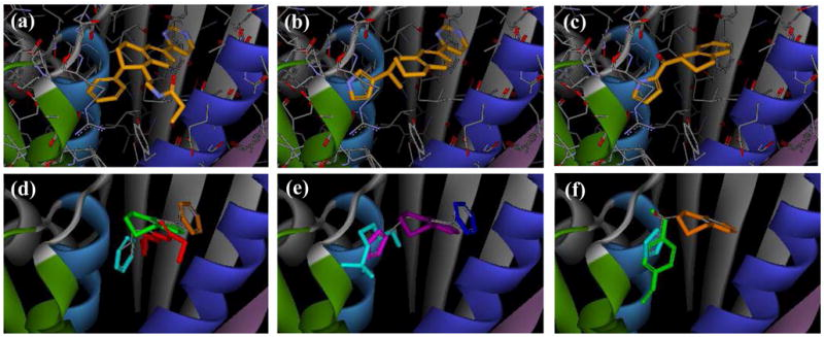

Fig. 7

Examples of hybrids found in PubChem with MED-Search. These molecules should be considered as scaffold to be decorated as they are often too hydrophobic to be tested.

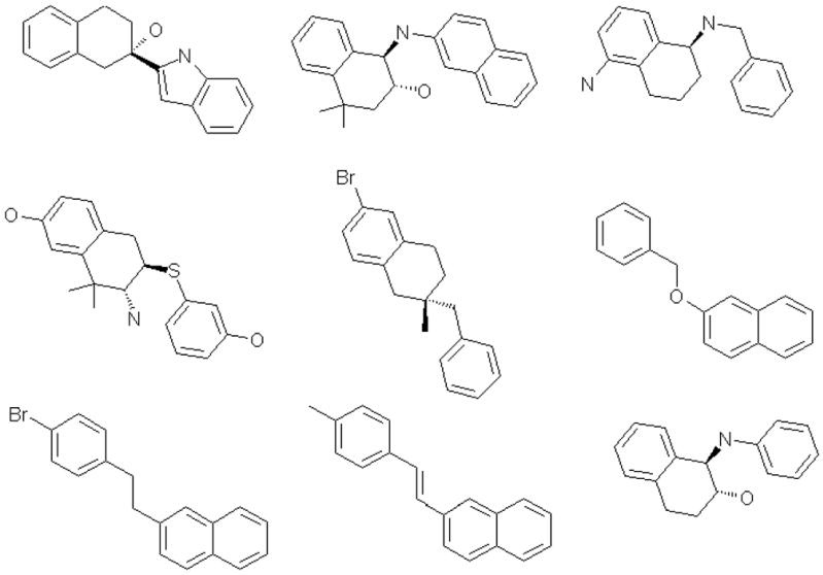

Fig. 8

(**a**) Example of a hybrid that matched a PubChem molecule. This hybrid is derived from three MED-Portions represented in their 3D coordinates (**b**) derived from the following PDB structures: 3pax in orange, 1q0z in green and 1gos in red 3pax: poly (ADP-ribose) polymerase, ligand 3-methoxybenzamide 1q0z: aclacinomycin methylesterase, ligand 10-decarboxymethylaclacinomycin a 1gos: human monoaminne oxidase b, ligand flavin-adenine dinucleotide

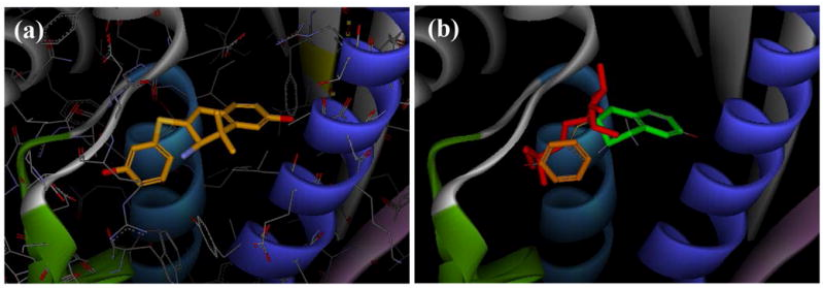

Fig. 9

Definition of the filtering options on a C-atom substructure inside the HYD pocket (violet sphere on the right) and the L5/α2 pockets (violet sphere on the left) of the Eg5 binding site surface from the 1x88 PDB structure. The ligand from the 1x88 structure, monastrol is represented in green. The mon-97 ligand (pale blue) from the 2ieh PDB structure that fills the HB pocket is represented in 3D coordinates after the alignment of the kinesin structures of 2ieh with 1x88 with MED-SuMo

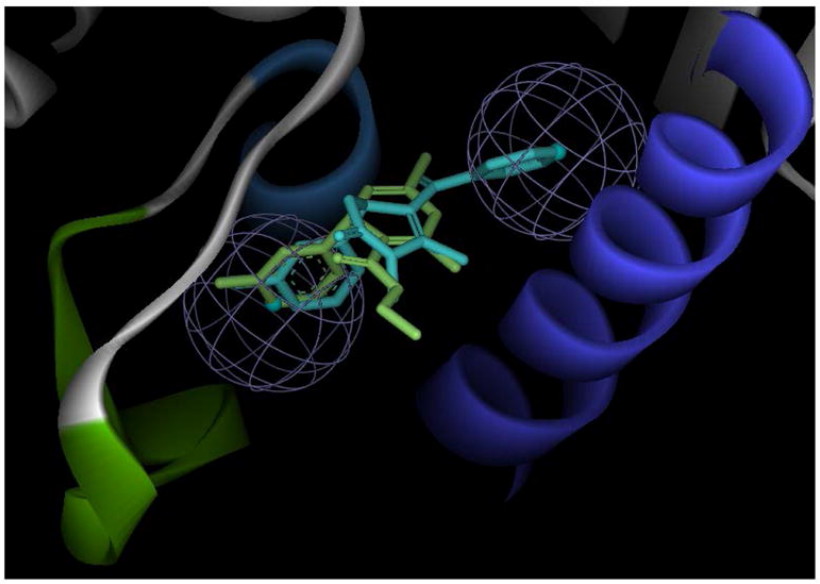

Table 1

Statistics on the hybrids. Hybrids were generated as described. They contain Du atoms and are filtered for 3D duplicates and 2D duplicates. The 2D filtered hybrids have been submitted to MED-Search and the listed sources of molecules. The numbers of found molecules represent the number of molecules in a particular library that matched the hybrids. Wildcard can match nothing or any atom type

Hybrid count (unique 3D)	4115
Hybrid count (unique 2D, Du wildcard)	1417
Scaffold (SC, unique 3D)	903
Scaffold (unique 2D, Du wildcard)	814
Hybrids found in PubChem	3390
Hybrids found in the PDB	2
Hybrids found in the chemical libraries (Chembridge diverset, Maybridge hitfinder)	114
Hybrids found in the a list of inhibitors (BindingDB)	6
SC found in PubChem	293
SC found in the PDB	15
Hybrids found in the chemical libraries (Chembridge diverset, Maybridge hitfinder)	57
Hybrids found in the a list of inhibitors (BindingDB)	2