J Comput Aided Mol Des (2009) 23:571–582
DOI 10.1007/s10822-009-9286-z

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

Ksenia Oguievetskaia · Laetitia Martin-Chanas · Artem Vorotyntsev · Olivia Doppelt-Azeroual · Xavier Brotel · Stewart A. Adcock · Alexandre G. de Brevern · Francois Delfaud · Fabrice Moriaud

Received: 16 February 2009 / Accepted: 19 May 2009 / Published online: 17 June 2009

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 protein data bank (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. Furthermore we have excluded intra-family MED-Portions derived from Eg5 ligands that occupy the hydrophobic pocket and predicted new potential ligands by hybridization that would fill simultaneously both pockets. Some of the latter having original scaffolds and substituents in the hydrophobic pocket are identified in libraries of synthetically accessible molecules by the MED-Search software.

Keywords Fragment-based · Drug design · PDB · Anti-mitotic · Kinesin · Allosteric pocket

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
ROC	Receiver Operating Characteristic

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

Ksenia Oguievetskaia and Laetitia Martin-Chanas contributed equally to this work.

K. Oguievetskaia (✉) · L. Martin-Chanas · A. Vorotyntsev · O. Doppelt-Azeroual · X. Brotel · S. A. Adcock · F. Delfaud · F. Moriaud
MEDIT SA, 2 rue du Belvedere, 91120 Palaiseau, France
e-mail: ksenia@medit.fr

O. Doppelt-Azeroual · A. G. de Brevern
INSERM UMR-S 665, Equipe DSIMB, Dynamique des Structures et Interactions des Macromolécules Biologiques, Institut National de Transfusion Sanguine (INTS), Université Paris Diderot—Paris 7, 6 rue Alexandre Cabanel, 75739 Paris Cedex 15, France

[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 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 (Fig. 1) [11–16]. Most of these inhibitors interact with a hydrophobic (HYD) pocket that is unoccupied by the monastrol (Fig. 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: *(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, probe active-site selectivity within a protein family, detect undesired off-target proteins, retrieve PDB ligand candidate for target hopping, and export all superimposed PDB ligands for post processing. 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 assigned into 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 (Fig. 3; [18, 19]). MED-SuMo is extremely efficient, allowing application to the entire Protein Data Bank (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 PDB is growing with 7,065 structures deposited in 2008 and more than 500 new entries currently released every month. In total, nearly

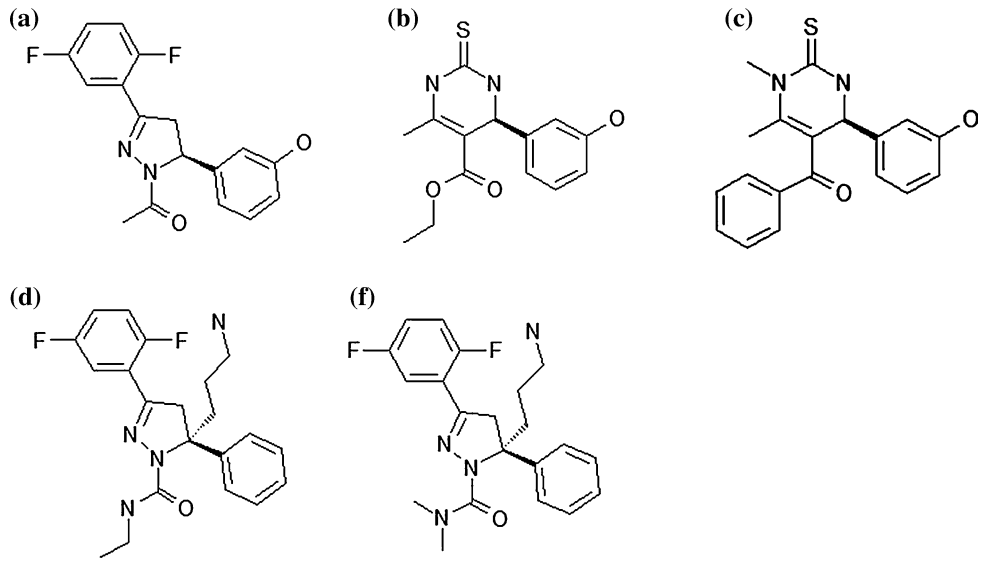

Fig. 1 2D representation of known Eg5 inhibitors. The 3,5-diaryl-4,5-dihydropyrazole (**a**) is a potent and selective inhibitor of Eg5 that is active in cells at low nanomolar concentrations. It was identified in a high throughput screening by Cox et al., Merck [12]. X-ray crystallographic evidence 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**)

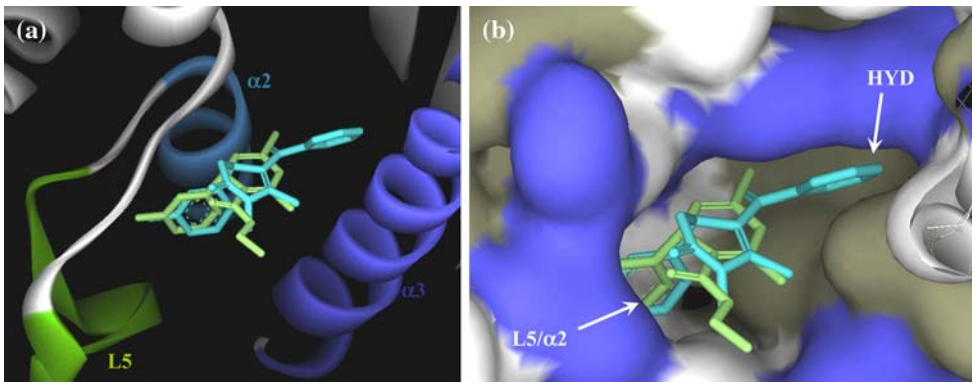

Fig. 2 **a** View of the Eg5 binding site surface from the 1x88 PDB structure showing the co-crystallized monastrol (*green*) and the mon-97 (*clear blue*) from the 2ieh PDB structure, retrieved from the superposition of the two protein structures with MED-SuMo. Both ligands occupy the sub-pocket formed by loop5 (L5) and the helix α2 (α2). The mon-97 occupies as well the hydrophobic pocket HB shown in **b** *Grey*: hydrophobic, *blue*: hydrophilic

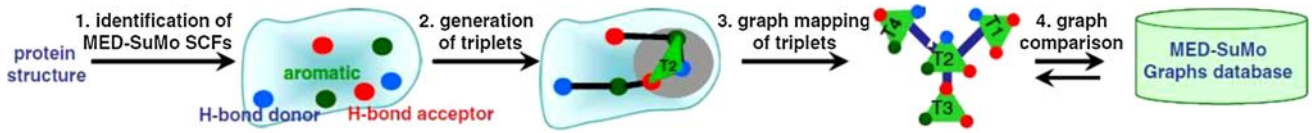

Fig. 3 MED-SuMo core algorithm: graph creation-comparison. 1. The protein molecular structure is scanned and mapped to a dictionary of physicochemical/geometric constructors that detect the MED-SuMo Surface Chemical Features (SCFs, represented by *colored circles*). 2. The geometric rules are applied 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, all matching macromolecular complexes are superimposed on the initial query protein

58,000 macromolecule structures are accessible as of June 2nd, 2009 (http://www.rcsb.org/pdb/statistics).

Recently, we developed a fragment-based drug design approach based on MED-SuMo/MED-SuMo-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 a chemical moiety represented in 3D, annotated with dummy atoms (Du) marking chemical attachment sites, 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 is to populate the HYD pocket of the allosteric binding site of Eg5 to design new potential inhibitors. Our fragment-based protocol is applied to populate the HYD pocket with MED-Portions chemical moieties by local similarity search upon the whole PDB.

The retrieved MED-Portions with MED-SuMo software 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. We applied this protocol on one hand to the MED-Portions that derived from the ligands co-crystallized in the Eg5 structures (intra-family MED-Portions). This allowed us to 'rebuild' the initial ligands, as well as a number of chimeras. On the other hand, in order to target the protocol towards the generation of new potential ligands that would fill the HYD pocket we have excluded intra-family MED-Portions derived from Eg5 ligands that occupy this pocket and applied the protocol to the resulting MED-Portions. This protocol allowed us to generate hybrid molecules that could be new potential ligands. Of these potential ligands, a certain number can be retrieved by MED-Search software from publicly available chemical libraries, such as the PubChem database [23].

Results and discussion

Definition of the custom SCF dictionary

Several water molecules are observed in the allosteric binding site of Eg5, several of them are 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. This is especially the case for superposition of inter-family hits where the structural similarity is local. We therefore decided to use in this study, in addition to the default database, a database built with a custom SCF dictionary.

To build this new database H-bond donors and H-bond acceptors were added 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 [21], only the water molecules in the near neighborhood 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. This increase in the number of SCF in the query signature generally results in hits having a larger number of SCFs and by consequence in a greater MED-SuMo score [24]. In this work, default dipolar SCF for H-bond donors and acceptors are replaced by non oriented objects.

This new SCF dictionary is able to better populate the binding sites or surfaces where water molecules are present with inter-family hits.

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

The goal of the study was to predict new categories of molecules that could fill the HYD pocket without using the knowledge of Eg5 PDB ligands that occupy this pocket. Thus we have chosen to use a structure (entry 1x88 in the PDB) containing an inhibitor that does not fill the HYD pocket.

We constructed the query at 10 Å around the monastrol of the 1x88 structure. The query graph is then submitted to MED-SuMo to be compared to the pre-calculated graphs in the MED-Portions database generated with both dictionaries, the classic one and the new dictionary taking water molecules into account.

During MED-SuMo comparison an automatic calculation step of steric clashes between query protein and retrieved MED-Portions chemical moieties has been added (as described in "Background and methods"). We kept only the MED-Portions chemical moieties that exhibited fewer than ten steric clashes with the query once superimposed inside its binding pocket.

In order to validate the contribution of this customized dictionary database to the study, we wanted to appreciate the ability to retrieve with this modified dictionary database additional pertinent MED-Portions. This was tested in the case of the MED-Portions chemical moieties that are derived from the Eg5 ligands from the PDB. We retrieved 75 Eg5 MED-Portions chemical moieties with our default database and 84 by taking the results from the default and the customized dictionary database out of 99 Eg5 total unique MED-Portions.

The retained MED-Portions are then grouped into two subsets. The first one, called "Eg5 only", contains only the fragments of the allosteric ligands co-crystallized in Eg5. This subset was used to validate the protocol against known ligands and was also used to score the final molecules.

A second subset, called "no Eg5 HYD", was designed to show that the protocol is capable of exploring the HYD pocket with only inter-family MED-Portions. To do so we have eliminated all the MED-Portions that are directly derived from co-crystallized ligands of Eg5 that are targeting this pocket (Fig. 2, see "Background and methods" for detailed description).

In both subsets, the MED-Portions have then been sorted by a new scoring function (described in the "Background and methods" section) taking better into account the variation of the MED-Portions' environment in the query binding site compared to the one in its original binding site. This score penalizes MED-Portions having a chemical environment very different from its original one. In this protocol, a threshold has been set to 1.2. It has been calibrated to retrieve all the intra-family MED-Portions. The MED-Portions having a score greater than 1.2 were kept and were considered as the starting point for the next steps. The "Eg5 only" was constituted from 48 MED-Portions and the subset "no Eg5 HYD" contained 2,812 MED-Portions (unique 2D). As shown in Table 1 the customized dictionary MED-Portion database brought a larger amount of new MED-Portions compared to the ones retrieved with the classic dictionary.

Hybridization and filtering of the MED-Portions into potential ligands

Both subsets of the retained MED-Portions chemical moieties were subjected to five hybridization and filtering steps with MED-Hybridise software [21]. Depending on

Table 1 Origin of the MED-Portions, input of the hybridization protocol

Database name	Number (3D at 2Å)	Number (2D)
Classic database	1,760	709
Customized database	8,640	2,485
Both databases	10,400	2,847

Only the MED-Portions having a score > 1.2 have been kept. The first column of the table corresponds to the number of MED-Portions unique in 3D. The second column to the number of unique MED-Portions in 2D

the method used, the hybridisation of all the MED-Portions together could 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 sub-list of rings populating one or the other pockets, called 'seeds', has been selected from the MED-Portions (see "Background and methods" for detailed description). The first step consisted of the hybridization of these 'seeds' with the whole list of MED-Portions chemical moieties using our Chain Combine algorithm implemented in MED-Hybridise (described in [21]). At each following step the hybrids from the previous step are hybridized to the original pool of retrieved MED-Portions. During each hybridization step, 3D duplicates (i.e., same chemical structure and similar 3D coordinates) are discarded.

In addition we applied a 3D sub-structure filter keeping only the hybrids that explored the HYD with at least one ring and the L5/α2 sub-pocket with an aromatic ring (as described in "Background and methods"). This rigorous filtering resulted in a limited number of hybrids that possessed the required characteristics (Table 2). Even though the number of MED-Portions in the beginning is lower in the case of the "Eg5 only" protocol, we have obtained a superior number of hybrids compared to the "no Eg5 HYD" protocol. This is not surprising given that the MED-Portions of the "Eg5 only" protocol, all coming from Eg5 structures are well fitted for the binding site and match perfectly to our filtering criteria.

Table 2 Statistics on the hybrids

	Eg5 only	No Eg5 HYD
Hybrids count		
Unique 3D RMSD 2Å without terminal Du	14,035	8,676
Unique 2D without terminal Du	8,385	6,986
Murcko Scaffold count unique 3D RMSD 2Å without terminal Du	547	384
Murcko Scaffold count unique 2D without terminal Du	505	376
Exact hybrids found in		
PubChem	559	223
PDB	11	1
Eg5 PDB ligands	11	1
Eg5 inhibitors from BindingDB	30	39
Murcko Scaffolds of the hybrids found in		
PubChem	76	173
PDB	17	13
Eg5 PDB ligands	7	1
Eg5 inhibitors from BindingDB	7	2

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. "Eg5 only": correspond to the results of the protocol run on solely MED-Portions chemical moieties derived from Eg5 ligands from the PDB. "No Eg5 HYD": corresponds to the results of the protocol run on all the retrieved MED-Portions chemical moieties after the exclusion of the Eg5 ligands from the PDB that were filling the HYD sub-pocket. Wildcard can match nothing or any atom type

Hybrids scoring and ranking

This fragment-based drug design protocol is generating hybrids from a set of MED-Portions chemical moieties selected with several criteria (as described above). These hybrids are thus likely (1) to have chemically reasonable structures, since they are generated from chemically accessible molecules, (2) to fit in the binding site since the selected MED-Portions chemical moieties have been selected to have a maximum number of tolerated steric clashes and (3) to potentially have favourable interactions with the protein since they have been co-crystallized with a protein containing locally similar biochemical features [18]. The generated hybrids have been analyzed and scored using an energy minimization step prior to the computation of standard scoring functions.

In order to optimise the geometry of the hybrids for the scoring step, their energy has been minimized in situ. The geometry of the hybrids has to be minimized since, due to the parameters used in the Chain Combine hybridization algorithm, the hybrids can be strained, some can also present internal steric clashes and after k steps of hybridisation the hybrids could have up to $10*k$ steric clashes with the protein. Energy minimized hybrids are filtered (as described in

Table 3 Hybrids filtering by removal of redundant and bad quality hybrids prior to scoring: first and second row concern the set of hybrids used to build the ROC curves; the third row concerns the filtering of PubChem molecules matching hybrid

	Hybrids count		
	Unique 3D	Energy	Steric clashes
Eg5 only	13,181	12,212	12,100
No Eg5 HYD	6,696	5,670	5,638
PubChem	248	248	248

From left to right 3 consecutive filters are applied. Unique 3D corresponds to remaining hybrids after Du atoms and 3D duplicates (RMSD 2Å) removal. Energy corresponds to the hybrids with an in situ energy $\leq$ 100 kcal/mol after minimization. Steric clashes corresponds to the hybrids with $\leq$ 5 steric clashes with the protein (protein and hybrids are protonated). NB: Most of the hybrids have few steric clashes and are fitting in the pocket. The Eg5 PDB ligand 2AZ from the PDB structure 2gm1 has 7 steric clashes and is filtered out (in agreement with the fact that an induced fit is needed to accommodate this ligand in the 1x88 structure used as a query in this protocol)

Table 4 Performance of a set of scoring functions to retrieve known actives from the set of all hybrids

Scoring function	ROC score	
	All hybrids	Non-actives from "No Eg5 HYD"
LigScore1	0.675	0.627
LigScore2	0.686	0.705
PLP1	0.755	0.896
PLP2	0.732	0.864
Jain	0.603	0.686
PMF	0.557	0.658
PMF04	0.707	0.812

As hybridization of known Eg5 PDB ligands is likely to generate positives, we have built the ROC curves by removing from the set of negatives the hybrids of Eg5 ligands. There are 23 2D unique actives for a total of 11,666 for the set containing all hybrids and 4,348 for the second set

"Background and methods" section) and the results are reported in Table 3 and scored using a set of scoring functions.

Hybrids that are exact matches (no Du atoms) of Eg5 PDB ligands (7 hybrids) or of BindingDB/ChemBank (16 hybrids) [25–27] are defined as positives; all other hybrids are treated as negatives even though some hybrids contain parts of Eg5 ligands and thus are a potential source of actives. In the absence of known decoys for this target we admitted that some could have been generated by the protocol. The filtering that has been applied to the MED-Portions and then to the hybrids, implies that the obtained molecules fit in the binding pocket. Therefore the obtained decoys present a challenging test for the scoring functions. This type of decoy generation was named 'modelled' in the article by Anthony Nicholls [28].

We have analysed the ranking of active ligands among all hybrids with a ROC curve score. Among the applied scoring functions we have selected PLP1 [29]. This scoring function was able to retrieve the known actives in the top of the list and had a ROC score of 0.896 (Table 4). Using this scoring function 80% (19/23 'actives') of the defined actives are ranked in the top 16% of all hybrids (710/4,353 hybrids).

Description of the scored hybrids

Protocol application on the "Eg5 only" subset

The fragment-based MED-SuMo approach is capable of retrieving most of the MED-Portions derived from the fifteen different ligands co-crystallized with Eg5 (84 out of 99 total in 2D). After only three steps of our hybridisation protocol, the generated hybrids matched thirteen of the Eg5 ligands when considering the Du atoms contained in the hybrids as wildcard atoms. Eight of them are perfect matches (no Du atoms). The two ligands (coming from the structures with the PDB codes: 2q2y and 2q2z) that are not retrieved by the protocol can be explained by the fragmentation method we used. Indeed all MED-Portions must be derived from known molecules in the PubChem database with a molecular weight less than 250 Daltons. In the case where no match is found with a PubChem molecule, this type of fragmentation results in an incomplete description of some atoms of some PDB ligands. The two ligands that have not been reconstituted have an incomplete description.

While filtering the obtained hybrids for those that fill the HYD and the L5/α2 sub-pockets two ligands that do not fill the HYD sub-pocket are lost as expected (co-crystallized in the structures with PDB codes 1x88, 1q0b and 2fme).

This fragment-based protocol is also able to generate chimeras of Eg5 ligands. These chimeras are of particular interest since they result from the combination of MED-Portions that have chemical features important for the binding in each sub-pocket. At the same time they generate chemical diversity (Fig. 4g, h).

Description of generated hybrids derived from the "no Eg5 HYD" subset

We have selected some of the hybrids that were ranked in the top 16% by the PLP1 scoring function as described above. They all result from the hybridization of inter-family MED-Portions chemical moieties inside of the HYD sub-pocket deriving from various PDB ligands (Fig. 4a–d). 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 and Arg119 [17]. All those residues are also involved in the predicted binding of the selected hybrids (Fig. 4). Given the restriction during the hybridisation protocol to keep hybrids targeting the HYD pocket, most of the hybrids are interacting with the hydrophobic residues, Leu160, Ile136 and several are forming T-stacking interactions with Phe239 (Fig. 4). Some hybrids are also predicted to be involved in H-bond interactions with the protein directly or through water molecules (Fig. 4a, c).

The hybrids selected at that step are hit-like and can be considered as original molecules compared to the known PDB ligands.

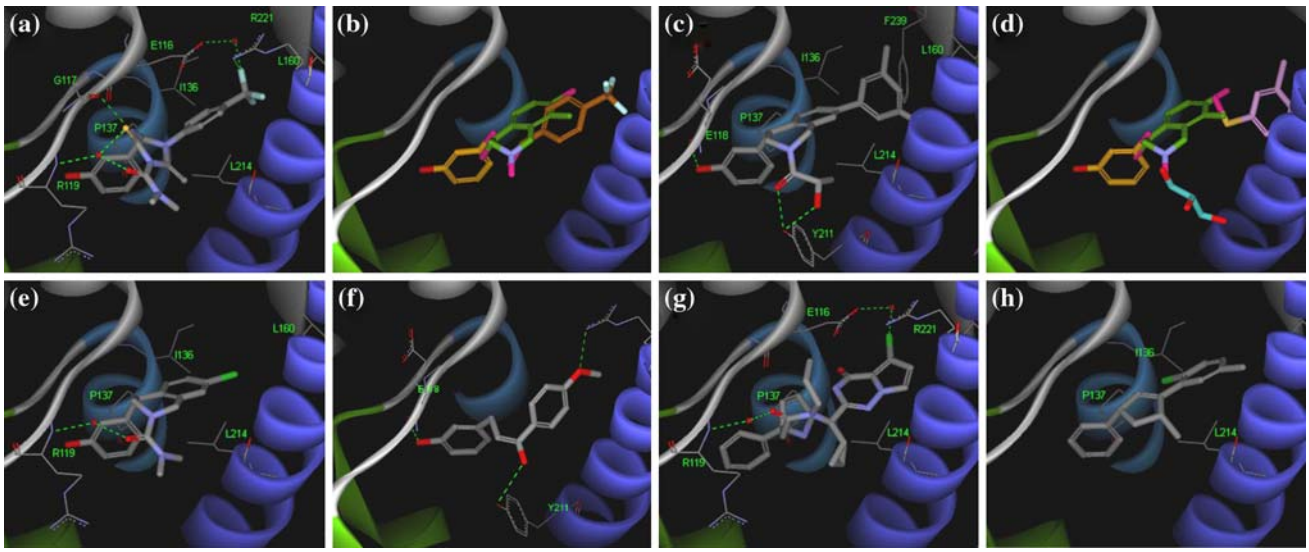

Fig. 4 Examples of the selected hybrids and the MED-Portions that have been used for their formation presented in 3D coordinates, inside the 1x88 allosteric binding site. In the L5/α2 pocket several conserved residues are involved in the inhibitor's binding in all the crystal structures of Eg5 from the PDB: Glu116, Glu118, Arg119, Pro137 and Leu214. All those residues are also involved in the predicted binding of the selected hybrids. **a** and **c** Two hybrids coming from the "no Eg5 HYD" protocol are shown inside the binding pocket of Eg5 (in stick, protein color code as in Fig. 2). The residues of the protein interacting with the obtained molecules are represented with their side chains in 'line' rendering and labeled. The shown water molecules (*red crosses*) are predicted to form H-bond between the protein and the hybrid. **b** and **d** The MED-Portions that have been used to form the hybrids represented in (**a** and **c**) respectively. Du atoms are shown in *pink*. The different MED-Portions chemical moieties derive from the following structures: **b** 1pxl (*brown*), 2fme (*yellow* and *green*), 1leg (not represented for the sake of clarity); **d** 1ep4 (*pink*), 1wvf (*blue*), 2fme (*green*), 1x88 (*yellow*). The MED-Portions that are filling the hydrophobic pocket are derived from solely inter-family MED-Portions. **e** and **f** Two examples of hybrids coming from the "no Eg5 HYD" protocol that have matched molecules in PubChem with the MED-Search module. The hybrid presented in **e** derives from 2fme MED-Portions. The hybrid presented in **f** groups MED-Portions derived from the following structures 1s0j, 1ppj and 1x88. **g** and **h** two hybrids obtained with the "Eg5 only" protocol **g**: chimera of 2gm1 and 3cjo; **h**: chimera of 1yrs, 2ieh, 2q2z and 2fky). They illustrate the diversity that can be obtained by our fragment-based approach with only 15 ligands as starting material (Eg5 PDB ligands). 1pxl: human cyclin dependent kinase 2; 1leg: crystal structure of h-2 kb bound to the dev8 peptide; 1ep4: crystal structure of hiv-1 reverse transcriptase; 1wvf: *p*-cresol methylhydroxylase; 1ppj: mitochondrial bc(1) complex; 1s0j: trypanosoma cruzi trans-sialidase

Comparison of the obtained hybrids with compound libraries using 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 by the fragment-based protocol were searched in several chemical libraries including known kinesin inhibitors and the entire PubChem database [23]. A new module called MED-Search has been developed for that purpose. As hybrids may 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 here we propose 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 quantity of comparisons it was crucial to efficiently pre-filter all possible matches. For that purpose, the SCINS representation of the molecules, described in the "Background and methods" section, was compared [30, 31]. Only the retained matches are submitted to an exact comparison by superposing the molecular graphs [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. The molecules coming out of MED-Search do not have Du atoms, and are known, chemically accessible molecules. Their 3D coordinates are kept, thus they are still aligned in 3D in the reference frame of the query binding site.

We have also generated and searched the scaffolds of the hybrids and the molecules of the chemical libraries using the Murcko rules described by Bemis and Murcko [32]. For the generation of the scaffolds exocyclic double bonds were kept. The results of the comparison are listed in Table 2.

From the 2,355 hybrids matching molecules in PubChem and 2,107 matching known Eg5 ligands in BindingDB/ChemBank, 2,124 and 1,996, respectively, are derived from MED-Portions retrieved from the customized database. This result indicates the utility of adding water molecules to the MED-SuMo graphs for the comparison.

We analyzed the matches in BindingDB/ChemBank using the previously selected scoring function (PLP1). Among 42 matches, 32 are scored in the range of the 16% best scored hybrids among which 80% known actives and

are potential ligands are found. The matching hybrids are analogs of the Eg5 PDB ligands. Their common scaffolds have a position very similar to the one in the co-crystallized PDB structure.

Among hybrids that have been found exactly in the PubChem library several are predicted to bind both in the HYD and in the L5/α2 pockets (Figs. 4e, f, 5). The hybrid presented in (e) derives from three 2fme MED-Portions. The hybrid presented in (f) possesses two MED-Portions derived from inter-family hits.

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 α2 (L5/α2), some occupy as well the hydrophobic pocket (HYD) (Fig. 2). The inhibitors that are binding to both pockets simultaneously have shown sometimes a ten fold higher affinity than those occupying only the L5/α2 pocket.

The goal of this study is to populate the HYD pocket of the allosteric binding site of Eg5 to design new potential inhibitors with our MED-SuMo/MED-SuMo-Fragmentor/MED-Hybridise fragment-based protocol. The MED-Portions retrieved with MED-SuMo software 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.

By considering MED-Portions that derived from the co-crystallized ligands in the Eg5 structures (intra-family MED-Portions) we were able to rebuild the initial ligands and chimeras. By excluding intra-family MED-Portions derived from ligands that occupy the HYD pocket, we were able to reproduce a situation where no known ligand is binding in the hydrophobic sub-pocket. However our protocol was able to predict that this sub-pocked could be filled with MED-Portions and produced new potential ligands that would fill simultaneously both sub-pockets. 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.

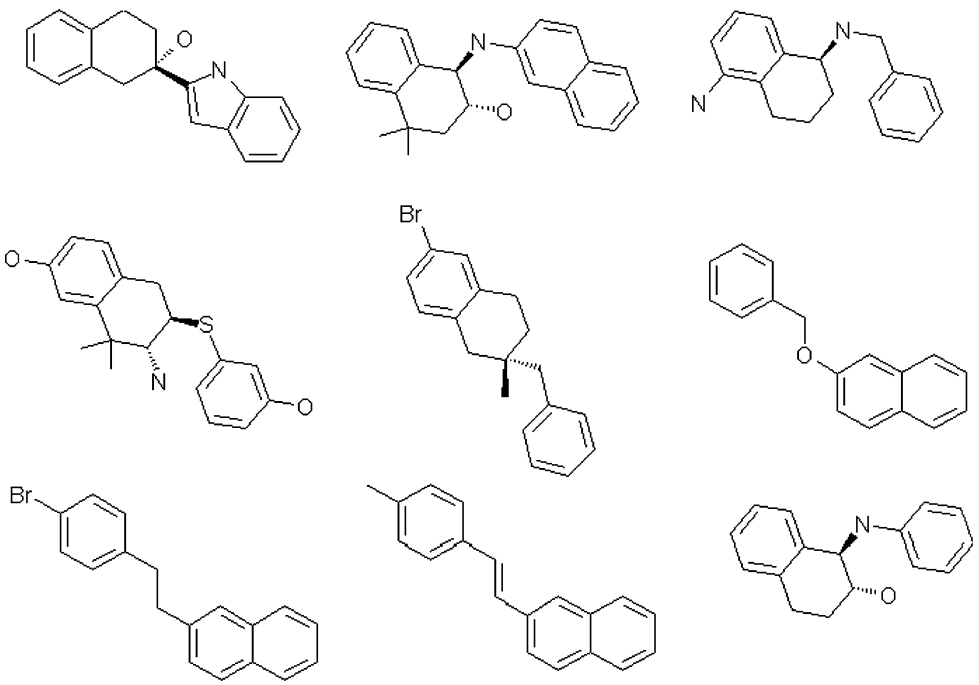

Fig. 5 Some examples of hybrids found in PubChem with MED-Search

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, 24] 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, thus each residue is described 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 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 neighborhood, 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. Once the comparison is made, hits are superimposed on protein query by minimizing distances between each pair of matching SCFs and the similarity is given by the MED-SuMo score which is calculated by taking into account the number of common SCFs with a weight depending on the nature of the SCF and the local shape similarity [18, 19, 24].

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, excluding 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 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-Portion 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-SuMo-Fragmentor [21]. MED-Portions are defined by several criteria: (1) a chemical moiety where atoms are topologically matching a molecule of less than 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 databases

The hydrophobic sub-pocket targeted by this protocol is filled with water molecules. The classical MED-SuMo dictionary does not consider them as part of the ligand environment. However they could be of great importance to retrieve inter-family hits that share only local similarities

in this sub-pocket involving water molecules. Therefore, in addition to the MED-Portion database generated using the classical MED-SuMo dictionary we generated another database based on a modified dictionary that consider the water molecules and accept their match with either water molecules, or H-bond donors and acceptors on the hit protein structure. Indeed, MED-SuMo software allows the advanced user to modify the dictionary of SCF definitions by editing a text file.

In the default 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 directly any orientation or projection for the H-bond donor or acceptor groups. Therefore we added punctual geometrical variant for 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. The SCFs too buried to interact with a potential ligand are discarded.

Selection of collected MED-Portions

The collected MED-Portions are split into two subsets. The first one, called "Eg5 only", is designed to validate the protocol on MED-Portions chemical moieties derived from the inhibitors co-crystallized with Eg5 (intra-family). It contains the MED-Portions derived from the 15 different Eg5 allosteric ligands present in the PDB. The second subset, called "no Eg5 HYD", contains all the MED-Portions collected from the MED-SuMo run from which have been 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 (1x88, 2fme and 1q0b) were kept in the second subset, since their ligands are not positioned in the HYD pocket. For each subset, the selected MED-Portions' atoms were typed with OpenBabel [33].

The quality of the superposition of each hit in a MED-SuMo run is evaluated with MED-SuMo score (described in the MED-SuMo technology section of "Background and methods") and/or terms on penalty, RMSD and fitting volume. Nevertheless this score is relative to the number of SCF pairs and then to the size of the MED-Portion (the larger the MED-Portion is the more SCF will be in its environment). As a consequence a small MED-Portion at the surface of a protein perfectly aligned in the query (same environment) scored after a larger MED-Portion positioned in a tight binding site even if the local SCF environment is very different between the query binding site and the original hit binding site.

To better take into account of the environment similarity of the MED-Portion in the new environment of the query binding site compared to the environment of the original binding site, we changed the default MED-SuMo score. The new "mediated MED-SuMo score" (mScore) corresponds to the MED-SuMo score of the retrieved MED-Portion ($Score_{hit}$) with the ratio between $Score_{hit}$ and the MED-SuMo score this MED-Portion would have in its original structure ($Score_{original}$): $mScore = Score_{hit} * (Score_{hit}/Score_{original})$. The mediated MED-SuMo score penalizes the MED-Portions having lost key structural features in their protein environment.

Selection of seeds and substructures for filtering

To create new ligands that would bind simultaneously the targeted HYD and the L5/α2 pocket, we selected a sub-list of the retrieved MED-Portions chemical moieties that explored these pockets as the starting points, the seeds. The seeds were selected as follows.

The retrieved MED-Portions were filtered 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. All Eg5 ligands co-crystallized in the PDB have an aromatic ring positioned in the L5/α2 pocket. The seeds positioned in the L5/α2 pocket are filtered using a hydrophobic and aromatic pharmacophoric feature with Discovery Studio 2.0 [34]. This pharmacophoric feature is positioned at the center of mass of the aromatic cycle of monastrol in 1x88 and has a radius of 3 Å. All selected seeds were decorated with Du atoms at every position consistent with the attached atoms' hybridization states and bonds, with arbitrary rotamers selected for terminal atoms.

In order to keep moieties of the selected seeds, we applied during the hybridization steps filtering with substructures that have been designed as follows. We deleted the Du atoms and fragmented the selected seeds into 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.

Hybridization parameters

We applied five 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 and to perform a crossover operation at such bonds. 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 molecules obtained in the previous step to the selected MED-Portions. During steps 2–5 we used the substructures without Du atoms for the 3D filtering, by keeping only the hybrids that were superstructures of the one of those substructures. After each step we filtered the generated hybrids for 3D duplicates (RMSD 2 Å). The resultant hybrid molecules have been submitted to the MED-Search module as explained later.

Scoring of the hybrids

The hybrids obtained from the hybridization protocol contain Du atoms. These are useful during the hybridization steps when the Chain Combine algorithm is applied. For the relaxation of the hybrids inside the protein binding site and for the scoring steps the Du are discarded because they are not considered as wildcard atoms by third party software. The hybrids are processed as follows: (1) terminal Du atoms are deleted, (2) remaining Du atoms are converted to carbon atoms, (3) explicit hydrogens are added and (4) 3D duplicates (matching molecules with RMSD < 2Å) are removed, by keeping only the one with the smallest energy, which is computed using an in-house implementation of MMFF94s [35].

The hybrids are scored in Discovery Studio 2.0 [34] as follows: (1) in situ hybrids minimization is achieved with the CHARMm forcefield [36] implemented in the "Ligand Minimization" protocol using the "Smart Minimizer" with default parameters. The protein structure is that of the query protein (PDB code 1x88) used for the MED-SuMo run after removing all ligands and water molecules. During the minimization the receptor is considered rigid. (2) We removed the hybrids with a very high final in situ energy (>100 kcal/mol), originating from strained conformations that can't be solved by energy minimization (3) We removed hybrids with more than five residual bumps because these are not expected to be resolved through flexibility in the protein side chains (4) all hybrids are scored using the available scoring functions: PLP1, PLP2, PMF, PMF04, Jain, LigScore1, LigScore2 [29, 37–41]; (5) computation of ROC curves and of the area under the ROC curve (ROC score) using the "Analyze Ligand Poses" protocol.

ROC curves for a given scoring function are built using all the scored hybrids as follows: positives are defined as the exact matches of the PDB Eg5 ligands (7 matches) and in BindingDB/ChemBank (16 matches) [25–27]. All the others are considered as negatives. In case of 2D duplicates, only the hybrid with the best score is kept. The ROC score is reported as a number between 0 and 1. The closer is the ROC score to 1 the better the model is at distinguishing known actives from all the hybrids.

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 encodes molecules' scaffold structures. The design of this mini-fingerprint was inspired by the Novartis SCINS classifier [30, 31]. 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 (using Sybyl-style atom types [42], ignoring hydrogens), whether they are in a ring and whether they are 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.

Acknowledgments 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 was 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".

References

1. Mitchison TJ, Salmon ED (2001) Nat Cell Biol 3:E17. doi:10.1038/35050656
2. Mitchison T, Kirschner M (1984) Nature 312:237. doi:10.1038/312237a0
3. Mitchison TJ (1989) J Cell Biol 109:637. doi:10.1083/jcb.109.2.637
4. Vale RD, Fletterick RJ (1997) Annu Rev Cell Dev Biol 13:745. doi:10.1146/annurev.cellbio.13.1.745
5. Amos LA, Cross RA (1997) Curr Opin Struct Biol 7:239. doi:10.1016/S0959-440X(97)80032-2
6. Wood KW, Cornwell WD, Jackson JR (2001) Curr Opin Pharmacol 1:370. doi:10.1016/S1471-4892(01)00064-9
7. Mayer TU, Kapoor TM, Haggarty SJ, King RW, Schreiber SL, Mitchison TJ (1999) Science 286:971. doi:10.1126/science.286.5441.971
8. Kwok BH, Kapoor TM (2007) Curr Opin Cell Biol 19:36. doi:10.1016/j.ceb.2006.12.003
9. Turner J, Anderson R, Guo J, Beraud C, Fletterick R, Sakowicz R (2001) J Biol Chem 276:25496. doi:10.1074/jbc.M100395200
10. Maliga Z, Kapoor TM, Mitchison TJ (2002) Chem Biol 9:989. doi:10.1016/S1074-5521(02)00212-0
11. Yan Y, Sardana V, Xu B, Homnick C, Halczenko W, Buser CA, Schaber M, Hartman GD, Huber HE, Kuo LC (2004) J Mol Biol 335:547. doi:10.1016/j.jmb.2003.10.074
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) Bioorg Med Chem Lett 15:2041. doi:10.1016/j.bmcl.2005.02.055
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) Bioorg Med Chem Lett 16:3175. doi:10.1016/j.bmcl.2006.03.040
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) Bioorg Med Chem Lett 16:6049. doi:10.1016/j.bmcl.2006.08.118
15. Tarby CM, Kaltenbach RF3, 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) Bioorg Med Chem Lett 16:2095. doi:10.1016/j.bmcl.2006.01.056
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) Bioorg Med Chem Lett 16:3937. doi:10.1016/j.bmcl.2006.05.037
17. Garcia-Saez I, DeBonis S, Lopez R, Trucco F, Rousseau B, Thuery P, Kozielski F (2007) J Biol Chem 282:9740. doi:10.1074/jbc.M608883200
18. Jambon M, Imberty A, Deleage G, Geourjon C (2003) Proteins 52:137. doi:10.1002/prot.10339
19. Jambon M, Andrieu O, Combet C, Deleage G, Delfaud F, Geourjon C (2005) Bioinformatics 21:3929. doi:10.1093/bioinformatics/bti645
20. Doppelt O, Moriaud F, Bornot A, de Brevern AG (2007) Bioinformation 1:357
21. Moriaud F, Doppelt-Azeroual O, Martin L et al. (2009) J Chem Inf Model
22. Doppelt-Azeroual O, Moriaud F, Delfaud F (2009) Infect Disord Drug Targets
23. The PubChem Project (2008) http://pubchem.ncbi.nlm.nih.gov. Accessed 9 June 2008
24. Jambon M (2003) A bioinformatic system for searching functional similarities in 3D structures of proteins. Université Claude Bernard Lyon 1, Lyon
25. Chen X, Lin Y, Liu M, Gilson MK (2002) Bioinformatics 18:130. doi:10.1093/bioinformatics/18.1.130
26. http://chembank.broad.harvard.edu/
27. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) Nucleic Acids Res 35:D198. doi:10.1093/nar/gkl999
28. Nicholls A (2008) J Comput Aided Mol Des 22:239–255
29. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST (1995) Chem Biol 2:317. doi:10.1016/1074-5521(95)90050-0
30. Czerminski R (2005) Conference presentation, http://www.eyesopen.com/about/events/cup6/
31. Jenkins JL, Glick M, Davies JW (2004) J Med Chem 47:6144. doi:10.1021/jm049654z
32. Bemis GW, Murcko MA (1996) J Med Chem 39:2887. doi:10.1021/jm9602928
33. The Open Babel Package (2008) http://openbabel.sourceforge.net/. Accessed 5 July 2008
34. Accelrys Software Inc.: 10188 Telesis Court, Suite 100 San Diego, CA 92121, USA San Diego
35. Halgren TA (1999) J Comput Chem 20:720. doi:10.1002/(SICI)1096-987X(199905)20:7<720::AID-JCC7>3.0.CO;2-X
36. Brooks BR, Bruccoleri RE, Olafson BD et al. (1983) J Comp Chem 187
37. Gehlhaar DK, Bouzida D, Rejto PA (1992) American Chemical Society: Washington DC 292
38. Jain AN (1996) J Comput Aided Mol Des 10:427. doi:10.1007/BF00124474
39. Krammer A, Kirchhoff PD, Jiang X, Venkatachalam CM, Waldman M (2005) J Mol Graph Model 23:395. doi:10.1016/j.jmgm.2004.11.007
40. Muegge I (2006) J Med Chem 49:5895. doi:10.1021/jm050038s
41. Muegge I, Martin YC (1999) J Med Chem 42:791. doi:10.1021/jm980536j
42. http://tripos.com/mol2/atom_types.html