MED-SuMo and MED-Hybridise: exploit all PDB macromolecule structures to design/optimize innovative leads

Moriaud F., Oguievetskaia K., Adcock S.A., Vorotynsev A.M., Martin L., Doppelt-Azeroual O. and Delfaud F.

MEDIT SA, 2 rue du Belvédère, 91120 Palaiseau France

Abstract

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

In this work, we've used **MED-SuMo** and **MED-Hybridise** [2] to query and mine the protein's Surface Chemical Functional groups (SCF) surrounding fragments of PDB ligands, seeking similarities with the kinase and GPCR of interest. MED-portions (augmented fragments) were used to design novel scaffolds (lead generation) or to optimize decoration of a given scaffold (lead optimization) for the GPCR and the kinase respectively.

Here we present the lead generation and lead optimization results after 3D hybridisation in 5 iterations suggested by our **MED-Hybridise** protocol. In this list, we've analysed the scaffolds in regard to their diversity and their presence in the PubChem, in the PDB and some other libraries.
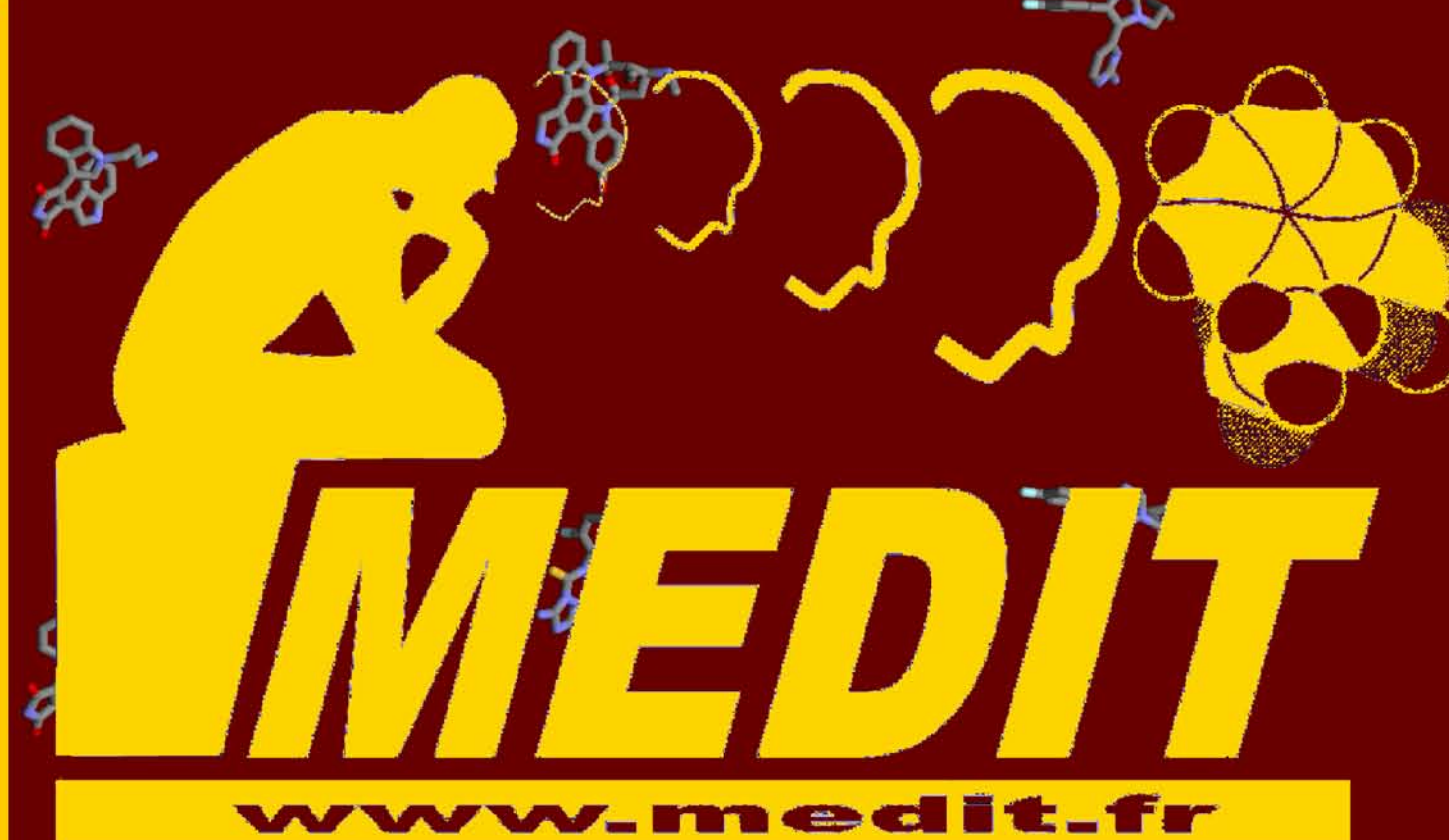
MED-SuMo / MED-Hybridise

Overview: MED-SuMo, innovative target-based drug design software, makes possible to compare any interaction surface against the whole Protein Data Bank in a few minutes. Independent from the notion of protein sequence or fold, **MED-SuMo detects and compares biochemical features on protein.** In order to rapidly browse hundred of hits, the **MED-SuMo** scoring function takes into account chemical features and shape overlaps.

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

Some MED-SuMo SCF:

- aromatic
- **H-bond donor**
- **H-bond acceptor**
- **H-bond donor**
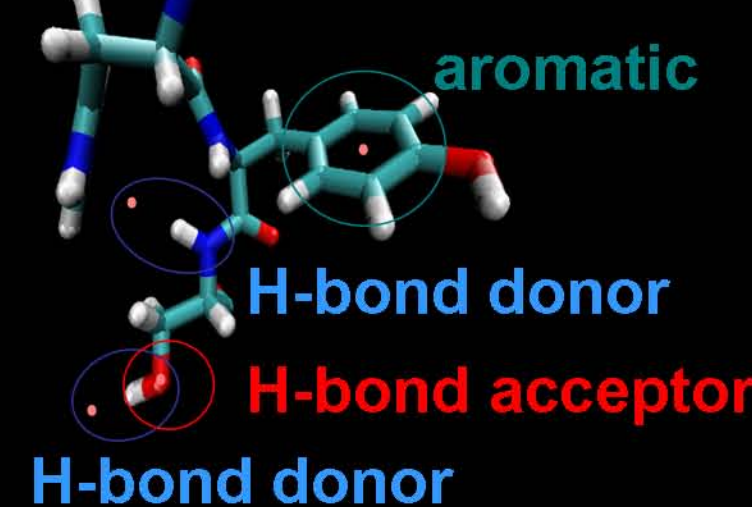
MED-SuMo SCF color code :

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

MED-Fragmentor
protein-MED-portion DB generation:

1 protein - ligand structure from the PDB

N protein – MED-portion structures
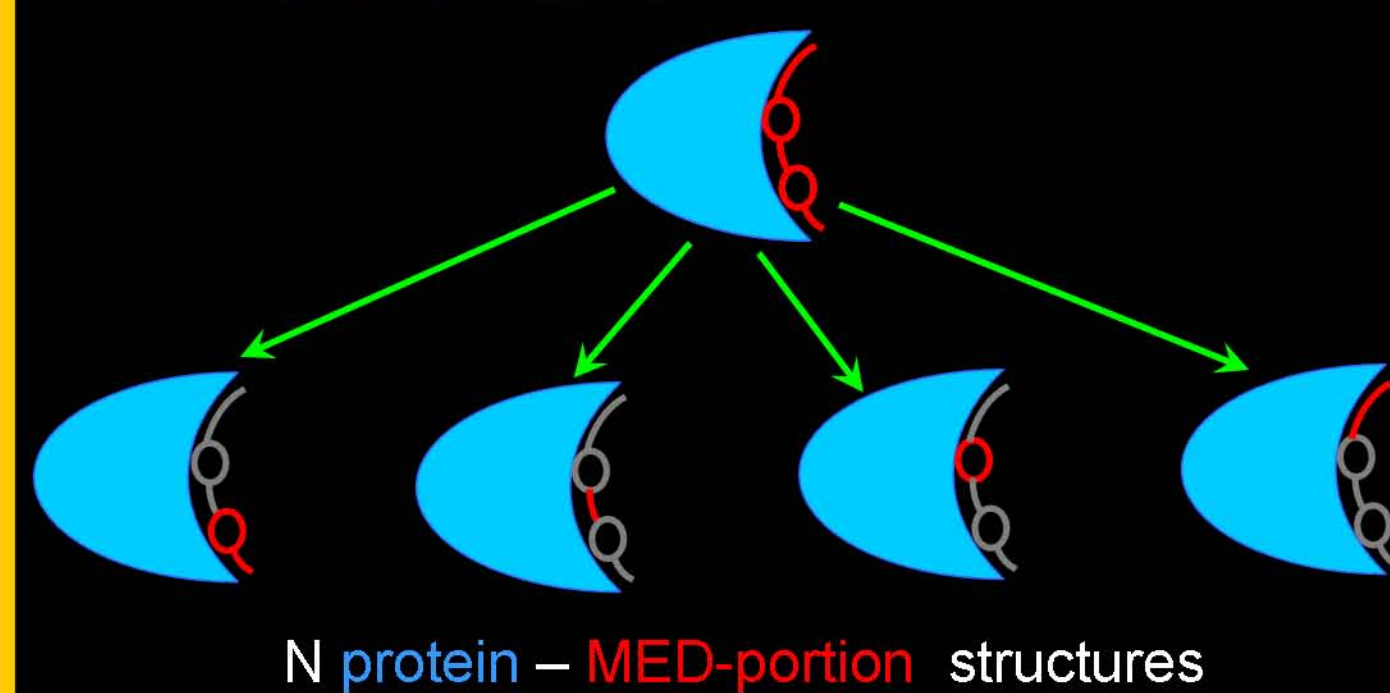
Advantages:
- The whole PDB is browsed
- Original 3D coordinates
- Protein flexibility is taken into account
- Interfamily hits
- One or more 3D local alignments

Lead optimization with MED-SuMo / MED-Hybridise protocol: application to the DFG-out VEGFR-2 kinase

We've kept the phenylamide moiety of its GIG ligand to decorate (lead optimization). The phenylamide is in the same atomic 3D coordinates as it is in 2oh4 (*Fig.1.A*). It is an interesting case study as inhibitors targeting the DFG-out conformation may exhibit a higher selectivity and less competition with ATP, as exemplified by the drug imatinib.

We run the **MED-SuMo** protocol using the protein-MED-portion database, followed by 5 hybridisation steps with the **MED-Hybridise**. In this study we allowed 2 possible positions for the 'decoration'.

We've obtained a set of 1129 hits positioned in sub-parts of the VEGFR-2 binding site (*Fig.1.B-D*). We've characterized the hybrids at each step. The hybrids are growing at each step, lead-likeness (rule of three) is decreasing quickly as the optimization proceeds and the drug-likeness (rule of 5) is higher than 60% within the hybrids (*Fig.2*).

To assess the diversity and potential affinity to protein kinases of the hybrids, we've analyzed the results in terms of unique scaffolds. We've found 3585 unique scaffold: 50 are found in the PDB and 298 in PubChem, 25 are found in the active molecules of the protein kinases biossays (*Tab.1*).
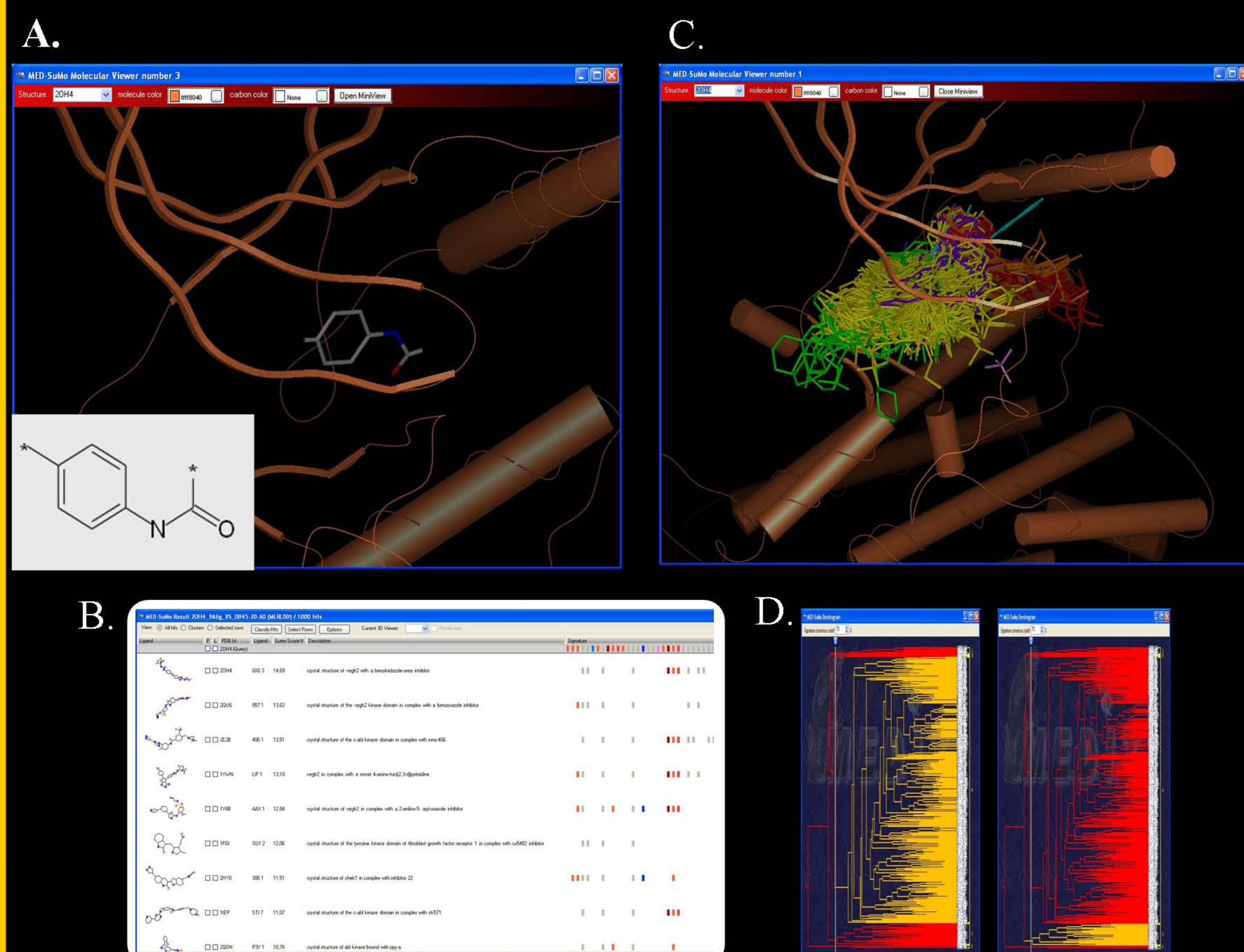

Fig. 1. A: 2oh4 (orange) with the phenylamide substructure of GIG. Attachment points for hybridisation are marked with a star. B. MED-SuMo result table showing the pdb hits: ligand depiction and their environment with a signature. C. MED-portions associated to pdb hits are shown in 3D in the reference frame of 2oh4. Coloring is done according to the MED-portions local environment clustering. D. Dendrogram represents the classification of the MED-portions according to their environment (SCF signature). Cluster 4/5 contain protein kinases with a correct fold alignment; 309/25 MED-portions aligned in the hinge region/ the allosteric pocket respectively.
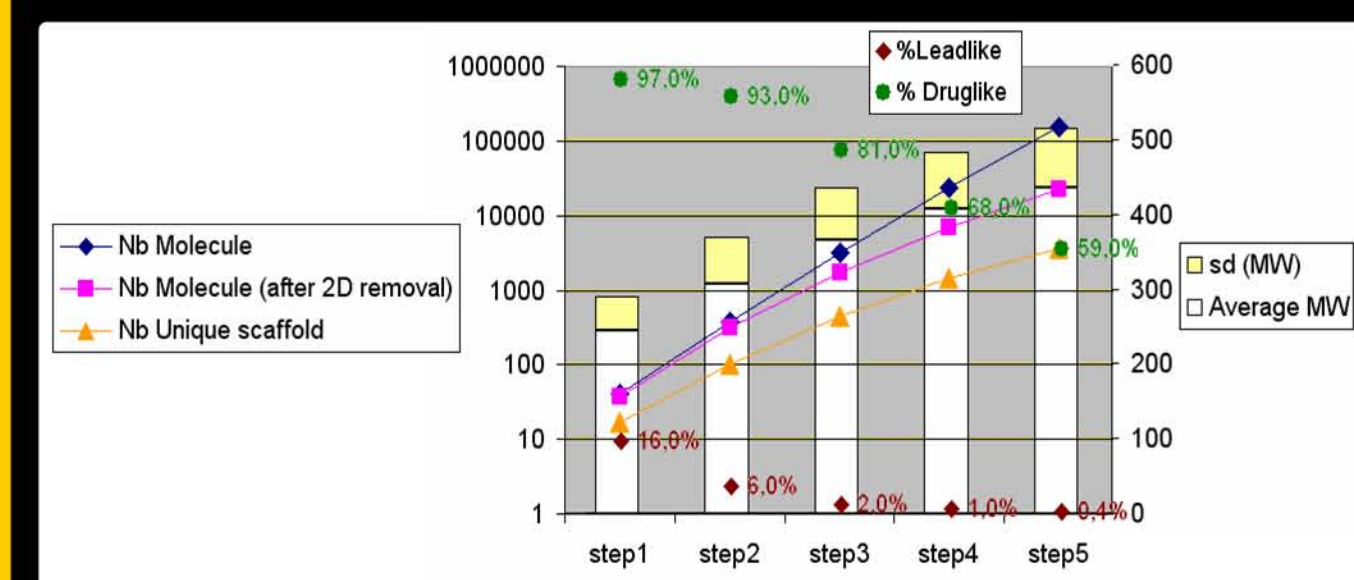

Fig. 2: Characterization of the hybrids in terms of molecular weight, diversity and drug-likeness.

PDB scaffolds	PDB scaffolds protein kinase	PubChem	PubChem kinases bioassays
97 (4977 others)	50	298	55

Tab. 1: Unique scaffolds found in PDB, PubChem and in protein kinases bioassay.

Lead Discovery with MED-SuMo / MED-Hybridise protocol: application to the GPCRs

G protein-coupled receptors (GPCRs) are probably the most important superfamily for drug design. The first human structure obtained was that of the β2-adrenergic receptor with corazolol, only a year ago [5]. We are applying here our **MED-SuMo / MED-Hybridise** protocol to this structure (*Fig. 3.A*).

We ran the **MED-SuMo** protocol using the protein-MED-portion database, then we selected the hits by their **MED-SuMo** score and the number of clashes (*Fig.3.B*). The obtained MED-portions were quite diverse and their 3D distribution resembled the one obtained by HTS by *Topiol et al. (Fig.3.C)* [5], some of them are depicted individually in *Fig. 3.D*.

The selected MED-portions were hybridised in 5 steps. We obtained 1480 unique scaffolds (17596 hybridized molecules). We compared these scaffolds to the PDB and to the PubChem (*Tab.2*). The scaffold that matched PubChem molecules are still diverse in terms of chemistry. Furthermore those molecules are exploring a large surface area of binding site pocket (*Fig.4.A*). Some of the final scaffolds are depicted in *Fig.4.B*.
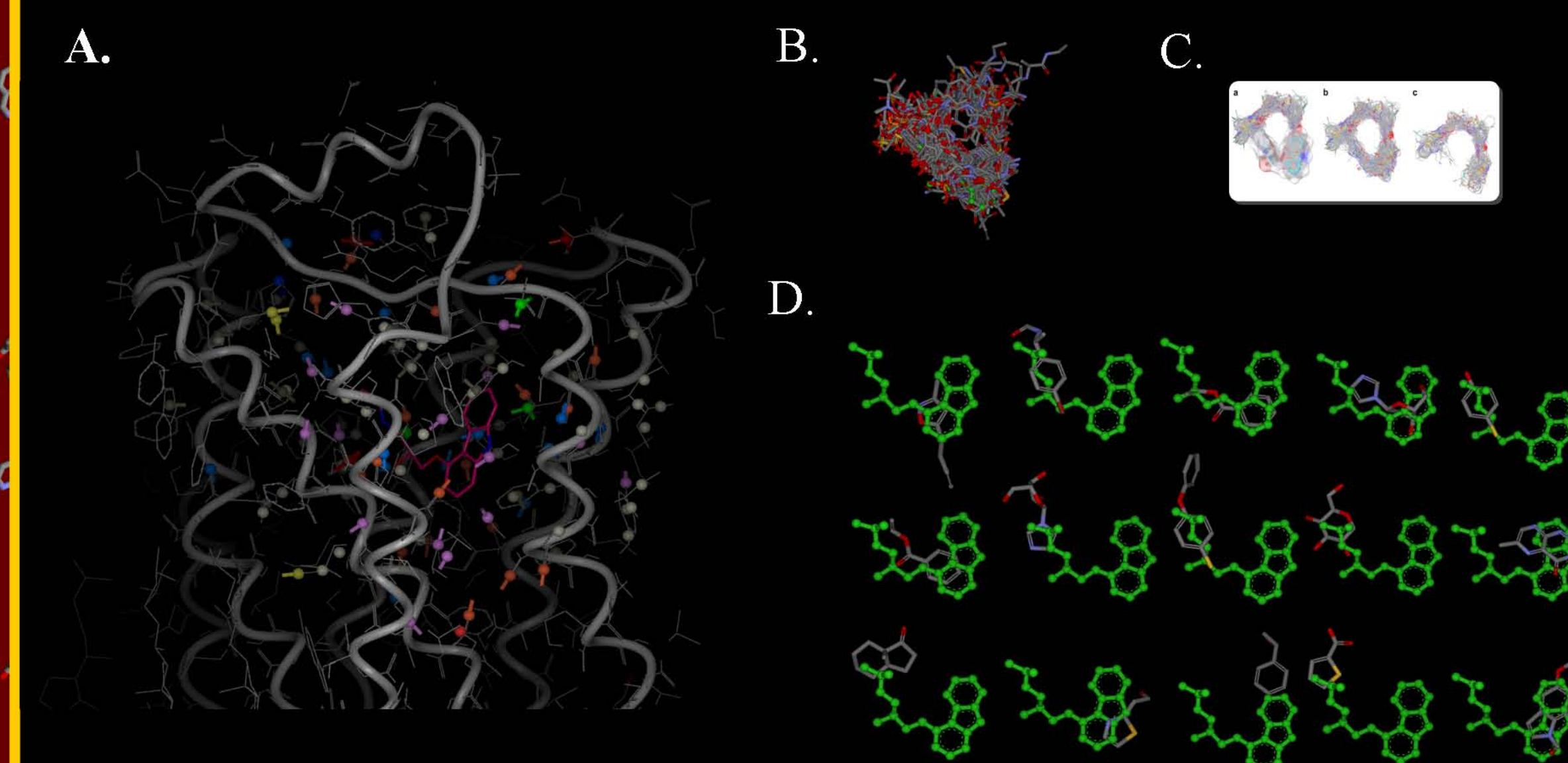

Fig.3. A. Query selection. 2rh1 is show in grey, the initial ligand is depicted. The retained SCF are shown. Superimposition of the hits selected by their MED-SuMo score and by the number of clashes (max. 5 clashes; 274 MED-portions). C. Results obtained by Topiol et al. in 2008 with a HTD approach. [5] D. Some of the MED-portions in their 3D coordinates.

PDB scaffolds	PubChem	PubChem GPCR	ChemLib
22	61	6	18

Tab.2: Unique scaffolds found in PDB, PubChem, molecules actives on the GPCRs from PubChem and in the GPCR chemical libraries.

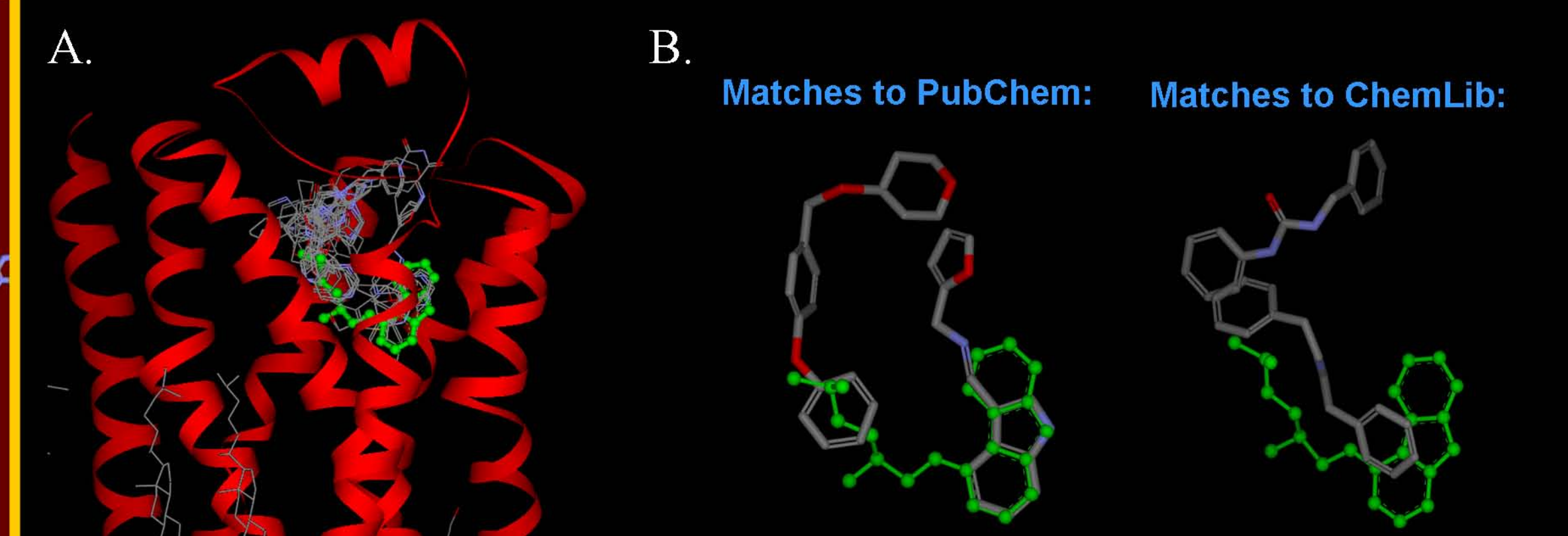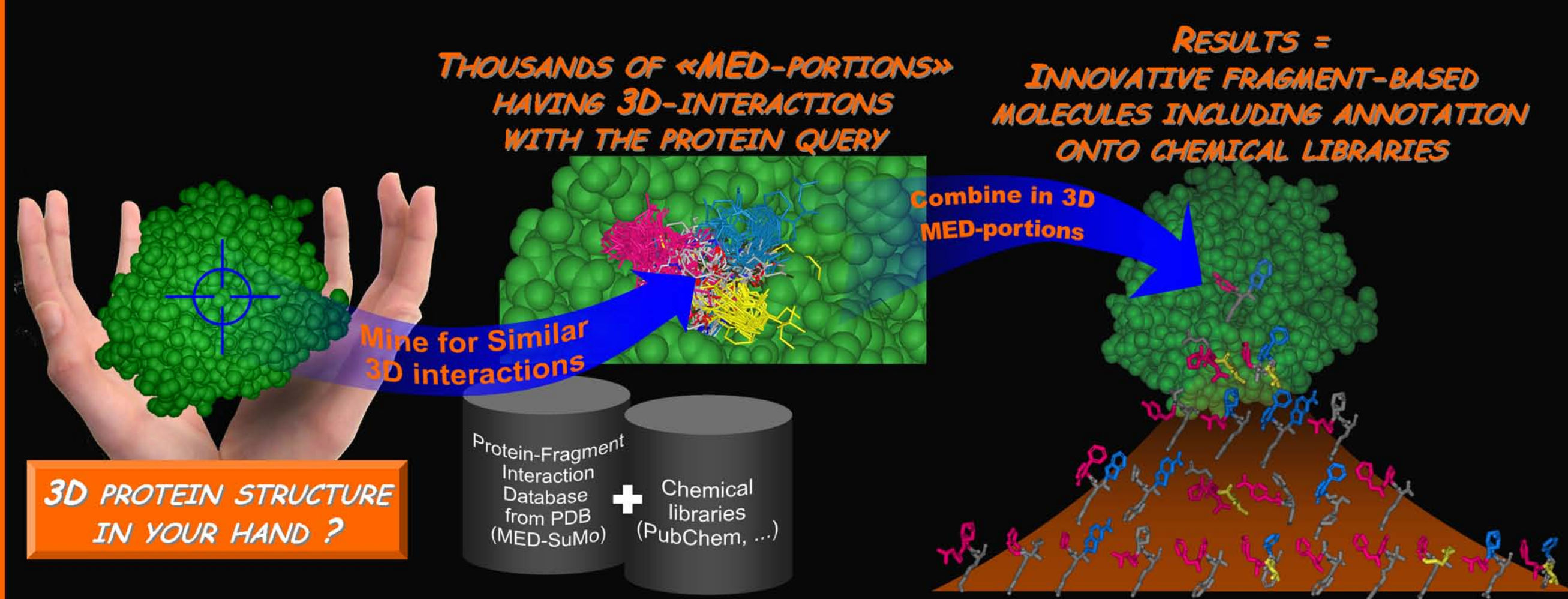**Matches to PubChem:**   **Matches to ChemLib:**

Fig.4.A. Obtained scaffolds that match PubChem. Molecules in their 3D coordinates. B. Some of the scaffolds that match different libraries in the 3D coordinates.
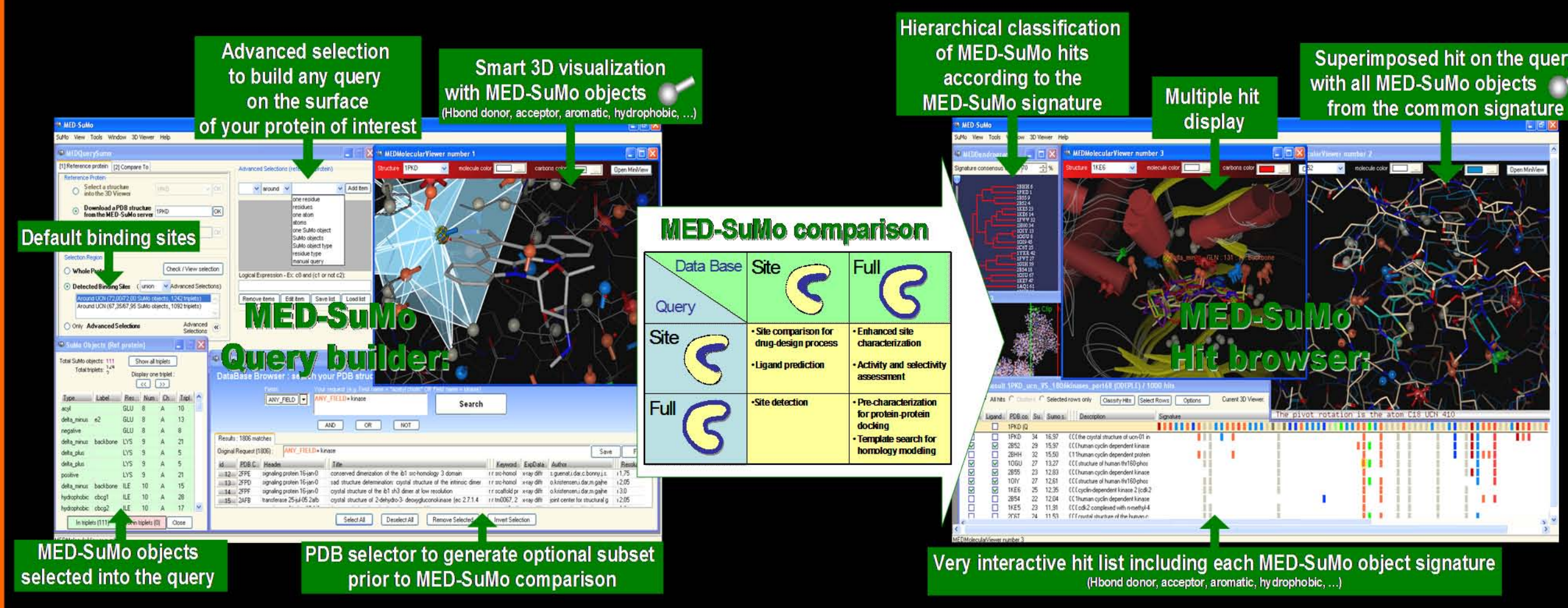
Conclusion

Based on the detection of similar 3D interactions between proteins and MED-portions from the PDB, our **MED-Hybridise** protocol is able to generate valuable hybrids for a lead discovery or lead optimization approach. Hybrids are diverse: 1% (50/3585 for the kinases and 22/1480 for the GPCRs) of the generated scaffolds are found within the PDB. In the case of the kinases, 50 scaffolds are found within the active molecules from the protein kinase PubChem bioassays, 3 of them are new compared to the PDB. Even for a difficult target like the GPCR we were able to find scaffolds that match other known ligands then the corazolol.

Such a protocol can be applied with success on either a binding pocket or a whole protein surface. The results are a pool of 3D positioned MED-portions to be hybridised.

Annexe on MED-Hybridise protocol

THOUSANDS OF «MED-PORTIONS» HAVING 3D-INTERACTIONS WITH THE PROTEIN QUERY

RESULTS = INNOVATIVE FRAGMENT-BASED MOLECULES INCLUDING ANNOTATION ONTO CHEMICAL LIBRARIES

Mine for Similar 3D interactions

Combine in 3D MED-portions

3D PROTEIN STRUCTURE IN YOUR HAND ?

Protein-Fragment Interaction Database from PDB (MED-SuMo) + Chemical libraries (PubChem, ...)

Annexe on MED-SuMo user interfaces

Advanced selection to build any query on the surface of your protein of interest

Smart 3D visualization with MED-SuMo objects (H-bond donor, acceptor, aromatic, hydrophobic, ...)

Hierarchical classification of MED-SuMo hits according to the MED-SuMo signature

Multiple hit display

Superimposed hit on the query with all MED-SuMo objects from the common signature

MED-SuMo Query builder

MED-SuMo comparison

MED-SuMo hit browser

Default binding sites

MED-SuMo objects selected into the query

PDB selector to generate optional subset prior to MED-SuMo comparison

Very interactive hit list including each MED-SuMo object signature (H-bond donor, acceptor, aromatic, hydrophobic, ...)

References

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

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

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

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

[5] Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science. 2007 Nov 23;318(5854):1258-65.

[6] Topiol S, Sabio M. Use of the X-ray structure of the Beta2-adrenergic receptor for drug discovery. Bioorg Med Chem Lett. 2008 Mar 1;18(5):1598-602.