THE 100K PDB BIOSTRUCTURES, A UNIQUE SOURCE OF BIOISOSTERIC REPLACEMENT BY POCKET MINING: FURTHER STATISTICAL ANALYSIS AND VALIDATION OF THE FC-BIOISOSTERE SOFTWARE SOLUTION

Jean-Yves Trosset[1,*], Christophe Tav[1], Raphael Berthier[1], Manon Réau[1], Lionel Chesnais[1], Christophe Linhard[1], Sindou Sougoumar[1], Stewart Adcock[2] and François Delfaud[2,3]

[1]*Laboratoire BIRL Sup'Biotech, 66 rue Guy Môquet, 94800 Villejuif France,*
[2]*Felix Concordia SARL, 400 av de Roumanille, 06906 Sophia Antipolis France,*
[3]*MEDIT SA, 2 rue du Belvédère, 91120 Palaiseau France*
*Correspondence to fdelfaud@felixc.eu

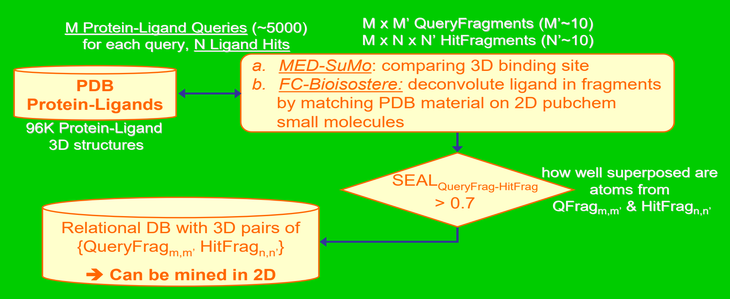
Bioisosteres in medicinal chemistry refers to structural changes on molecule that are not affecting existing biological activities. It's a powerful tool to optimize the pharmaco-dynamic profile. Source of bioisosteric replacement are provided by (1) manual inspection of structure-activity literature (2) collection over such literature, (3) automatic analysis of Structure-activity data to detect chemical substitution that are maintaining the activity profile, (4) data-mining of 3D molecular data such as the CSD or the PDB databases.

Here we are exploring bioisosteric rules extracted by crossmining in 3D/2D the PDB and small Pubchem molecules to detect local pair of similar fragment-protein 3D interactions, as implemented in FC-Bioisostere software [Moriaud2011]. First, new statistics are measured on the overall distribution of those pairs of superposed PDB-based chemical moieties. 2D duplicates pairs of bioisoscteric replacement are detected and sorted according to cases where at least one pair is having the same functional annotation in the protein binding cavity, in order to score the chemical mutation suggested by other pairs having the same 2D fragments. Second structure-activitiy data related to bioisosterism mutation are qualitatively compared with FC-bioisostere pairs

This work on *FC-Bioisostere* explores one step further the overall chemo-proteomic challenge as initiated in our C2P Chemo-Proteomic Platform to better understand and predict interactions between, on one hand, ligands and all related fragments and, on the other hand, binding sites and all related subpockets

Material & Methods:

1)FC-Bioisostere protocol[2,3]:



We build a database of bioisosteric pairs by driving from *FC-Bioisostere* the *MED-SuMo* MEDIT software to superpose binding pocket accross the PDB : a whole set of 5139 protein-ligand binding site queries (Res<2,5A ; 350<MW<550 ; PDB_occurencies<11 ; Sept2013) are compared toward all PDB, a maximum of 100 Hits with a *MED-SUMo*-score above 6.0 are retreived

From those ligand superpositions, a database of 376096 3D/2D pairs of bioisosteric fragments is generated with :

- the Fragmentor mode to deconvolute PDB ligand in fragments with a list of 281472 Pubchem fragments having 3 to 13 atoms
- filter on FragmentCandidate > 5Heavy(or dummy) Atoms
- 2D filter on *MED-SuMo* Ligand hit per binding site
- Seal Score (to detect overlaping fragment) > 0,7

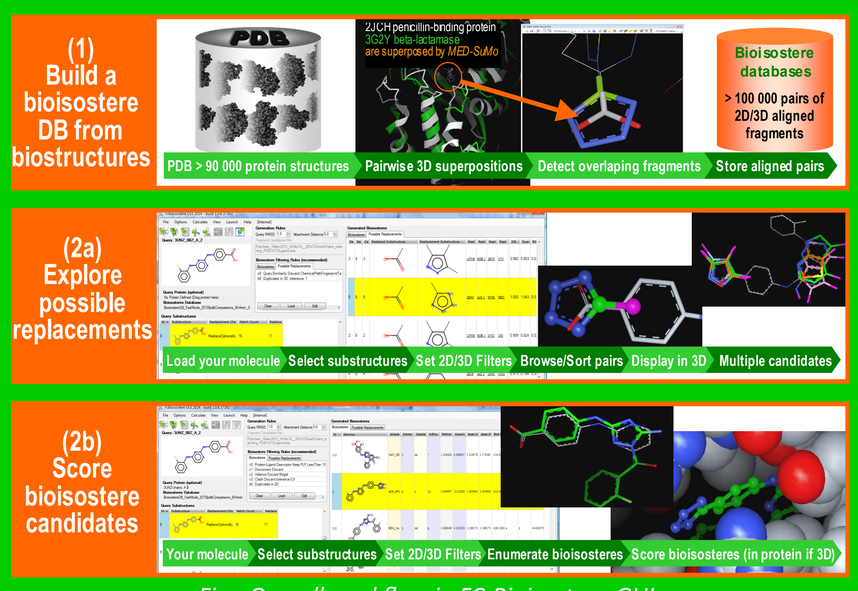


Fig.: Overall workflow in FC-Bioisostere-GUI

2) Qualifying frequent bioisosteric pairs[1]:

Canonicle smiles were used to detect and enumerate 2D duplicates pairs of {FragA,FragB} corresponding to different pairs of biostructures having a local 3D interaction similarities. 2DduplicateID and 2DduplicateOccurency are stored in relational database.

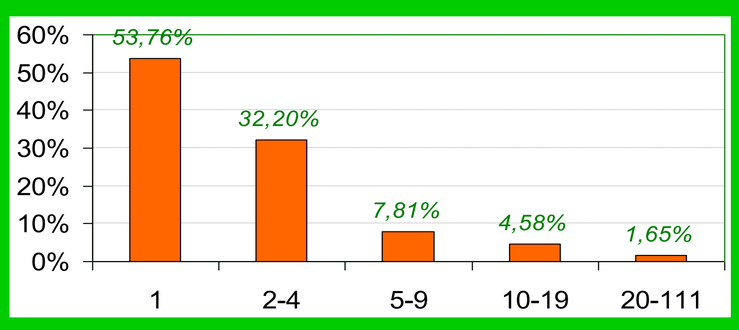
Uniprot protein sequence ID and PFAM (Protein Family annotation) are imported in 2 separate tables

SQL expressions make possible to modulate the filtering of the whole database upon Frequent/NoFrequent AND/OR intra/inter Pfam AND/OR intra/inter Uniprot

3) Exploring biostructural bioisostere pairs in Structure-Activity bioisosteric pairs[1]

A subset corresponding to the smaller biostructural fragments in the FC-Bioisostere database were submitted to SwissBioisostere web site [Wirth2013] to evaluate qualitatively the overlaps between the Biostructural and the Structure-Activity approaches to detect bioisostere candidates.

Qualifying frequent bioisostere pairs:



Distribution chart of 2D duplicate pairs found amongst the 376096 biostructural bioisosteres pairs of the FC-bioisostere database.

→ 56% of the fragment pairs are related to protein partners sharing the same Uniprot ID

→ 87% of the fragment pairs are related to protein partners sharing the same PFAM ID

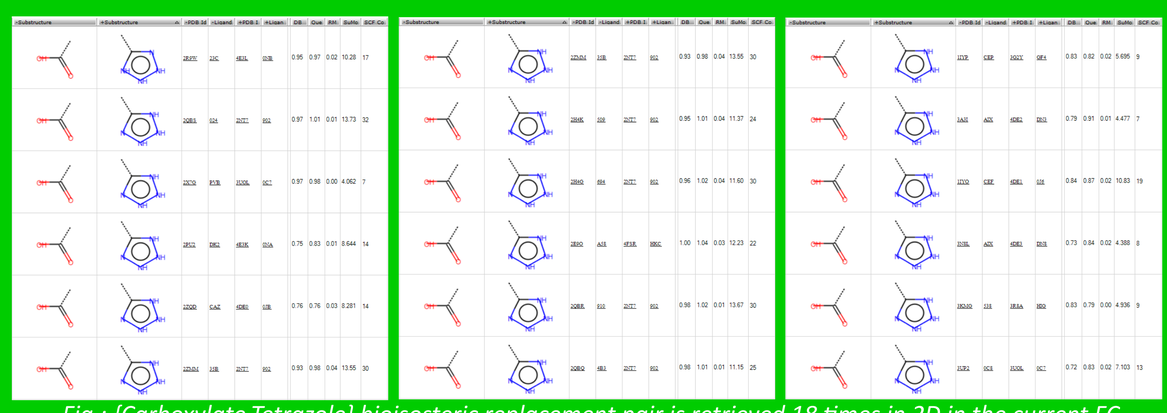
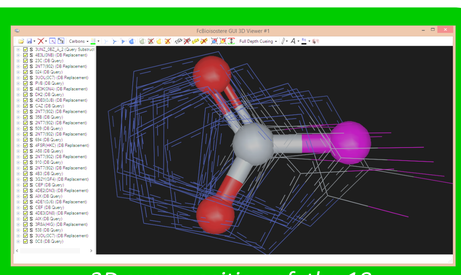
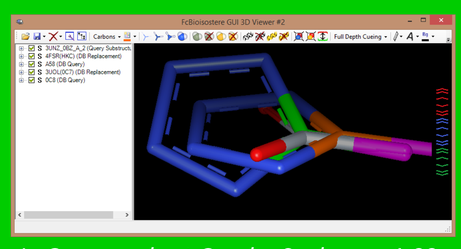


Fig.: {Carboxylate,Tetrazole} bioisosteric replacement pair is retrieved 18 times in 2D in the current FC-bioisostere 3D database

Columns from the left to the right are: QueryFragment, HitFragment, QueryPDB ID, QueryLigand ID, HitPDB ID, HitLigand ID, OverlapSeal score as stored in the Database, OverlapSeal score after the 3D superposition of the {QueryFragment,HitFragment} onto the structure of the input molecule, RMSD between the input molecule and the QueryFragment, MED-SuMoScore, Number of chemical features shared between 2 binding sites



3D superposition of the 18 {Carboxylate,Tetrazole} pairs from the current FC-bioisostere database



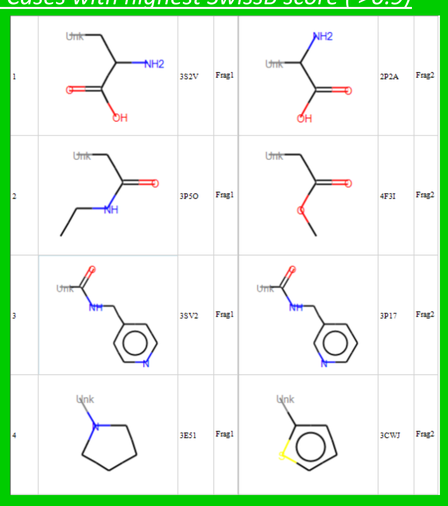
In Orange = best OverlapSeal score 1.00
In Green = worst OverlapSeal score 0.72

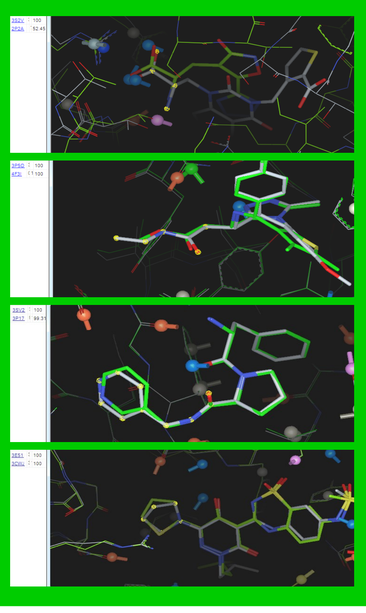
Exploring biostructural FC-Bioisostere pairs in Structure-activities data:

Fig.: set of 2D matches between biostructural bioisostere pair and structure-activity SwissBioisostere bioisostere pair

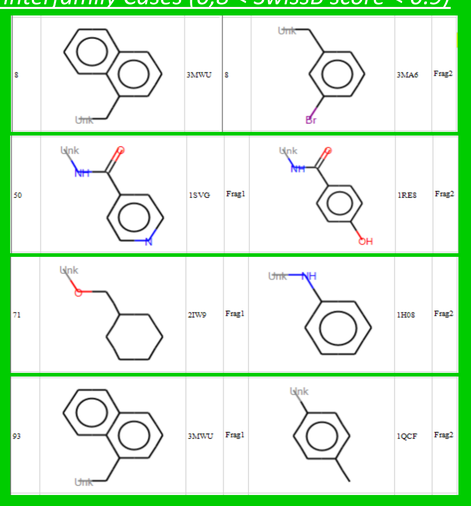
On the right: corresponding MED-SuMo superposition shows the 3D overlap of bioisostere fragments ; protein-ligand structure in white is the MED-SuMo query while in green is the protein-ligand structure showing 3D interaction similarity ; in stick mode are the MED-SuMo based superposed ligand ; selected in yellow are the query fragment of the bioisostere ; % sequence identity is on left of the 3D view

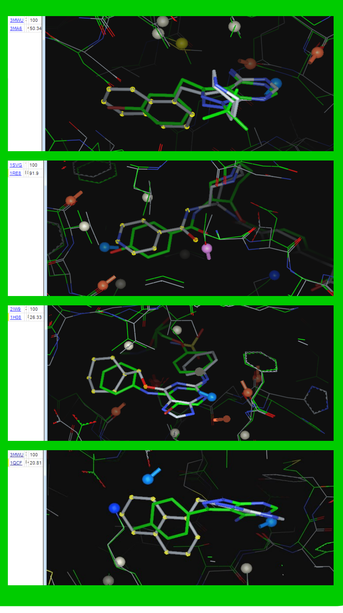
4 Cases with highest SwissB score (>0.9)





4 interfamily Cases (0,8 < SwissB score < 0.9)



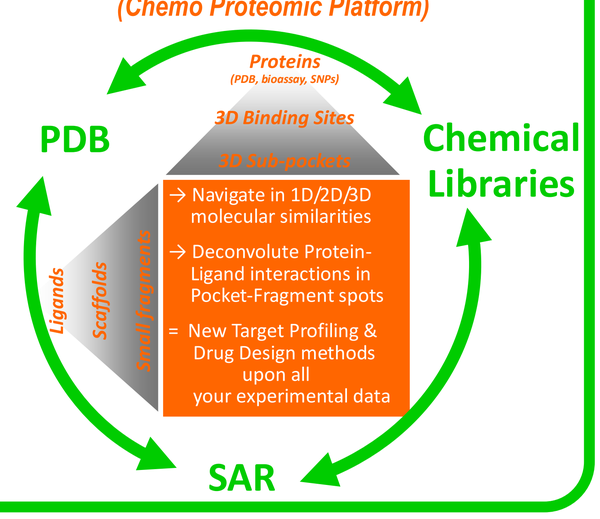


Conclusion:

FC-Bioisostere is proposing a new protocol to identify putative bioisostere pairs of chemical moieties by detecting 3D overlapping fragments from superposition of binding pocket across the PDB. Multiples occurrences may be used to evaluate the confidence of the proposed 3D bioisostere substitution. Mining these pairs against Pfam or Uniprot annotations allows to focus on inter-family protein-based material.

FC-Bioisostere proposes 3D bioisostere pairs that is some case are already identified in 2D by structure-activity methods, and some other cases new to the area. The 3D superposition of Fragment pairs (as observed by pocket mining with MED-SuMo software) allows to better understand 3D interaction mechanisms and eventually to orient the construction of a 3D model of interaction of the bioisostere into a defined 3D protein target.

To facilitate the FC-Bioiosstere experience and to provide the best performance, a SQL query makes possible to generate a separate version of the 3D bioisostere database where duplicates are removed to keep the best pairs according to the quality of the local 3D superposition



References:

Moriaud, F., Adcock, S. A., Vorotyntsev, A., Doppelt-Azeroual, O., Richard, S. B., and Delfaud F. (2011) "A Computational Fragment Approach by Mining the Protein Data Bank: Library Design and Bioisosterism", ACS Symposium 2011, series1076, Chapter book 5:71-88

Wirth, M., Zoete, V., Michielin, O., & Sauer, W. (2013) SwissBioisostere: a database of molecular replacements for ligand design, Nucleic Acids Research, 41 (D1), D1137-1143.