

Fast software development in drug design with the C2P-API toolkit : Rebuilding ligand bound conformation from PDB subpocket superpositions

François Delfaud (fdelfaud@felixc.eu - Tel +33 6 7513 0847)
FELIX CONCORDIA SARL, 400 av de Roumanille, BP309, 06906 Sophia Antipolis France



Literature in the FBDD field reported cases where the binding location and orientation of small co-crystallized fragments is preserved when these fragments are chemically evolved into larger ligands. Crystallography also revealed the central role of some weak fragment binders. With the benefit of the PDB growth, pocket superposition illustrated some ligand substructures having conserved binding modes towards multiple targets. Recently, we showed that 48% of 2,241 ligands structures uniquely represented in the PDB could be reproduced at 80 % of their atom coordinates within a 1.0 Å deviation by using chemical material from 3D pocket superpositions.

The development of an heuristic to rebuild ligand bound conformation from PDB pocket superpositions requires at least methods to detect Maximum Common Substructures between the compound to 3D-rebuild and all PDB ligands from superposed pockets, and methods to 3D-hybridise those MCS fragments. It was possible to develop a rebuildier prototype including a graphical interface in about 3 months by combining (a) the richness of the C2P-API Chemo-Proteomic Advanced Programming Interface, (b) the agility of the .Net C# language including reconfigurable dictionaries to include molecular rules, quasi-automatic parallelization and safe memory management, and (c) the comfort of the Microsoft Visual Studio Community IDE in developing graphical interface and debugging software.

Our auto-converging knowledge-based iterative hybridiser software is 3D-rebuilding as much as possible any input 2D-ligands into any biostructural binding site. We selected inhibitors co-crystallized in multiple PDB proteins to validate our protocol.

Context

1) Fragments compared to larger binders:

Kosakov et al.¹ reviewed multiple crystallography FBDD cases where 2D fragments are found into larger molecules within the protein binding site with the same or not bound conformation.

In similar incremental fragment-to-molecules strategy, Davies et al.² validated with crystallography that the free fragment 3-(4-Chlorophenyl) propanoic acid (MW 185 Da) maintains the same bound conformation when inserted in three active compounds with only modest adjustments of the binding site of KEAP1-Kelch-NRF2.

Behnen et al.³ observed from the two biostructures that the pose of 1,2-propanediol probe (MW 76 Da) coincided remarkably well with those of much larger CDPME superstructure ligand in 4-diphosphocytidyl-2C-methyl-d-erythritol synthase target.

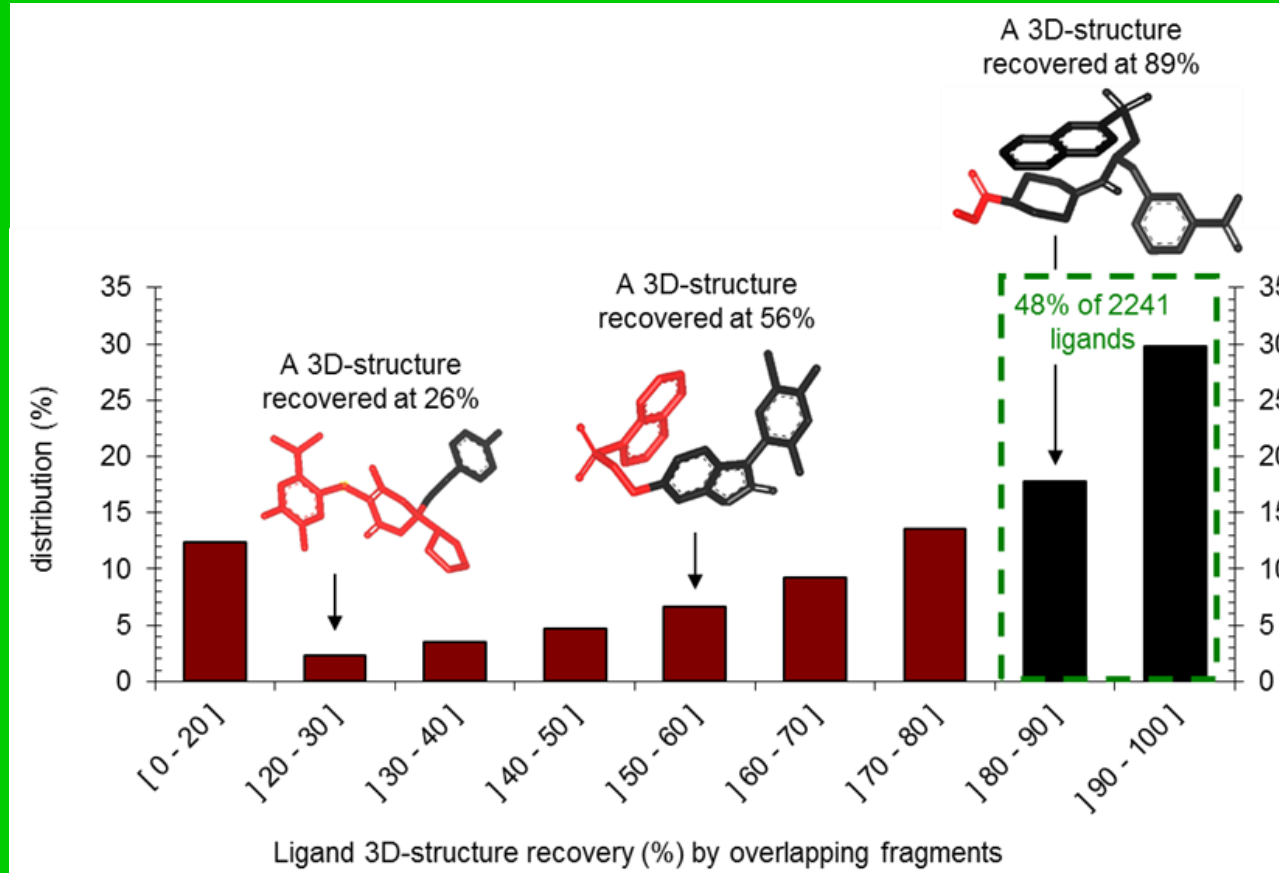
In Astex data where fragments and ligands binding modes were experimentally determined on HSP90, PKB, uPA, CDK2 and BACE targets, Murray et al.⁴ observed on successful 39 fragment-to-lead campaigns that 80% of these atoms are topologically retained in the lead molecule with a similar position (mean shift of 0.79 Å).

In a dual inhibitor drug design strategy, Taylor et al.⁵ found an unchanged binding mode between an initial benzodiazepinones fragment (MW 176 Da) and five enlarged compounds 12, 14, 23, 27 and 28 (up to MW 386 Da) all co-crystallized in CPB bromodomains while affinities gain 4 logs .

2) Mining similar PDB pockets to 3D rebuild ligand?

In a recent work we evaluated to which extend the binding conformation of co-crystallized ligands can be inferred from the sole knowledge of the 3D-structure of their binding site [M. Vieillevoye et al, Submitted for publication]. Our protocol extracts matching ligand substructures from any ligand complex of the PDB which binding pocket could be 3D aligned on the reference binding site.

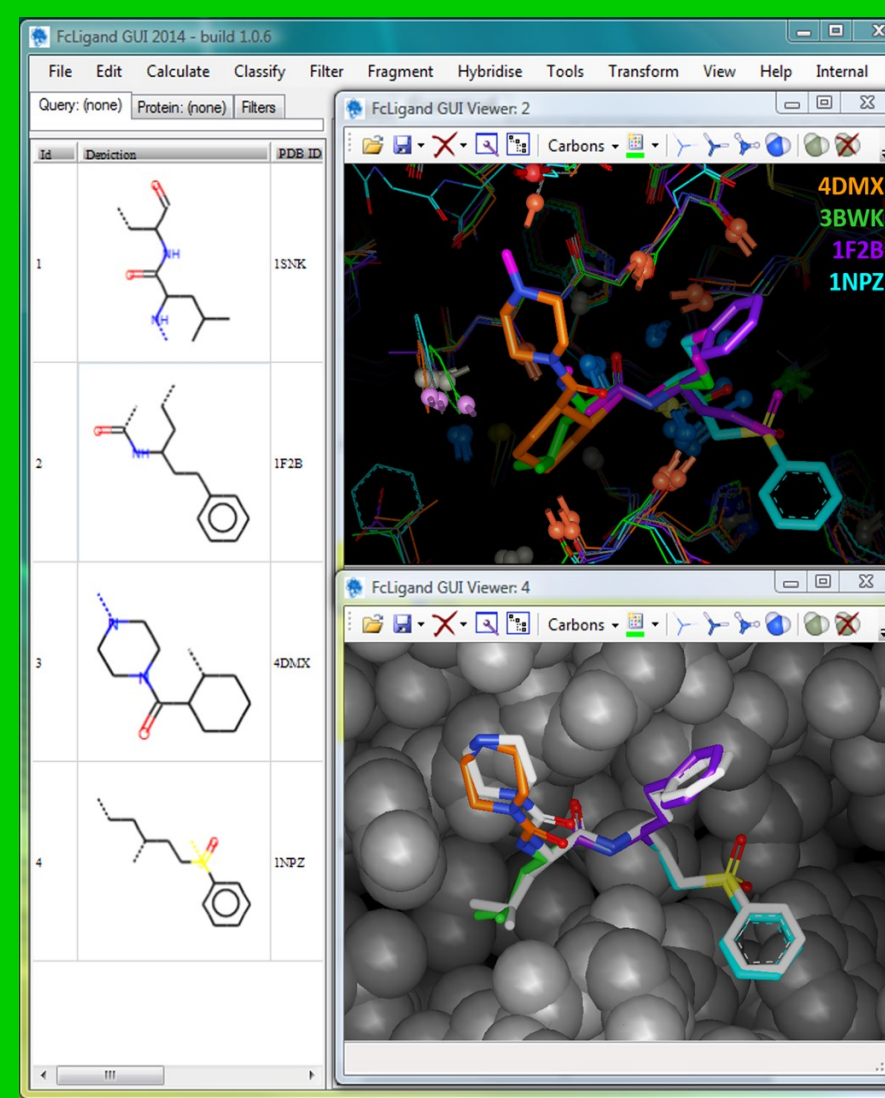
Fig. Using a dataset of 2,241 protein-ligand complexes uniquely represented in the PDB, we show that this biostructural database hosts enough information to reconstruct nearly half (48%) of the ligands with more than 80% of their input coordinates within a 1 Å distance deviation from constitutive substructures existing in other PDB entries.



3) Example of successful ligand reconstruction

FcLigand software was able to successfully rebuild vinyl sulfone inhibitor, a ligand of cathepsin K (ID: 1MEM) by coupling 3D-hybridisation of fragments from sub-pocket superposition, and a substructure constraint to the inhibitor. 4 fragments which are associated with different proteins in the PDB (cathepsin K, falcipain-3, cruzain protein, cathepsin S) were necessary for this reconstruction.

Fig. Top panel: the 4 fragments superposed by MED-SuMo software (MEDIT SA) onto their targets before hybridisation. Lower panel: The hybrid molecule (in colour) and vinyl sulfone inhibitor (in grey) in the binding site of cathepsin K. [M. Vieillevoye et al, Submitted for publication]



Autoconverging heuristic to rebuild ligand bound conformations from PDB pocket superpositions

1) Protocol:

The proof of concept on a cathepsin K inhibitor (see context section) required more than an hour of CPU time due to the combinatorial explosion upon the 3D hybridisation of pool of aligned PDB ligands from the pocket superposition.

This time, we set up an auto-converging knowledge-based iterative hybridisation algorithm to 3D-rebuild any input 2D-ligand into any biostructure binding site.

Our protocol provides an optional form to control MED-SuMo software (MEDIT SA) which is aligning PDB protein-ligand complexes onto the targeted binding site. Then clashing atoms with the binding site in this pool of aligned PDB ligands are removed. A Maximum Common Substructure search is performed with the 2D ligand to rebuild in order to keep only substructures that can contribute to the ligand reconstruction. A simple fingerprint to encode atom and bond alignments between MCS fragment and the 2D ligand to reconstruct is computed. Iteratively, MCS fragments sharing same bonds of the ligand 2D are tentatively 3D hybridised. 2D duplicates and substructure duplicates for RMSD below a defined threshold are removed after each hybridisation cycle.

2) Results on a HIV-RT inhibitor:

We tested our protocol on a set on 100 PDB ligands, here is the result for a HIV-1 RT with non-nucleoside inhibitor annulated Pyrazole (ligand code PLZ, in PDB code 3DYA) that was almost reconstructed (0.37 Å deviation with PDB)

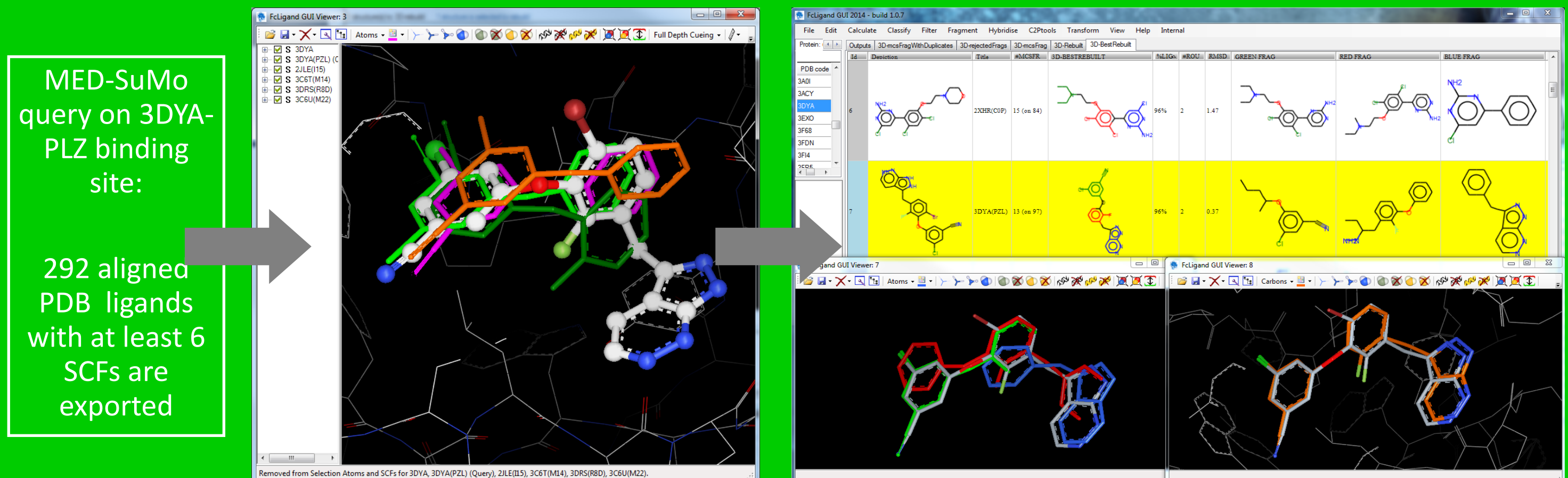


Fig. : left view shows from PDB MCS-fragments a chloro phenoxybenzonitrile in 4 close positions almost in agreement with the PDB PLZ experimental position ; chemical spreadsheet shows from left to right the 2D ligand, the best 3D reconstruction and the hybridised MCS fragments ; middle viewer displays the 3 elected MCS fragments to rebuild PZL in green, red and blue (binding site superposition with 3ATM-MO6, 3D16-PDZ and 3E01-PZ2) ; right viewer shows in orange the rebuilt PLZ.

3) Developed with the C2P-API:

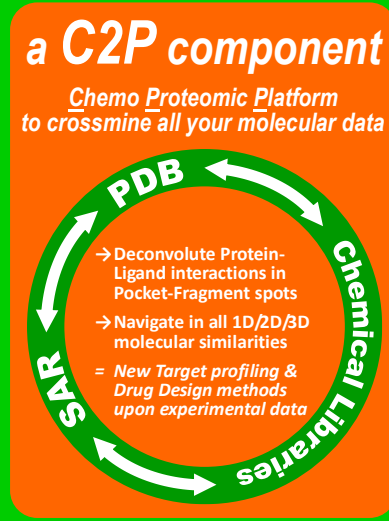
```
// 2. LOOP over each Lig3D ...to extracting 2D-MCS chemical moiety
//
Parallel.ForEach(Lig3Ds, lig3D =>
{
    // Calculate the MCS between Frag3D and lig3D
    MaximumCommonSubstructureSearch mcslig = new MaximumCommonSubstructureSearch();
    mcslig.MinimumMcsSize = mcslig2D.Lig3DMinimumSize;
    Molecule mcsFrag3D = lig3D.Clone(); // Cloning the object as
    int mcsSize = mcslig.CalculateMcs(mcsFrag3D, lig3D); // gets number of 2D over

    // If MCS exist, then...
    if (mcsSize != 0)
    {
        var frag3DlistMatchingAtomID = mcslig.Results; // gets the maximum matching
    }
}
```

Fig. : Parallel.ForEach .Net method distributes over the 8 logical cores of 17 Intel processor the MCS Maximum Common Substructure searches between each PDB ligand from pocket superposition and the 2D ligand to rebuild

C2P-API (Chemo-Proteomic Advanced Programming Interface)

Overview:



The C2P Advanced Programming Interface (API) provides a large set of chemo-proteomic methods in C# language under a L-GPL license to develop very quickly new expert chemo-proteomic software applications. Combined with the Microsoft Visual Studio IDE (integrated development environment), Computational Chemists are able to access a rich and convenient environment to develop small to large applications for MS-Windows or Linux platforms.

Main Features:

- 1D/2D/3D descriptors
- 2D fingerprints and similarity searches
- Substructure/superstructure searches
- Clustering
- Ligand fragmentation
- Fragment hybridisation
- 1D/2D/3D interactive spreadsheet and 3D viewer
- File parsers
- MMFF minimisers
- ...

Main Benefits:

- Rich toolset to manage in 1D/2D/3D macromolecules and ligands;
- Fast methods, some are parallelized for multi-cpu/core computer node;
- Robustness (core architecture in FcLigand, FcBioisostere and FcCutlass);
- Comfortable software development environment when used in Microsoft Visual Studio IDE (code completion, debugger, CPU/RAM profiler, GUI designer, rich C# methods, ...);
- Allow users to focus on new ideas rather than struggling with technical low level software/mathematical issues;
- Open sources benefit (flexibility, security, auditability, cost, ...);
- Interactive documentation (all C2P-API methods are fully documented and searchable thanks to Doxygen documentation generator);
- Fully Object oriented, all C# classes are organized in different namespaces with interfaces and inheritances to provide consistent methods

License: Lesser-GPL license

Free to download: please contact support@felixc.eu

API documentation:

Fig. : subset of the C2P class list (left), detail of a class with methods & arguments (right)

Conclusion:

We have developed here an innovative protocol to tentatively rebuild bound conformation of any input 2D ligand by mining PDB knowledge, it's 3D-hybridising iteratively PDB ligands from pocket superpositions across the full PDB onto the targeted binding site. By importing only 3D Maximum Common Substructures with the ligand to rebuild from the pool of PDB aligned ligands, and by limiting 3D hybridisation trials to list of same topological bonds, our protocol is becoming auto-convergent and provides performance in the ranges of seconds. Thanks to the richness of the C2P advanced programming interface, prototyping and releasing were performed in 3 months. We observed qualitatively a rate of reconstruction above 90% on half of a set of 100 PDB ligands we tested to rebuild.

Now available into the FcLigand software, the graphical interface of the new module allows (1) to drive the pocket mining toward multiple targets, (2) to review all intermediate PDB MCS fragments, (3) to relax with molecular mechanics any hybridised molecules, and (4) to use standard protein-ligand scoring functions.

Such ligand 3D reconstruction strategy should be considered as complementary to crystallography and virtual screening technologies.

In order to increase the number of 3D predicted atoms, we are planning to implement the integration of PDB-based bioisosteric fragment pairs⁶ to bridge existing PDB pocket/fragment similarities to the topology of the ligand to rebuild.

References:

- [1] Kosakov, D.; Hall, D. R.; Jehle, S.; Jehle, S.; Luo, L.; Ochiana, S. O.; Jones, E. V.; Pollastri, M.; Allen, K. N.; Whitty, A.; Vajda, S. Ligand Deconstruction: Why Some Fragment Binding Positions Are Conserved and Others Are Not. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (20), E2585-E2594.
- [2] Davies, T. G.; Wixted, W. E.; Coyle, J. E.; Griffiths-Jones, C.; Hearn, K.; McMenamin, R.; Norton, D.; Rich, S. J.; Richardson, C.; Saxty, G.; Willems, H. M. G.; Woolford, A. J.-A.; Cottom, J. E.; Kou, L.-P.; Yonchuk, J. G.; Feldser, H. G.; Sanchez, Y.; Foley, J. P.; Bolognese, B. J.; Logan, G.; Podolnik, P. L.; Yan, H.; Callahan, J. F.; Heightman, T. D.; Kerns, J. K. Monocyclic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1/NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J. Med. Chem.* **2016**, *59* (8), 3991-4006.
- [3] Behnen, J.; Köster, H.; Neudert, G.; Craan, T.; Heine, A.; Klebe, G. Experimental and Computational Active Site Mapping as a Starting Point to Fragment-Based Lead Discovery. *ChemMedChem* **2012**, *7* (2), 248-261.
- [4] Murray, C. W.; Verdonk, M. L.; Rees, D. C. Experiences in Fragment-Based Drug Discovery. *Trends Pharmacol. Sci.* **2012**, *33* (5), 224-232.
- [5] Taylor, A. M.; Côté, A.; Hewitt, M. C.; Pastor, R.; Leblanc, Y.; Nasveschuk, C. G.; Romero, F. A.; Crawford, T. D.; Cantone, N.; Jayaram, H.; Setser, J.; Murray, J.; Beresini, M. H.; de Leon Boenig, G.; Chen, Z.; Conery, A. R.; Cummings, R. T.; Dakin, L. A.; Flynn, O. W.; Kaufman, S.; Keller, P. J.; Kiefer, J. R.; Lai, T.; Li, Y.; Liao, J.; Liu, W.; Lu, H.; Pardo, E.; Tsui, V.; Wang, J.; Wang, Y.; Xu, Z.; Yan, F.; Yu, D.; Zawadzke, L.; Zhu, X.; Zhu, X.; Sims, R. J.; Cochran, A. G.; Bellon, S.; Audia, J. E.; Magnuson, S.; Albrecht, B. K. Fragment-Based Discovery of a Selective and Cell-Active Benzodiazepine CBP/EP300 Bromodomain Inhibitor (CPI-637). *ACS Med. Chem. Lett.* **2016**, *7* (5), 531-536.
- [6] Moriaud et al. "A Computational Fragment Approach by Mining the Protein Data Bank: Library Design and Bioisosterism", ACS Symposium 2011, series1076, Chapter book 5:71-88