Design and Application of **LEF**, a library of chemical fragments with different **L**ocal **E**nvironment of **F**luorine

Anna Vulpetti

FBLD Conference 2010

Philadelphia, 11 October 2010

NOVARTIS

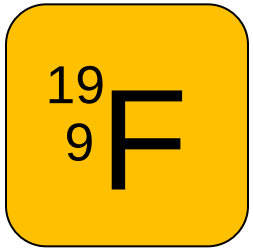
Outline

- Fluorine in **drug discovery**
- **^{19}F-NMR Screening**
- Development of a **new descriptor** (F-FP-L)
- **Design a library** of chemical fragments with different *L*ocal *E*nvironment of *F*luorine (LEF)
- Predict **^{19}F chemical shift** of new molecules
- Analyse different protein environments of **fluorine in protein-ligand** crystal structures

Fluorine in Drug Discovery

- Conformational influences (binding affinity)
  - Proline conformation / Ar-OCF_3 vs Ar-OCH_3/ gauche effect / ...
- Protein-ligand interactions (binding affinity)
  - H-bond, lipophilic contacts, multipolar interactions
- Influences on physico-chemical properties (logP, pK_a)
- Bioavailability (absorption and transport)
- Metabolic stability (increased oxidative stability against enzymatic attack)
- ^{19}F NMR Spectroscopy

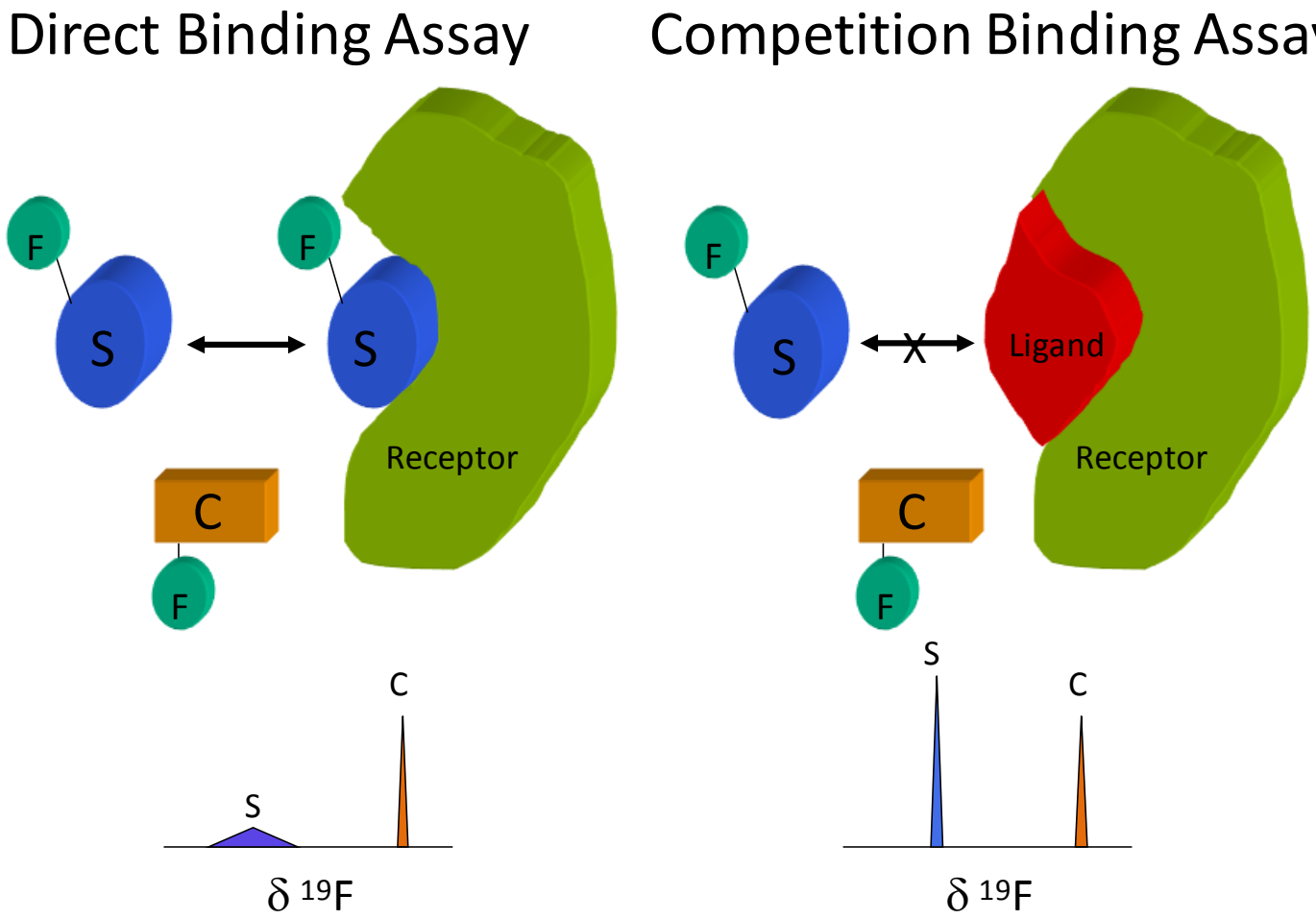
Screening with ^{19}F NMR Spectroscopy



- Good sensitivity (0.83 ^{1}H)
- 100% Natural Abundance ^{19}F NMR Active Isotope
- No problems with overlap
- No interferences from buffer, solvent, detergent signals
- ~ 15% of ACD molecules contain F
- Favourable Transverse Relaxation

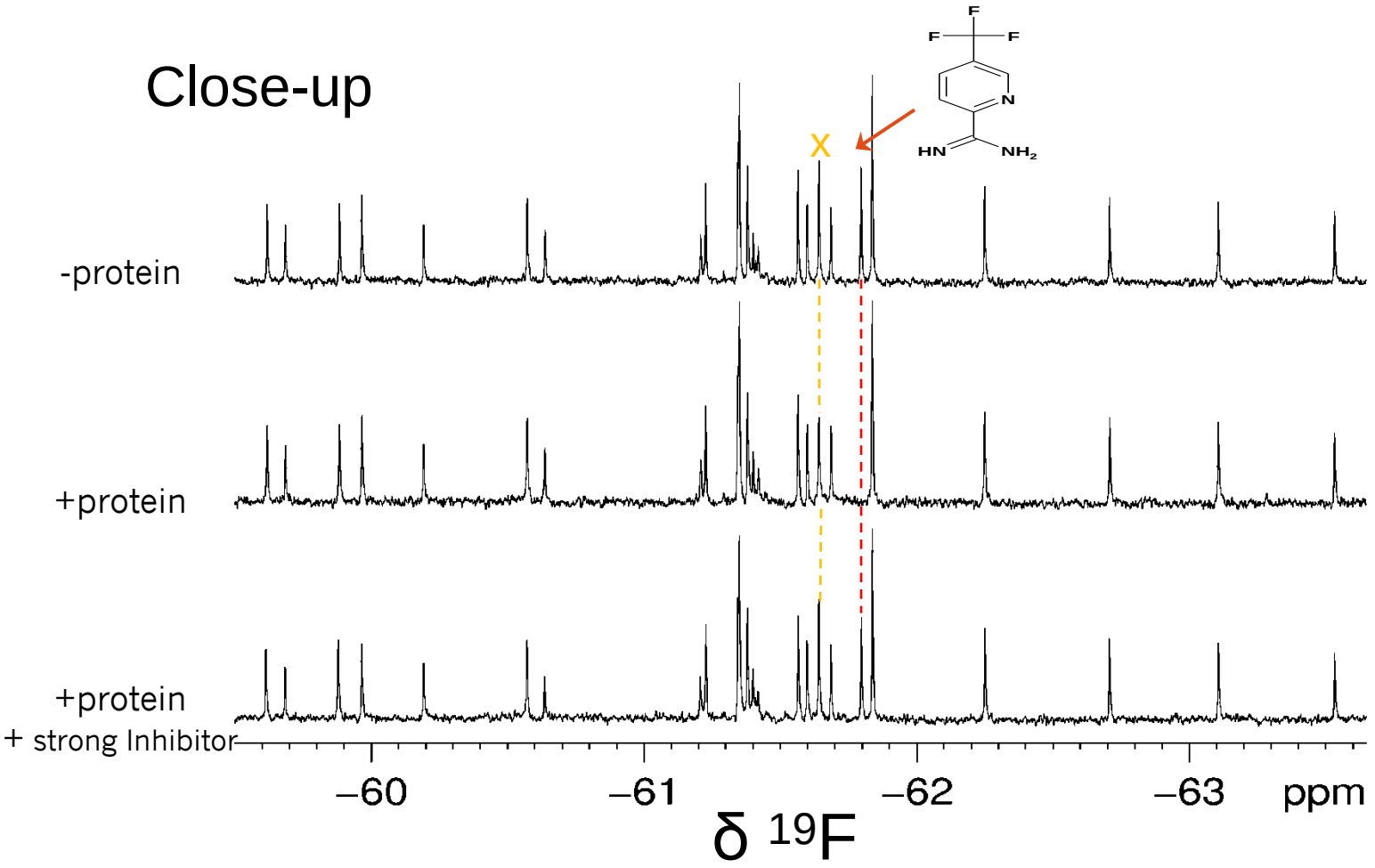
Dalvit C., Progress in NMR Spectroscopy 51, 243 (2007)
Dalvit C., Concepts in Magnetic Resonance 32A, 341 (2008)

FAXS (**F**luorine chemical shift **A**nisotropy and e**X**change for **S**creening)



Dalvit et al. Comb. Chem. & HTS 5, 605 (2002)
Dalvit et al. JACS 125, 7696 (2003)

Screening with **Large Mixtures** (36 Molecules)
+ **Validation of the Hits** via Competition Binding Experiments



Vulpetti, A., Hommel, U., Landrum, G., Lewis, R., Dalvit, C., J. Am. Chem. Soc. 131 (2009) 12949-12959

NOVARTIS

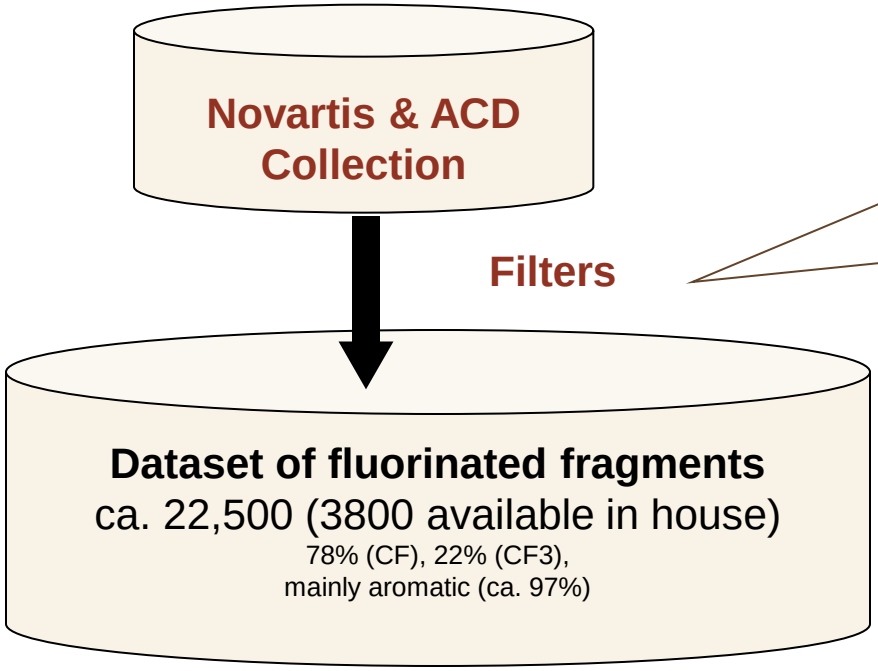
LEF (Library of Environment of Fluorine)

- **Goal:** design, construction and implementation of a diverse fragment library of fluorinated compounds for ^{19}F NMR-based screening

- **Why Fluorinated fragments:**
  - Exploit high sensitivity ^{19}F-NMR Screening technology
  - Identification of spy molecules for subsequent screening of compounds containing (or not) fluorine atom(s)

- **Why Fluorine in different local environment:**
  - Fluorine Local Environment is relevant for the protein-fluorine interactions. Identification of fluorophilic hot-spots by presenting the fluorine with different chemical and structural features.

Key steps

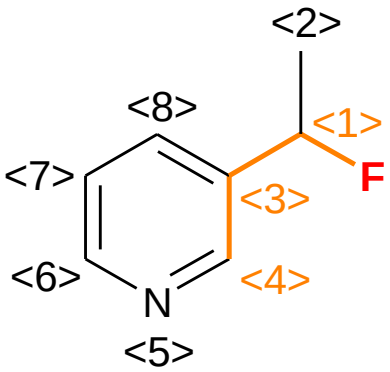


- One **CF3** or one **CF**
- Available amount > 20 mgr
- 140 ≤ MW ≤ 300
- HB acc.+ HB don. ≥ 3
- Exclusion of not organic atoms
- AlogP ≤ 3
- Rotatable bonds ≤ 4
- Rings ≤ 4
- Unknown stereoatoms/bonds = 0
- Unwanted chemical functionalities

1. **Clustering** according to two types of diversity: *global* (on the whole molecule) and *local* (based on the description of the 2D local chemical environment around the F atom and the CF_3 group).
2. **Annotation** of other calculated properties (synthetic tractability, in silico water solubility)
3. **Selection** by visual inspection

Fluorine local-environment fingerprints (F-FP-5)

- New descriptor derived from the topological-torsion descriptor concept
- Paths consisting of between one and five bonds and only starting from the fluorine atom or CF3 moiety are included



E.g. highlighted path: F-<1>-<3>-<4>

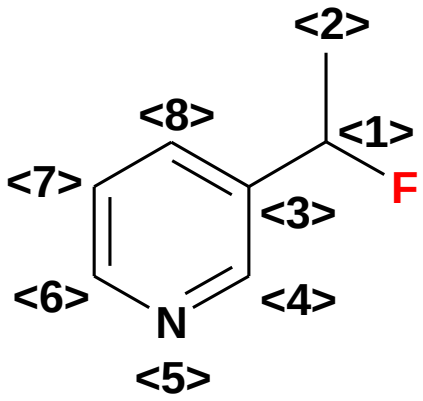
For each atom:
Element,
Number of π electrons,
Number of heavy neighbors (not counting those in the torsion).

F-0-0,C-0-1,C-1-1,C-1-1

Vulpetti, A., Hommel, U., Landrum, G., Lewis, R., Dalvit, C., J. Am. Chem. Soc. 131 (2009) 12949-12959

Fluorine local-environment fingerprints (F-FP-5)

- The set of all paths of length one to five bonds rooted at the fluorine atom are enumerated



Length	Path	Atom Codes	Bit Id
1	F-C1	(F-0-0,C-0-2)	b1
2	F-C1-C2	(F-0-0,C-0-1,C-0-0)	b2
2	F-C1-C3	(F-0-0,C-0-1,C-1-2)	b3
3	F-C1-C3-C4	(F-0-0,C-0-1,C-1-1,C-1-1)	b4
3	F-C1-C3-C8	(F-0-0,C-0-1,C-1-1,C-1-1)	b4
4	F-C1-C3-C4-N5	(F-0-0,C-0-1,C-1-1,C-1-0,N-1-1)	b5
4	F-C1-C3-C8-C7	(F-0-0,C-0-1,C-1-1,C-1-0,C-1-1)	b6
5	F-C1-C3-C4-N5-C6	(F-0-0,C-0-1,C-1-1,C-1-0,N-1-0,C-1-1)	b7
5	F-C1-C3-C8-C7-C6	(F-0-0,C-0-1,C-1-1,C-1-0,C-1-0,C-1-1)	b8

- Similarity

$$Sim(FP_i, FP_j) = \frac{2 * \sum_b \min(FP_{ib}, FP_{jb})}{\sum_b FP_{ib} + \sum_b FP_{jb}}$$

molA: 3 1 0 2 molB: 2 0 0 2

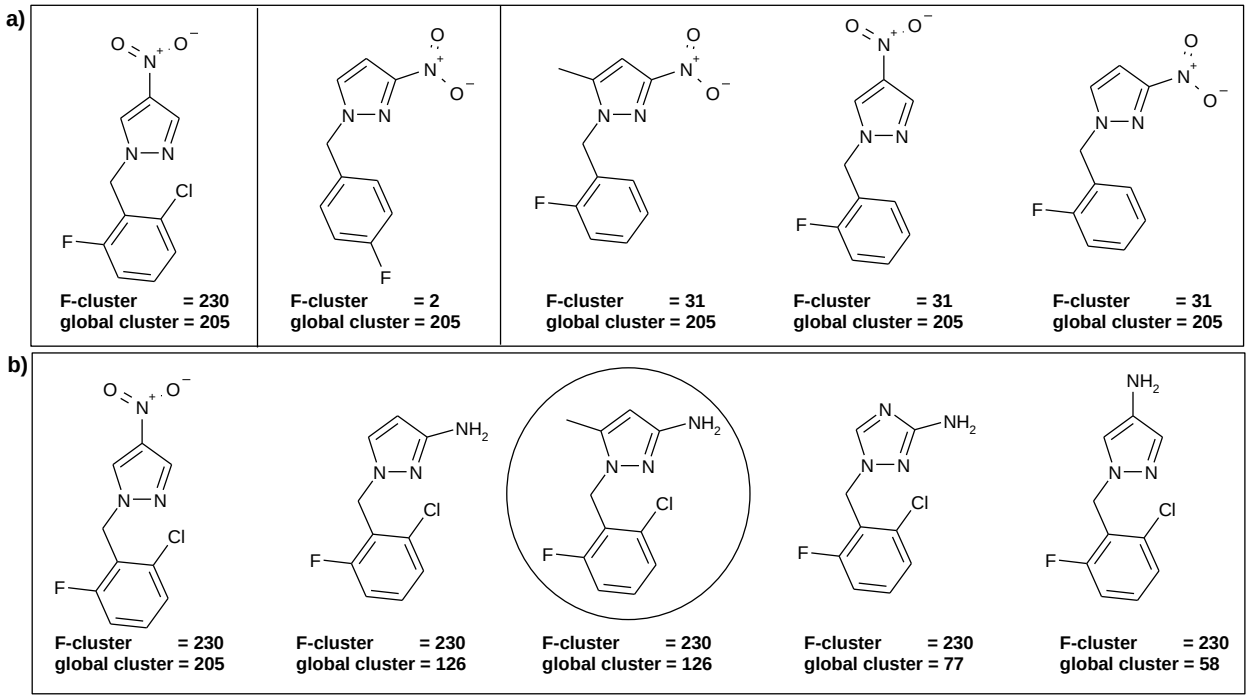
$$Sim = \frac{2.0 * (2+0+0+2)}{6+4} = 8/10 = 0.8$$

Vulpetti, A., Hommel, U., Landrum, G., Lewis, R., Dalvit, C., J. Am. Chem. Soc. 131 (2009) 12949-12959

NOVARTIS

Clustering

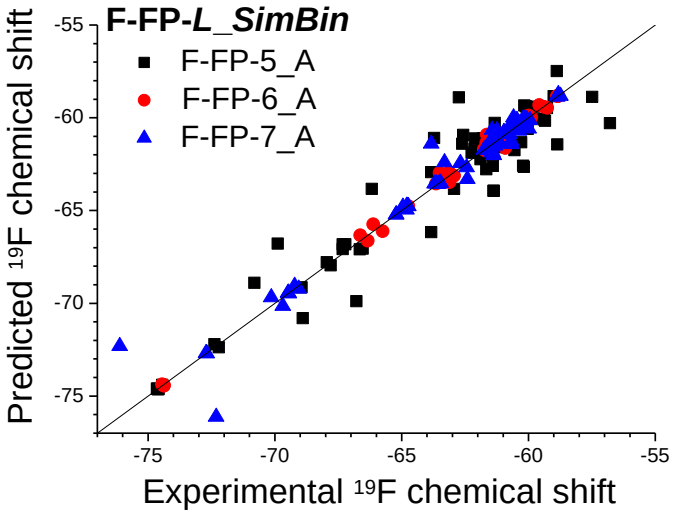
- Based on:
  - *local* fluorine-environment fingerprints (F-FP-5),
  - whole molecular structure (*global* description, FCFP-4).



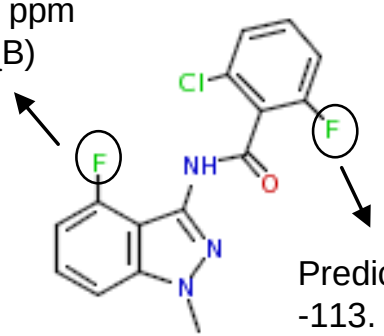
Molecules of the same F-cluster have similar ^{19}F chemical shifts

Extension of F-FP-*L* descriptor to ^{19}F chem. shift prediction

- New extended fluorine fingerprint and a distance-weighted *k*-nearest neighbors algorithm applied on a training set of known chemical shifts.



Predicted ^{19}F δ:
-119. 83 ppm
(F-FP-7_B)



Predicted ^{19}F δ:
-113. 98 ppm
(F-FP-7_B)

MW = 321.7

Experimental ^{19}F δ: -113.51, -120.00 ppm

- The quality of the prediction depends on both the training set size and its diversity. The expansion of our current chemical shift training set is on going.

Vulpetti, A., Landrum, G., Rüdisser, S., Erbel, P., Dalvit, C., J. of Fluorine Chem., 131 (2010) 570-577

NOVARTIS

Experimental Conditions

- **Current LEF**: 120 mixtures (54 CF_3 and 66 CF) (12 molecules per mixture) (expansion in progress)
- **Mixture concentration** : CF_3 and CF mixtures are tested at 15-20 μM and 35-40 μM, respectively with 600 MHz without cryoprobe and ~ 5 μM and ~ 15 μM, respectively with 600 MHz with cryoprobe
- **Protein concentration** : 1-10 μM
- **Acquisition time** for each mixture is 10-15 minutes [one 96-wells rack (1152 molecules) in < 24 hours]
- **Spectral analysis** performed with peak peaking
- **Tested with success** on different targets

LEF Screening in **Trypsin**

LEF

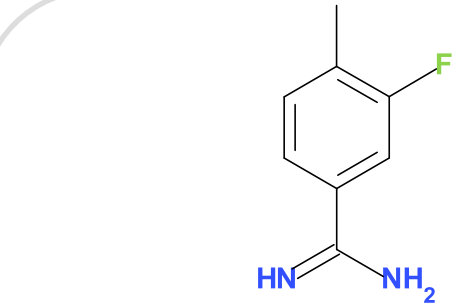
^{19}F-NMR Screening

31 NMR HITS

IC_{50} at 1 mM against trypsin

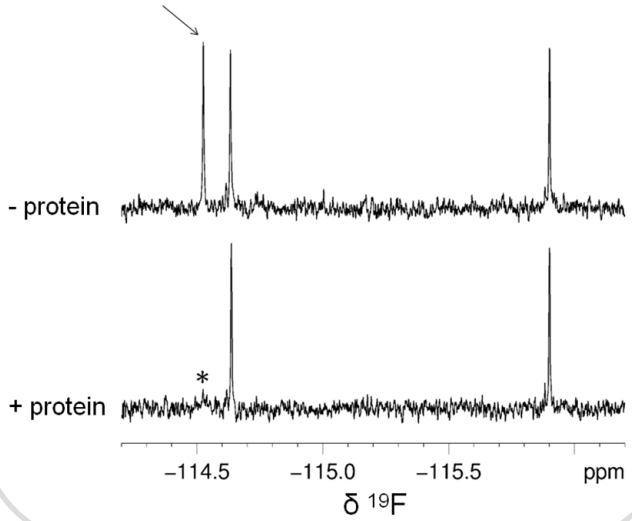
15 hits soaked in trypsin

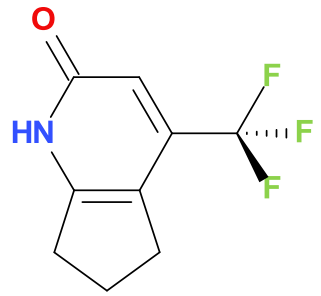
6 trypsin X-ray solved



(1)

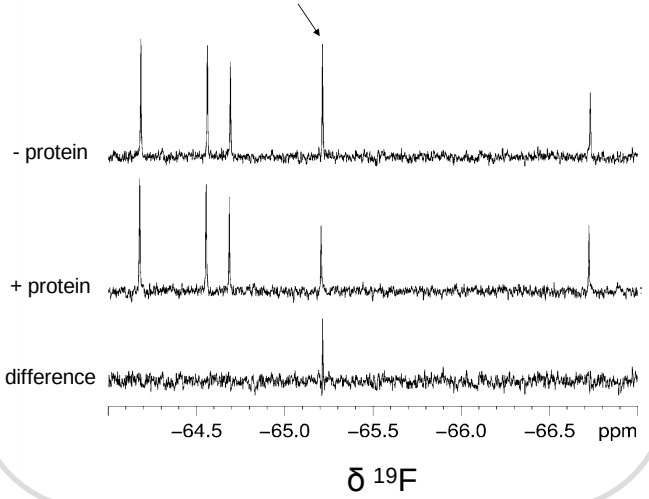
IC_{50} = 8.6 µM





(2)

0% Inhibition at 1mM



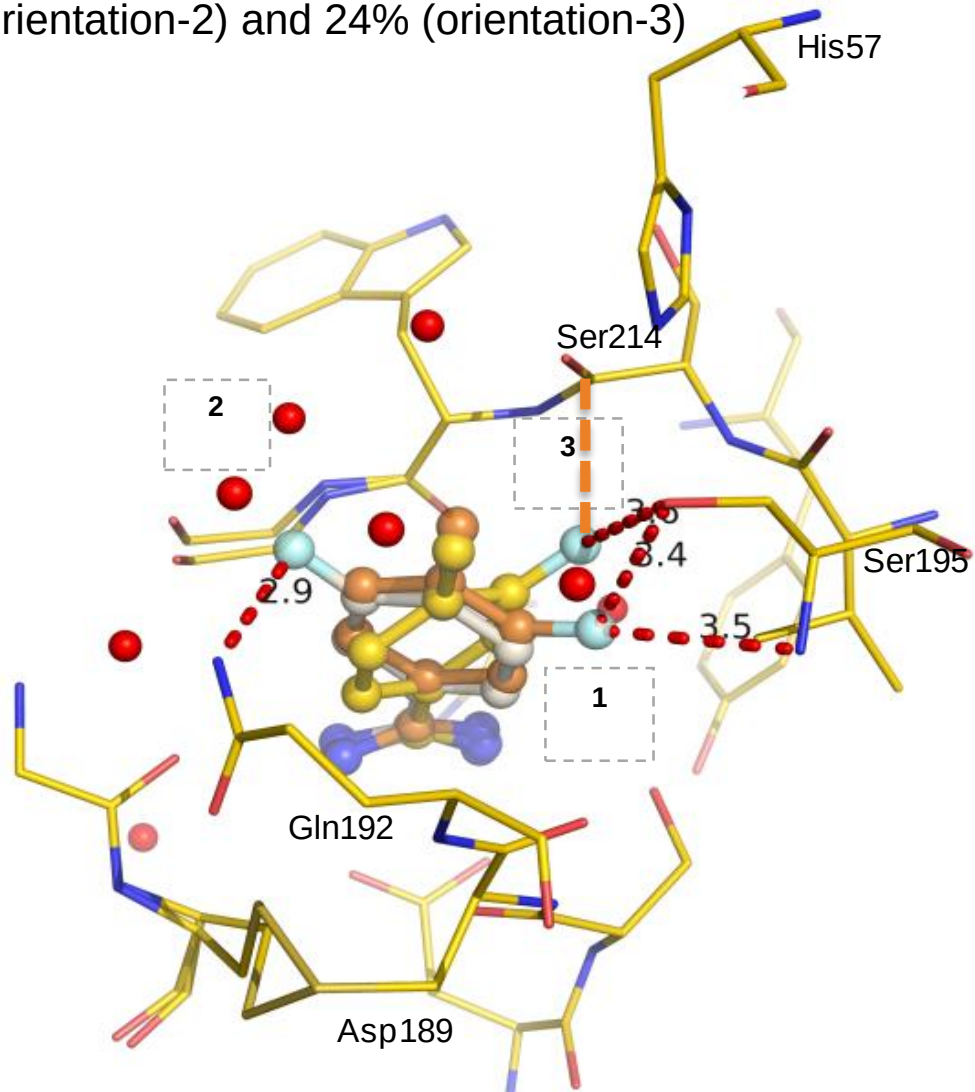
Vulpetti A., Schiering N., Dalvit C. Proteins: Structure, Function, and Bioinformatic Journal (2010) - DOI: 10.1002/prot.22836

NOVARTIS

Compound (1)

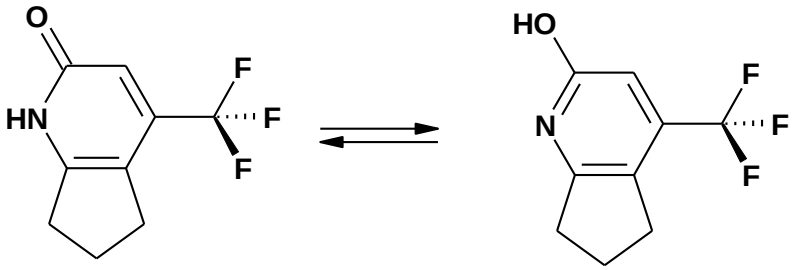
IC_{50} = 8.6 µM

- Compound (**1**) is modeled in one major and two minor orientations of the ring with occupancies of 54% (orientation-1), 22% (orientation-2) and 24% (orientation-3)
- Orientation-1 *(orange):*
  - the fluorine points towards *Ser195 (OG and NH)*
- Orientation-2 (*white):*
  - the aromatic ring is flipped by ca. 180º and the fluorine is pointing in the opposite direction, at a distance of ca. 3 Å to the *Gln192 side chain NE2*.
- Orientation-3 *(yellow)*:
  - At 3.6 Å to *Ser195 OG atom*, at 3.2 Å to the carbon of the *carbonyl of Ser214*, (α_1 = C-F…C of 115.5º ; angle α_2 = F… C=O of 69.9º (favorable multipolar interaction)

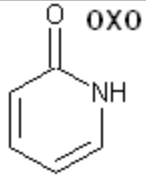
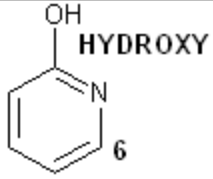


Compound (2)

- Compound (2) can be drawn in two tautomeric forms: 2-pyridone and 2-hydroxypyridine.



  - The 2-pyridone tautomeric form is estimated to be in general the preferred tautomer considering the occurrences of the two tautomeric forms in CSD and PDB databases.

Name	ΔG (kcal mol^{-1}) / medium / relevant form(s)	Major form in Water (M)			Minor form in Water (m)			PDB with identical HB score
		Structure/Name	CSD	PDB	Structure/Name	CSD	PDB	
2-pyridone	4.2/water/M; -0.65/gas (calc.)/m; ND/solid/m.		172	24		12	2	2

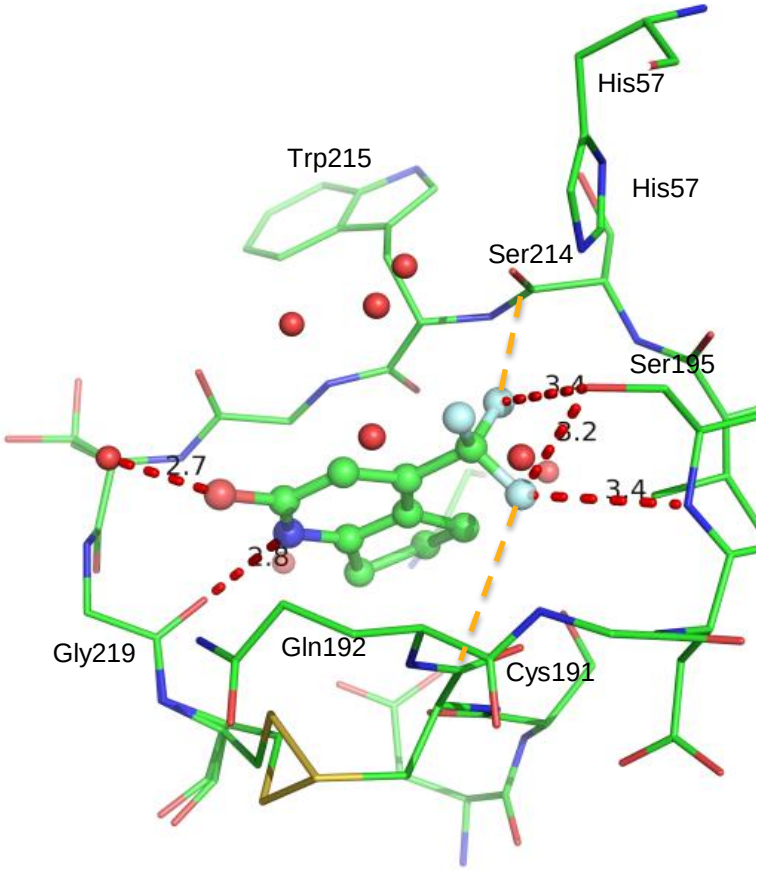
Milletti F. ,Vulpetti A. J.Chem.Inf.Model., 50 (2010) 1062-1074

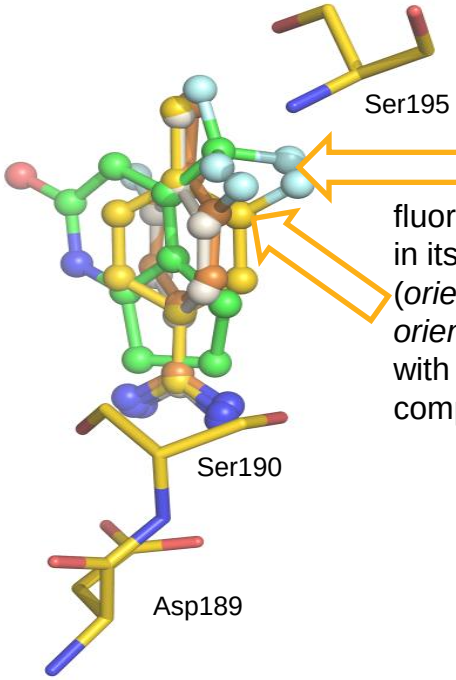
NOVARTIS

Compound (2)

0% Inhibition at 1 mM

- the trifluoromethyl group in contact with:
  Ser195 (OG and the NH);
  long-range contact (3.7 Å) to Cys191 C=O;
  3.2 Å to Ser214 C=O (α_1 = C-F…C of 168.4° , α_2 = F… C=O of 62.3°) (favorable multipolar interaction).



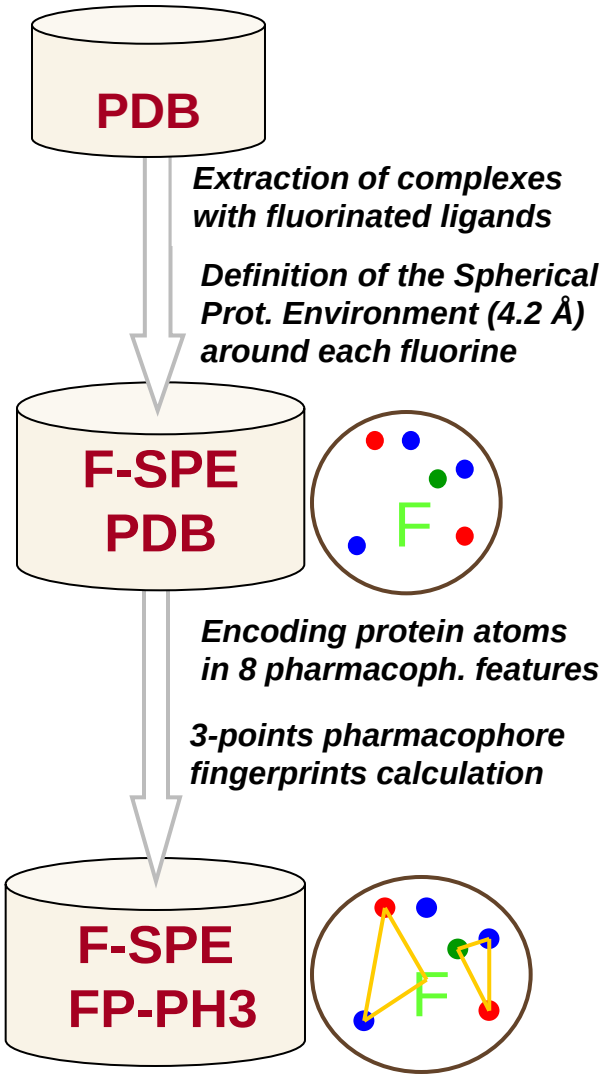


fluorines of compound (**1**) in its major orientation (*orientation-1*) and in *orientation-3* superimpose with two fluorines of compound (**2**)

NOVARTIS

F-SPE-PDB database

preparation



Protein sub-sites are extracted around each fluorine (F-SPE) contained in the PDB and compared to the query of interest by using a pharmacophoric description

Feature	Description	Description
D	donor attribute	NZ.LYS
CA	alpha carbon	CA
C=	carbon of carbonyl of amides / C guanidinium	C, CG.ASP, CG.ASN, CD.GLU, CD.GLN, CZ.ARG
P	donor and acceptor (-OH)	OG.SER, OG1.THR, OH.TYR, ND1.HIS, NE2.HIS
H	hydrophobe (-CH_2,-CH_3)	all aliphatic CB, CG, CD and SD.MET, SG.CYS
D=	π donor (e.g., sp^2 NH)	N, NE2.GLN, ND2.ASN, NH1.ARG, NH2.ARG, NE.ARG, NE1.TRP
A=	π acceptor (e.g., sp^2 O)	O, OD1.ASP, OD2.ASP, OE1.GLU, OE2.GLU
H=	π hydrophobe	all aromatic carbons of Phe, Tyr, Trp, His
X	Ligand fluorine atom	F

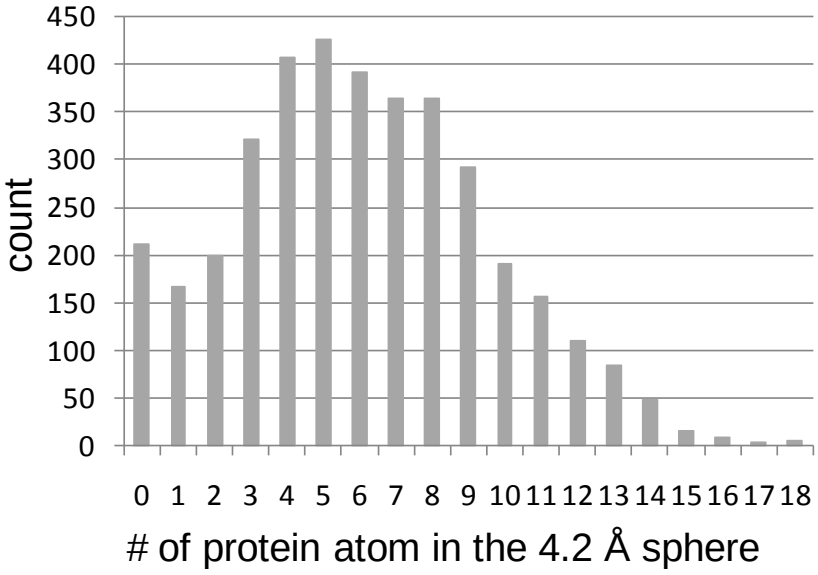
F-SPE-PDB database

details

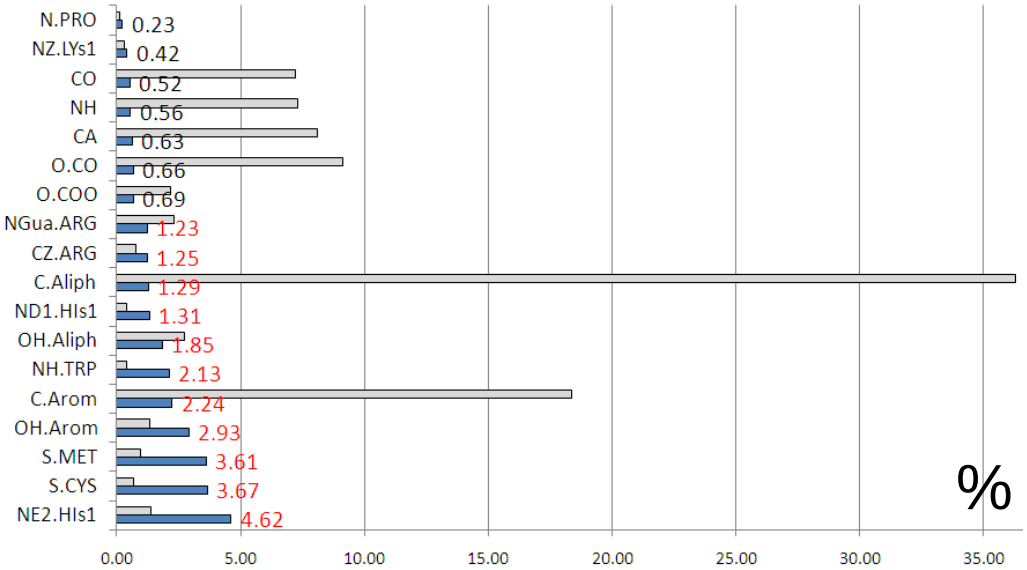
A. Number of complexes with 0 to 18 **protein atoms** in the sphere of radius 4.2 Å**.** [Ave= 6.14 , Std= 3.49]

B. Frequency / Relative Propensity of the **18 protein atom types**

A



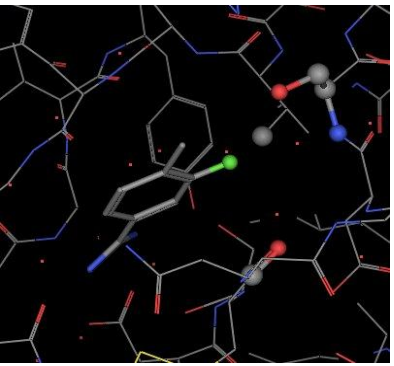
B



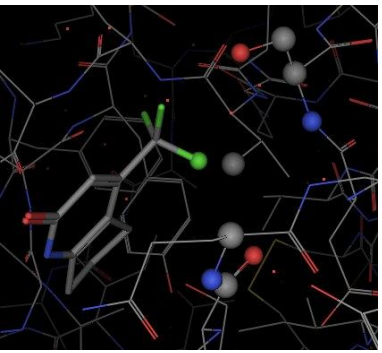
- The **number of triplets** is on average 39 with a standard deviation of 13
- **Clustering** : 1543, 2177 and 2414 clusters (Tanimoto coefficient 0.6, 0.7, 0.8)

Overlay of Compound (2) bound to Trypsin with TFE and Inhibitors bound to Elastase

Query



*Compound (**1**)*



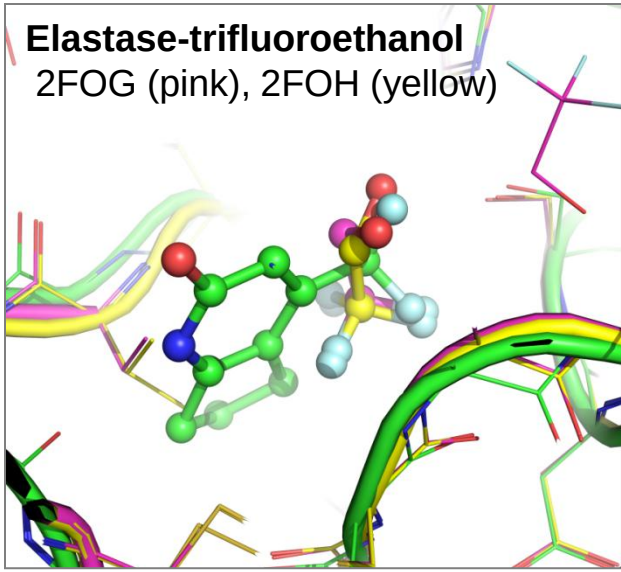
*Compound (**2**)*

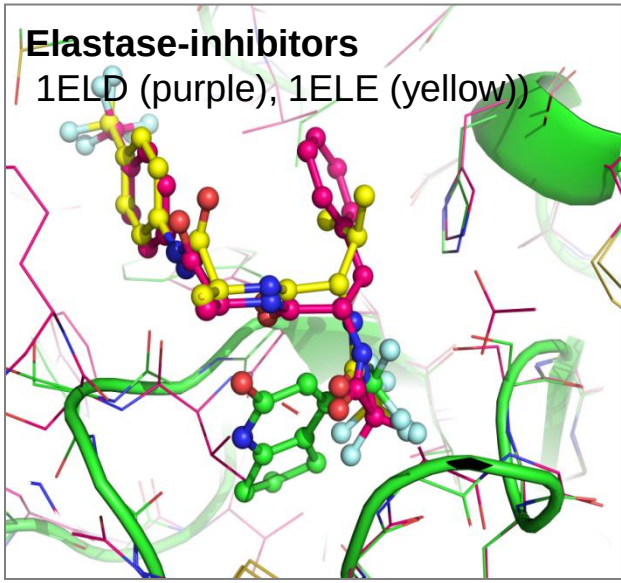
Database

F-SPE
FP-PH3

Similarity Search

Proteins with Similar
3D F-Spherical Protein Environment





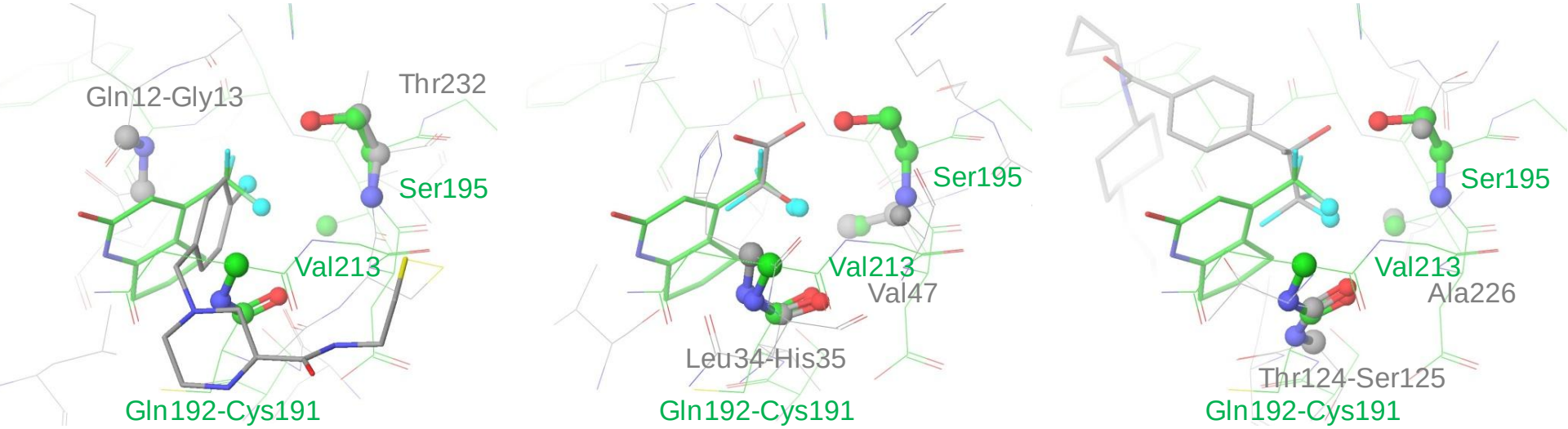
Overlay of Compound (2) bound to Trypsin with Ligands bound to different Proteins

A. 2ZJJ (BACE1),

B. 3CEE (oxylipin-conjugated barley lipid transferase protein, LTP1) and

C. 3D3E (11beta-hydroxysteroid dehydrogenase type 1, 11beta-HSD1).

A B C



Despite their different global fold and their different active binding sites these proteins have an environment around the ligand fluorine atoms that is quite similar

NOVARTIS

Summary

- **^{19}F-NMR screening of LEF**
  - Identification of useful starting fragments for fragment-based drug discovery projects
- **Characterization of Fluorophilic Protein Environments** (synergistic combination of ^{19}F-NMR, X-Ray, CADD)
  - Generation of the F-SPE-PDB database
  - Similarity searches into the F-SPE-PDB database
    - Identify proteins accommodating fluorine atoms with a similar pattern recognition.
    - Virtual F-scan.

Acknowledgments

NMR: **Claudio Dalvit,** Paul Erbel, Simon Rüdisser

XRay: **Nikolaus Schiering**, Alan D' Arcy, Frederic Villard

CADD: Greg Landrum, Francesca Milletti, Guido Kirsten (CCG)

Richard Lewis & Ulrich Hommel