J|A|C|S ARTICLES

Published on Web 08/25/2009

Design and NMR-Based Screening of LEF, a Library of Chemical Fragments with Different Local Environment of Fluorine

Anna Vulpetti,*,§ Ulrich Hommel,§ Gregory Landrum,§ Richard Lewis,§ and Claudio Dalvit*,§,#

Novartis Institute for Biomedical Research, Novartis Pharma AG, CH-4002 Basel, Switzerland, Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy

Received June 25, 2009; E-mail: anna.vulpetti@novartis.com; claudio.dalvit@iit.it

Abstract: A novel strategy for the design of a fluorinated fragment library that takes into account the local environment of fluorine is described. The procedure, based on a fluorine fingerprints descriptor, and the criteria used in the design, selection, and construction of the library are presented. The library, named LEF (Local Environment of Fluorine), combined with ^{19}F NMR ligand-based screening experiments represents an efficient and sensitive approach for the initial fragment identification within a fragment-based drug discovery project and for probing the presence of fluorophilic protein environments. Proper setup of the method, according to described theoretical simulations, allows the detection of very weak-affinity ligands and the detection of multiple ligands present within the same tested mixture, thus capturing all the potential fragments interacting with the receptor. These NMR hits are then used in the FAXS experiments for the fragment optimization process and for the follow-up screening aimed at identifying other chemical scaffolds relevant for the binding to the receptor.

Introduction

Fragment-based drug discovery (FBDD) represents an efficient approach to lead identification and optimization. FBDD is now widely recognized for its impact on drug discovery projects and is strongly endorsed by industry as well as academia.[1−7] In the initial phase of this approach a small (few thousands) collection of tiny and simple (low-complexity) compounds, usually consisting of no more than one or two rings with a few substituents, is screened against the macromolecular target of interest.

The small collection of fragments covers a much greater proportion of all of the possible compounds that could exist, termed "chemical space", than large (million) corporate compound collections for high-throughput screening. Better coverage of chemical space derives from the fact that the number of possible compounds that can, in principle, be constructed is a very steep exponential function of the number of atoms from which they are constituted.[8]

The design of these small libraries is crucial for the success of the approach. The physicochemical properties of fragments to be included, the aqueous solubility and purity, an assessment of molecular diversity, the chemical tractability of the fragments for follow-up, which chemical functionalities are disallowed, the drug-likeness of the fragments, and sampling of privileged medicinal chemistry scaffolds are some of the considerations that have been taken into account in the library design.[9−18]

Owing to their small size and limited functionality, fragments typically interact weakly with the protein and, as a result, display equilibrium dissociation binding constants in the range $10^{-4}-10^{-2}$ M.

§ Novartis Institute for Biomedical Research.
# Italian Institute of Technology.

(1) (a) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531–1534. (b) Hajduk, P. J.; et al. *J. Am. Chem. Soc.* **1997**, *119*, 5818–5827.
(2) Jahnke, W., Erlanson, D. A., Eds. *Fragment-based approaches in drug discovery*; Wiley-VCH: Weinheim, Germany, 2006.
(3) Jhoti H., Leach, A. R., Eds. *Structure-based Drug Discovery*; Springer: The Netherlands, 2007.
(4) Hajduk, P. J.; Greer, J. *Nat. Rev. Drug Discov.* **2007**, *6*, 211–219.
(5) Albert, J. S.; Blomberg, N.; Breeze, A. L.; Brown, A. J.; Burrows, J. N.; Edwards, P. D.; Folmer, R. H. A.; Geschwindner, S.; Griffen, E. J.; Kenny, P. W.; Nowak, T.; Olsson, L. L.; Sanganee, H.; Shapiro, A. B. *Curr. Top. Med. Chem.* **2007**, *7*, 1600–1629.
(6) Zartler, E. R., Shapiro, M. J., Eds. *Fragment-Based Drug Discovery*; John Wiley & Sons, Ltd.: New York, 2008.
(7) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. *J. Med. Chem.* **2008**, *51*, 3661–80.
(8) (a) Siegal, G.; AB, E.; Schultz, J. *Drug Discovery Today* **2007**, *12*, 1032–1039. (b) Fink, T.; Reymond, J. L. *J. Chem. Inf. Model.* **2007**, *47*, 342–353.
(9) (a) Lepre, C. A. *Drug Discovery Today* **2001**, *6*, 133–140. (b) Lepre, C. A.; Moore, J. M.; Peng, J. W. *Chem. Rev.* **2004**, *104*, 3641–3676.
(10) Erlasson, D. A.; Wells, J. A.; Braisted, A. C. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 199–223.
(11) Baurin, N.; Aboul-Ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2157–2166.
(12) Schuffenhauer, A.; Ruedisser, S.; Marzinizik, A. L.; Jahnke, W.; Blommers, M.; Selzer, P.; Jacoby, E. *Curr. Top. Med. Chem.* **2005**, *5*, 751–762.
(13) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. *J. Med. Chem.* **2005**, *48*, 403–413.
(14) Card, G. L.; Blasdel, L.; England, B. P.; Zhang, C.; Suzuki, Y.; Gillette, S.; Fong, D.; Ibrahim, P. N.; Artis, D. R.; Bollag, G.; Milburn, M. V.; Kim, S. H.; Schlessinger, J.; Zhang, K. Y. *Nat. Biotechnol.* **2005**, *23*, 201–207.
(15) Davies, T. G.; van Montfort, R. L. M.; Williams G.; Jhoti, H. In *Fragment-Based Approaches in Drug Discovery*; Jahnke, W.,Erlanson, D. A., Eds.; Wiley-VCH: Weinheim, 2006; pp 193−214.
(16) Petros, A. M.; et al. *J. Med. Chem.* **2006**, *49*, 656–663.
(17) Blaney, J.; Nienaber, V.; Burley, S. K. In *Fragment-Based Approaches in Drug Discovery*; Jahnke, W., Erlanson, D. A., Eds.; Wiley: Weinheim, 2006; pp 215−248.

 © 2009 American Chemical Society

NMR has been extensively used in the initial phase of FBDD projects for the identification of fragments that interact with the receptor.[19−31] Although NMR has an intrinsic low sensitivity, thus allowing only a limited throughput, it has one of the largest relative sensitivity to protein binding. This is due to the large dynamic range, defined as the difference of the NMR measured response in the free and protein-bound states. ^{19}F R_2 filter NMR experiments are among the most sensitive techniques for binding detection,[32−35] and a theoretical explanation of this effect was recently given.[36]

In this approach, a library of fluorinated fragments is first screened, and the identified binders are then used as reporters for the FAXS (fluorine chemical shift anisotropy and exchange for screening) experiments and for measuring the binding constant of the molecules binding to the receptor.[32,36]

A novel strategy for the design of a fluorinated fragment library that takes into account the local environment of fluorine is described. In addition, both a theoretical evaluation and an experimental application of the novel library to ^{19}F NMR-based fragment screening of mixtures are presented.

Results and Discussions

Why Fluorine Is Used in Medicinal Chemistry. Introduction of fluorine in the lead-optimization phase of drug discovery programs has become a common strategy. Historically this approach aimed mainly at improving physicochemical and absorption, distribution, metabolism, and excretion (ADME) properties of the lead compound. In fact, the substitution of hydrogen by fluorine can cause a change of physicochemical properties, such as pK_a[37] and lipophilicity,[38] and thus modify the membrane permeability[39] and bioavailability of a compound. Metabolic stability can also be favorably affected, as the site of metabolism (via oxidative mechanism) can be often blocked by the introduction of a fluorine atom. This well-established strategy for modulating pharmacokinetic properties has resulted in a large number of fluorinated drugs in clinical use.[40] Approximately 20% of MDDR (v. 2009.01) entries contain fluorine atom(s).[41]

A single hydrogen/fluorine exchange in the molecule can also have a large effect on the protein−ligand binding affinity as a direct consequence of specific interactions of fluorine atom(s) with the protein environment or as an indirect consequence of the change of the conformational preference of the compound. This is due to the different size and stereoelectronic effect of fluorine vs hydrogen. Favorable interactions of C−F moieties with H-bond donors have been reported in literature.[42] Favorable interactions can also be formed between C−F and lipophilic side chains. Moreover, a detailed analysis of binding data and crystal structures of fluorinated thrombin inhibitors in which a fluorine atom was systematically placed at different positions (also known as "F-scan") revealed the evidence of orthogonal favorable C−F···C═O and C−F···H−C_α interactions.[43] An outstanding review summarizing the influence of fluorine substituents as drug component can be found in the paper published by Müller at al.[44] The advantage of fluorine in drug design has stimulated the development of synthetic procedures[45] for the introduction of fluorine into organic molecules and has contributed to the increase in the number of fluorinated compounds and building blocks in the ACD[46] (v. 2009.01, ca. 15%).

A Novel Fluorine Library for NMR Screening. In this paper, the process followed in generating a fragment library of molecules with different Local Environment of Fluorine (LEF library) is described. The local environment of fluorine concept, recurring often in this work, represents the basis for the design of the library.

Based on the literature data reported above and in-house findings, the experimental identification of a potential fluorine binding site on a protein surface (also known as "fluorophilic

(18) Leach, A. R.; Hann, M. M.; Burrows, J. N.; Griffen, E. J. *Mol. BioSyst.* **2006**, *2*, 429–446.
(19) (a) Stockman, B. J. *Prog. NMR Spectrosc.* **1998**, *33*, 109–151. (b) Stockman, B. J.; Dalvit, C. *Prog. NMR Spectrosc.* **2002**, *41*, 187–231.
(20) Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Q. Rev. Biophys.* **1999**, *32*, 211–240.
(21) (a) Moore, J. M. *Curr. Opin. Biotechnol.* **1999**, *10*, 54–58. (b) Peng, J. W.; Lepre, C. A.; Fejzo, J.; Abdul-Manan, N.; Moore, J. M. *Methods Enzymol.* **2001**, *338*, 202–230. (c) Peng, J. W.; Moore, J.; Abdul-Manan, N. *Prog. NMR Spectrosc.* **2004**, *44*, 225–256.
(22) (a) Diercks, T.; Coles, M.; Kessler, H. *Curr. Opin. Chem. Biol.* **2001**, *5*, 285–291. (b) Coles, M.; Heller, M.; Kessler, H. *Drug Discovery Today* **2003**, *8*, 803–810.
(23) Van Dongen, M.; Weigelt, J.; Uppenberg, J.; Schultz, J.; Wikström, M. *Drug Discovery Today* **2002**, *7*, 471–478.
(24) Wyss, D. F.; McCoy, M. A.; Senior, M. M. *Curr. Opin. Drug Discovery Develop.* **2002**, *5*, 630–647.
(25) (a) Pellecchia, M.; Sem, D. S.; Wüthrich, K. *Nat. Rev. Drug Discovery* **2002**, *1*, 211–219. (b) Pellecchia, M.; Becattini, B.; Crowell, K. J.; Fattorusso, R.; Forino, M.; Fragai, M.; Jung, D.; Mustelin, T.; Tautz, L. *Exp. Opin. Ther. Targets* **2004**, *8*, 597–611.
(26) Meyer, B.; Peters, T. *Angew. Chem., Int. Ed.* **2003**, *42*, 864–890.
(27) Zartler, E. R.; Yan, J.; Mo, H.; Kline, A. D.; Shapiro, M. J. *Curr. Top. Med. Chem.* **2003**, *3*, 25–37.
(28) Salvatella, X.; Giralt, E. *Chem. Soc. Rev.* **2003**, *32*, 365–372.
(29) (a) Jahnke, W.; Widmer, H. *Cell. Mol. Life Sci.* **2004**, *61*, 580–599. (b) Fernandez, C.; Jahnke, W. *Drug Discovery Today: Technologies* **2004**, *1*, 277–283.
(30) Ludwiczek, M. L.; Baminger, B.; Konrat, R. *J. Am. Chem. Soc.* **2004**, *126*, 1636–1637.
(31) Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G. *Nat. Rev. Drug Discovery* **2008**, *7*, 738–745.
(32) (a) Dalvit, C.; Flocco, M.; Veronesi, M.; Stockman, B. J. *Comb. Chem. High Throughput Screening* **2002**, *5*, 605–611. (b) Dalvit, C.; Fagerness, P. E.; Hadden, D. T. A.; Sarver, R. W.; Stockman, B. J. *J. Am. Chem. Soc.* **2003**, *125*, 7696–7703.
(33) Tengel, T.; Fex, T.; Emtenas, H.; Almqvist, F.; Sethson, I.; Kihlberg, *Org. Biomol. Chem.* **2004**, *2*, 725–731.
(34) Klages, J.; Coles, M.; Kessler, H. In *Exploiting Chemical Diversity for Drug Discovery*; Bartlett, P. A., Etzeroth, M., Eds.; RSC Publishing: Cambridge, UK, 2006; Vol. 12, pp 263−290.
(35) Poppe, L.; Harvey, T. S.; Mohr, C.; Zondlo, J.; Tegley, C. M.; Nuanmanee, O.; Cheetham, J. *J. Biomol. Screening* **2007**, *12*, 301–311.
(36) (a) Dalvit, C. *Prog. NMR Spectrosc.* **2007**, *51*, 243–271. (b) Dalvit, C. *Concepts Magn. Reson. Part A* **2008**, *32A*, 341–372.
(37) Morgenthaler, M.; Schweizer, E.; Hoffmann-Röder, A.; Benini, F.; Martin, R, E,; Jaeschke, G.; Wagner, B.; Fischer, H.; Bendels, S.; Zimmerli, D.; Schneider, J.; Diederich, F.; Kansy, M.; Müller, K. *ChemMedChem* **2007**, *2*, 1100–1015.
(38) Leroux, F.; Jeschke, P.; Schlosser, M. *Chem. Rev.* **2005**, *105*, 827–856.
(39) Gerebtzoff, G.; Li-Blatter, X.; Fischer, H.; Frentzel, A.; Seelig, A. *ChemBioChem* **2004**, *5*, 676–684.
(40) Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V. *Chem. Soc. Rev.* **2008**, *37*, 320–330.
(41) MDDR (MDL Drug Data Report) data set is available from Symyx Technologies, Inc.
(42) (a) Kashima, A.; Inoue, Y.; Sugio, S.; Maeda, I.; Nose, T.; Shimohigashi, Y. *Eur. J. Biochem.* **1998**, *255*, 12–23. (b) Carosati, E.; Sciabola, S.; Cruciani, G. *J. Med. Chem.* **2004**, *47*, 5114–5125.
(43) Schweizer, E.; Hoffmann-Röder, A.; Schärer, K.; Olsen, J. A.; Fäh, C.; Seiler, P.; Obst-Sander, U.; Wagner, B.; Kansy, M.; Diederich, F. *ChemMedChem.* **2006**, *1*, 611–621.
(44) Müller, K.; Faeh, C.; Diederich, F. *Science* **2007**, *317*, 1881–1886.
(45) Kirk, K. *Org. Process Res. Dev.* **2008**, *12*, 305–321.
(46) ACD (Available Chemicals Directory) data set is available from Symyx Technologies, Inc.

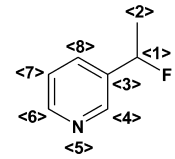
protein environment") is an important finding for drug design. The local environment around the fluorine atom plays an important role for presenting the fluorine moiety in the proper conformation into the fluorophilic protein environment.

Also, the ^{19}F isotropic chemical shift σ_{iso}, and especially the magnitude of the principal components of the chemical shift tensor σ_{11}, σ_{22}, and σ_{33}, are determined by the local atomic environment in which the fluorine is embedded. Small substitution patterns close to the fluorine atom have a great influence on the ^{19}F chemical shift. For example, the isotropic chemical shifts for a series of 13 fluorobenzenes with different substitutions cover a 63 ppm chemical shift range, and the shielding tensor elements (derived from solid-state NMR) cover a 237 ppm range.[47] The isotropic ($\sigma_{iso} = (\sigma_{11} + \sigma_{22} + \sigma_{33})/3$) and anisotropic ($\Delta\sigma = \sigma_{33} - (\sigma_{11} + \sigma_{22})/2$) components of the ^{19}F chemical shift tensor are responsible for the high sensitivity of the FAXS experiments.[36]

The ^{19}F NMR screening of small fluorinated molecules represents a powerful technique for studying the weak interactions involving fluorine atoms and for identifying binders with fluorine(s) in their pharmacophore. Once a binder is detected, the role of the fluorine atom(s) in the binding can be assessed by testing available close analogues (i.e., SAR by archive), with or without the fluorine moiety, before the crystal structure determination.

Source of Fragments. A database of fluorinated fragments containing a single CF_3 or CF moiety was compiled from the Novartis compound collection. Molecules with more than one nonequivalent fluorine atom were not selected in order to keep the complexity of the ^{19}F fluorine spectra low (one ^{19}F singlet resonance for each molecule) and to avoid signal intensity reduction due to the resonance splitting originating from $^{19}F-^{19}F$ short- or long-range homonuclear scalar couplings. In addition, molecules with multiple fluorine atoms could have low solubility in an aqueous environment. Experimental solubility data have shown that CF_3 and OCF_3 substituents always increase lipophilicity, while a single fluorine atom can alter this property in either direction.[48] If the fluorine is in close proximity to an oxygen atom (hydroxyl, alkoxy, carbonyl oxygen atom), it lowers lipophilicity (likely by enhancing the solvation energy in water more than in 1-octanol). Conversely, a fluorine atom placed near a basic nitrogen center increases the lipophilicity (by reducing pK_a and thus increasing log D).

The CF/CF_3 molecules collected from an internal archive were in large part proprietary (~80%), with the remaining ~20% being acquired from various commercial vendors. A number of filters were applied to this set of compounds to identify small fragments with desirable properties. We focused our selection on molecules with a molecular weight between 140 and 300 Da, A log P[49] (the implementation of log P calculation in Scitegic's Pipeline Pilot[50]) ≤ 3, number of rotatable bonds ≤ 4, number of any type of rings ≤ 4, and with at least a total number of H-bond acceptors/donors ≥ 3. We did not consider fragments with unknown stereochemistry. The resulting fragments were then filtered to ensure the absence of reactive and other undesirable chemical functionalities by means of multiple substructure searches.



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

Figure 1. (Top) Sample molecule with atoms numbered using angle brackets. (Bottom) Set of all paths of length one to five bonds rooted at the fluorine atom.

The same procedure applied to the ACD collection of commercial compounds identified the presence of about 22 500 fluorinated fragments: 78% of these contain one CF and 22% contain one CF_3. These fragments are mainly aromatic, with the CF/CF_3 moieties directly attached to aromatic rings of various nature (ca. 97%). Very few CF_3-containing fragments have the CF_3 bound to oxygen, sulfur, nitrogen, or carbonyl (ca. 5% of the CF_3 fragments).

Fluorine Local-Environment Fingerprints. All unique in-house and commercial fluorinated fragments were combined and clustered using a new type of fingerprint to characterize the local chemical environment around the F atoms and CF_3 groups. These fluorine environment fingerprints were derived from the familiar topological torsion descriptors published by Nilakantan et al.[51] Topological torsion fingerprints are based on hashing and counting all paths of three consecutive bonds (four neighboring atoms) in a molecule. Atoms are characterized using the atom-typing scheme developed for atom-pair fingerprints;[52] an atom's type is determined by its atomic number, number of π electrons, and number of heavy-atom neighbors (not counting those in the torsion). This gives atom types of the form C-0-1 (carbon, no π electrons, one non-hydrogen neighbor not in the torsion); a topological torsion is then a sequence of four atom types, e.g., (C-0-1, C-0-0, C-1-1, C-1-1). Our fluorine environment fingerprints differ from standard topological torsions in that they include paths consisting of between one and five bonds and only include paths that start from the fluorine atom or CF_3 moiety.

The construction of this easily interpretable fingerprint is illustrated in Figure 1. Each path of one to five bonds starting at the F atom is enumerated, atom types are generated for the atoms in the path, and these atom-typed paths are hashed to generate integer bit id's. The bit id's themselves, which can be quite large, are schematically indicated in Figure 1 as b1, b2, ..., b8. Notice that both of the length 3 paths generate the same bit id, b4; in the resulting fingerprint, b4 would thus have a count of two, while the other bits would have a count of one.

(47) (a) Mehring, M. *High Resolution NMR Spectroscopy in Solids*; Springer-Verlag: Berlin, 1976. (b) Raber, H.; Mehring, M. *Chem. Phys.* **1977**, *26*, 123–130.
(48) Böhm, H.-J.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Müller, K.; Obst-Sander, U.; Stahl, M. *ChemBioChem* **2004**, *5*, 637–643.
(49) Ghose, A. K.; Viswanadhan, V. N.; Wendolowski, J. J. *J. Phys. Chem.* **1998**, *102*, 3762–3772.
(50) http://accelrys.com/products/scitegic/.
(51) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 82–85.
(52) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.

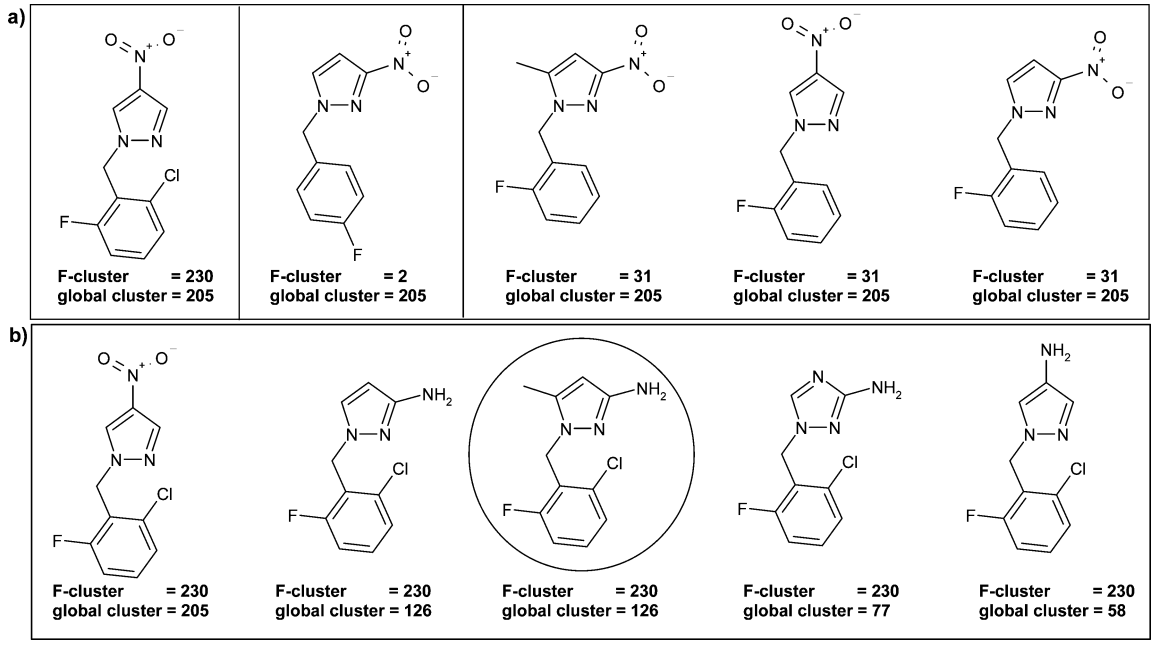


Figure 2. Fragments taken from Fluorochem supplier clustered by two different descriptions. (a) The five molecules belong to the same global cluster number 205, but to different F-clusters generated using the fluorine local environment fingerprints. (b) The five molecules belong to the same F-cluster number 230, but to different global clusters generated using the whole molecular structure. The molecule in the circle corresponds to the centroid of the global cluster 126.

The fluorine environments were clustered using the Butina clustering algorithm.[53] Similarity was calculated using the Dice metric, as is usual for topological torsions. The Dice similarity between two counts-based fingerprints FP_i and FP_j is defined as

$$\mathrm{Sim}(FP_i, FP_j) = \frac{2\sum_b \min(FP_{ib}, FP_{jb})}{\sum_b FP_{ib} + \sum_b FP_{jb}} \quad (1)$$

where FP_{ib} is the count for bit b in FP_i; FP_{jb} is the count for bit b in FP_j, and $\min(FP_{ib},FP_{jb})$ is the lower count for bit b in the two fingerprints.

By using a Dice similarity threshold of 0.9, 3821 F-clusters (i.e., characterized by a different local chemical environment around the F atom) and 1825 CF_3-clusters (characterized by a different local chemical environment around the CF_3 moiety) were obtained. For the initial proprietary LEF library, the selection was done from the set of in-house fragments. The large number of commercial fragments are considered for hit follow-up and future expansion of the LEF library. Generation of the local fingerprints and subsequent clustering were carried out using the open-source cheminformatics toolkit RDKit.[54]

Selection Process. The resulting in-house sets for CF and CF_3 fragments, already clustered using the *local* fluorine environment fingerprints, were also clustered on the basis of the whole molecular structure (*global* description). The clustering based on the whole molecule was performed using FCFP-4 fingerprints as descriptors and the Tanimoto coefficient as the metric of similarity (Scitegic Pipeline Pilot[50]). Each compound was thus assigned to membership in two clusters: one based on the *global* and one based on the *local* description. The two criteria for grouping molecules are very different, as illustrated in Figure 2.

Figure 2a shows five molecules that cluster together using the whole-molecule description (*global* cluster number 205). However, the five molecules do not belong to the same cluster based on *local* environment around the fluorine atom. The five molecules are divided among three *local* clusters (F-cluster labels). As expected, the fluorine fingerprint emphasizes the locally different topology around the fluorine. Each F-based cluster contains members which are dissimilar, considering the whole structure (i.e., they have different "global" cluster membership); see, e.g., F-cluster number 230 in Figure 2b.

As described in the Introduction, the library design process aims at generating a fragment library of molecules with different local environment of fluorine. For that reason, the selection of fragments was biased toward a maximal coverage of number of clusters generated using the fluorine fingerprints: to ensure the presence of each fluorine chemical environment in the library, at least one compound was selected from each F/CF_3-based cluster. In selecting compounds from each F/CF_3-based cluster, the goal was to cover as much of the chemical space as possible. In order to do so, the first priority was given to molecules that were cluster centroids in the global clustering (the circled molecule in Figure 2b). This ensures that global diversity is covered in the final selection. In addition to the centroids, the selection of the representatives of each F/CF_3-cluster was based on visual inspection of the molecules. In more detail, molecules with solubilizing moieties as well as fragments with functional groups amenable to chemical derivatization were prioritized if available. About 1400 fragments were selected, 45% of which contain one CF_3 and 55% contain one CF.

Characterization of the LEF library. The LEF library has a good coverage of the different fluorine chemical environments embedded in the available fluorinated fragments.

(53) Butina, D. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747–750.
(54) RDKit: Open-source cheminformatics; http://www.rdkit.org, accessed May 2009.

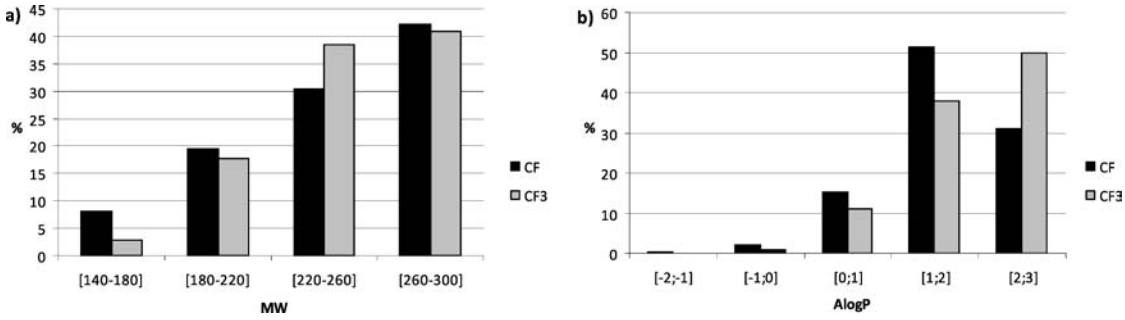


Figure 3. Histograms showing the distribution of molecular weight (a) and *A* log *P* (b) for the LEF library. The profiles of CF-containing fragments are shown by black bars and the profiles of CF_3-containing fragments by gray bars. As expected, the distribution of CF_3-containing fragments is shifted slightly toward the right-hand part of the *A* log *P* plot due to the increased lipophilicity of CF_3-containing molecules.

The presence in the fragment of synthetic handles and possible retrosynthetic routes is very important to facilitate the rapid synthesis of analogues. We define as "synthetically tractable" those fragments having one of the 16 chemical environments described by the BRICS approach (Breaking of Retrosynthetically Interesting Chemical Substructures),[55] which is an extension of the popular RECAP (Retrosynthetic Combinatorial Analysis Procedure)[56] approach. Moreover, the presence of a number of functional groups commonly used in organic synthesis (such as Cl, Br, I, OH, NH_2, COOH, and N═C═O) was also considered as favorable. BRICS decomposition and identification of other useful functional groups were carried out using the open-source cheminformatics toolkit RDKit.

Based on these criteria, the potential synthetic tractability of the fragments present in the LEF library is good: ca. 90% of the fragments fulfill one or both of the criteria. All other descriptors, such as MW, *A* log *P*, number of H-bond donors/acceptors, rotatable bonds, etc., were calculated with Scitegic's Pipeline Pilot. Figure 3 shows the profile of MW and *A* log *P* for the selected CF and CF_3 fragments.

Mixtures Generation. The molecules selected for the LEF library were dissolved in DMSO at a concentration of 100 mM using a Xiril robot. Subsequently, an equal aliquot from all these solutions was taken for generating mixtures consisting of 12 molecules each. Mixtures contained exclusively either CF_3 (54 mixtures) or CF (66 mixtures) molecules. This was necessary for two experimental reasons. First, the concentration at which the CF_3 molecules are tested is lower when compared to that of the CF molecules. Second, the large dispersion in chemical shift of the ^{19}F signals requires a different carrier frequency and spectral width for the acquisition of the CF_3 and CF signals. The concentration of the molecules in the mixtures stock solution is 8.33 mM. Each column of the 96-well rack of the mixtures corresponds to an entire 96-well rack of the single compounds. Molecules belonging to the same fluorine cluster were put in different mixtures in order to reduce the likelihood of signal overlap, thus allowing the creation of even larger mixtures.

NMR Screening. Molecules were tested at a concentration of 18 μM for the CF_3 mixtures and 35 μM for the CF mixtures. Typical ^{19}F spectra recorded with ^1H decoupling for a CF_3 (mixture 46) and a CF (mixture 93) mixture are shown respectively in Figure 4a,b.

The signals are well separated, and the dispersion in chemical shifts is 20.53 ppm for the CF_3 mixture and 32.58 ppm for the CF mixture. Despite the smaller chemical shift range for the CF_3 mixtures, resonances with very similar chemical shifts can easily be resolved because all the signals appear as sharp singlets, as shown in the expanded regions of Figure 4a.

Molecules of the same cluster typically have similar chemical shifts, as shown in Figure 5. However, significant chemical shift scatter is observed for a few clusters, as shown, for example, for cluster 9 in Figure 5. Cluster 9 contains para-substituted fluorobenzene. This scatter is likely due to long-range effects deriving from atoms located at a distance farther from the fluorine atom than the five bonds used for the fluorine descriptor. The clustering of molecules by considering path lengths larger than the five bonds would probably result in a smaller difference in chemical shifts for the molecules of the same cluster. However, the number of generated clusters would also dramatically increase. The use of extended fluorine environment fingerprints and/or other descriptors could find an important application as a tool for predicting the ^{19}F isotropic chemical shift. Molecules of different clusters do not have always different chemical shifts. This is evident in Figure 5, where the isotropic chemical shifts of cluster 5 are similar to those of clusters 6 and 7, and the chemical shifts of cluster 13 are similar to those of cluster 14. This is likely due to the fact that the observed isotropic chemical shift is the average of the three principal components of the chemical shift tensor, σ_{11}, σ_{22}, and σ_{33}. Molecules of different clusters could have different values for the components of the chemical shift tensor and nevertheless have a very similar isotropic chemical shift. Only solid-state NMR experiments and/or, when is feasible, liquid-state CSA/DD cross-correlation experiments performed on fluorinated molecules belonging to different ^{19}F-clusters, but with similar isotropic chemical shifts, could provide a definitive answer to these experimental observations.

The screening protocol requires the acquisition of ^{19}F R_2 filter experiments with the CPMG (Carr−Purcell−Meiboom−Gill)[57] scheme and with a long delay τ between the 180° pulses in order to take advantage of the contribution deriving from the exchange between free and bound states and from faster relaxation of antiphase ^{19}F single-quantum coherence (for the fluorine spins scalar-coupled to the protons).[36b] CPMG-based dispersion studies can be performed after the screening run with the identified hits. On the basis of the dispersion profiles (at different field strengths) it is possible, in favorable cases, to

(55) Degen, J; Wegscheid-Gerlach, C.; Zaliani, A.; Rarey, M. *ChemMedChem* **2008**, *3*, 1503–1507.
(56) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522.
(57) Meiboom, S.; Gill, D. *Rev. Sci. Instrum.* **1958**, *29*, 688.

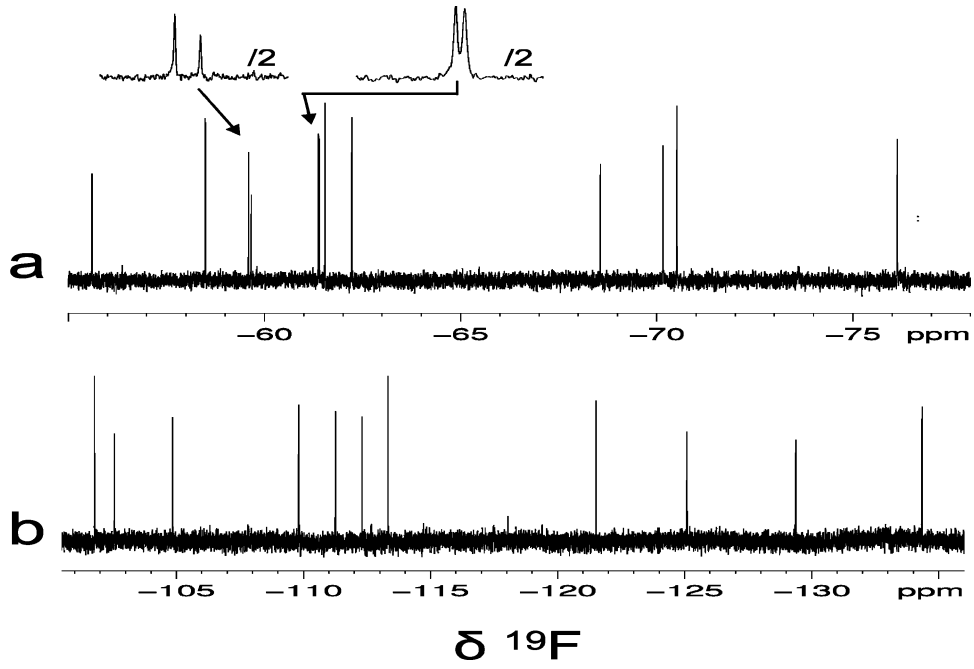


Figure 4. ^{19}F NMR spectra (564 MHz) recorded with proton decoupling of (a) mixture 46 (CF_3-containing molecules) and (b) mixture 93 (CF-containing molecules) of the LEF library. Close-ups for two narrow spectral regions of (a) indicated by the arrows are shown. These two expanded regions are plotted at half the vertical scale with respect to (a). A total of 80 (a) and 96 (b) scans were acquired with a repetition time of 3.8 s. The concentration of the molecules was 18 μM (a) and 35 μM (b).

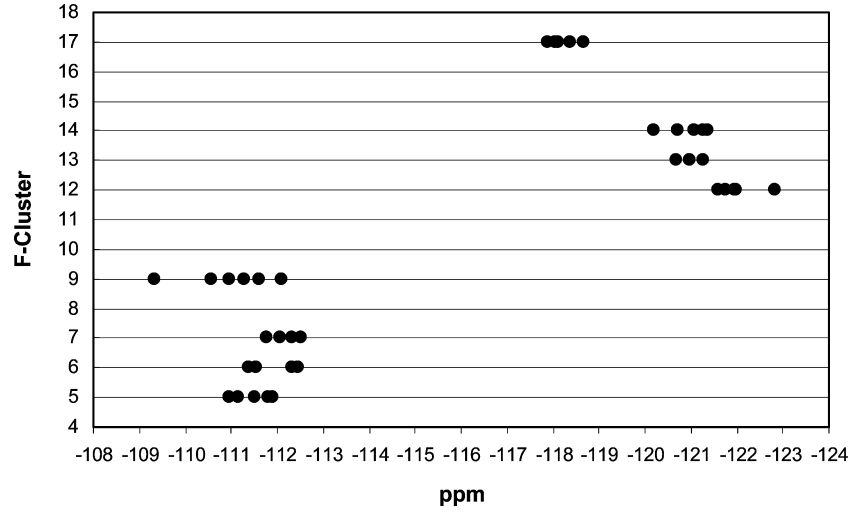


Figure 5. ^{19}F NMR experimental chemical shifts expressed in ppm (x-axis) vs F-cluster number generated using the fluorine local environment fingerprints (y-axis) on a subset of CF-containing fragments.

extract populations, changes in chemical shift, and lifetimes of bound states.

Typical screening results for two active mixtures, a CF and a CF_3 mixture of the LEF library, against the protein bovine trypsin are shown in Figure 6. The identification of the binding molecules in the mixture is a straightforward process: by comparing the spectra of the mixtures recorded in the absence and in the presence of the protein (respectively in the upper and lower traces of Figure 6), the signals of the molecules interacting with protein are significantly reduced in intensity in the presence of the protein. The ^{19}F chemical shift of each molecule is known, and therefore the perturbed signal is immediately associated with the molecule that interacts with the receptor. The large dispersion in ^{19}F chemical shift and the low concentration of the molecules used in the experiments allow the screening of even larger mixtures. This can be seen in Figure 7, where the results of screening a mixture of 36 different CF_3-containing fragments against trypsin are reported.

Despite the number of components in the mixture, the 36 signals are well resolved. Visual inspection allows the rapid identification of the signal(s) that are significantly perturbed in the presence of the protein. The signal at −61.79 ppm disappears upon addition of the protein, as shown in the expanded region of the spectrum. The chemical shift of this signal corresponds to the fragment 5-(trifluoromethyl)-2-pyridylamidine (MW = 189.14) (**1**), thus identifying the fragment as a ligand for trypsin.

The use of large mixtures and on-the-fly identification of the fragments interacting with the protein allows rapid screening

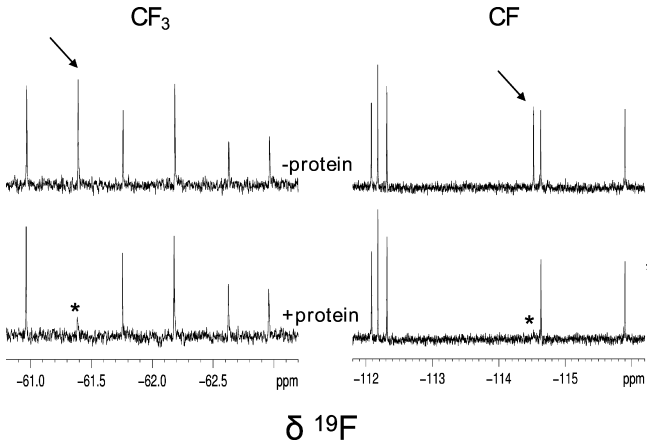
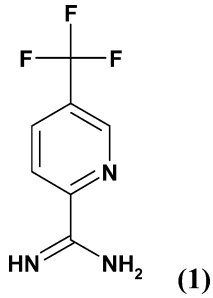


Figure 6. Expanded region of the ^{19}F NMR (564 MHz) R_2 filter spectra recorded with proton decoupling for (left) mixture 40 (CF_3-containing molecules) and (right) mixture 101 (CF-containing molecules) in the absence (upper traces) and in the presence (lower traces) of 9 μM bovine trypsin. The length of the R_2 filter was 0.4 s (left) and 0.32 s (right) with a τ period of 40 ms. A total of 80 scans (left) and 96 scans (right) were acquired with a repetition time of 3.8 s. The concentration of the molecules was 18 μM (left) and 35 μM (right). The arrows indicate two signals that are significantly reduced in intensity in the presence of the protein, and the asterisks indicate the position of the affected signals.



(1)

and data analysis. Fast recognition is possible even when the signal of the active ingredient of the mixture overlaps with other signals. In this case it is sufficient to test individually, in the absence and in the presence of the protein, only the molecules of the mixture with that same isotropic chemical shift. With our current setup, the screening of a 96-well rack of mixtures of the LEF library (12 compounds per mixture, corresponding to a total of 1152 molecules) requires about 24 h measuring time; that time can be reduced to 8 h if mixtures of 36 molecules are used. This time can be further reduced with the use of cryogenic probe technology optimized for ^{19}F detection.[58] It is evident that, with this ^{19}F NMR method, it is now possible to perform medium-throughput screening in a reasonable measuring time, despite the low intrinsic sensitivity of the NMR technique, and reduce the protein consumption for the screening.

After the completion of the screening run, a known ligand is added into the samples containing the active mixtures to validate the identified hits. In this particular case, the ligand was a potent inhibitor that binds to the S1 and S4 pockets of trypsin. This can be seen in Figure 8 for the 36-molecules mixture when a strong inhibitor of trypsin (K_D in the nanomolar range) is added to the sample in the presence of the protein.

The ^{19}F R_2 filter experiments are recorded with the same parameters to allow a direct comparison. The signal of **1**, absent

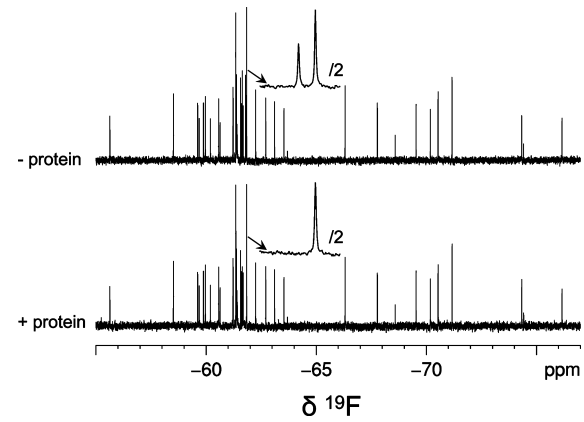


Figure 7. ^{19}F NMR (564 MHz) R_2 filter spectra recorded with proton decoupling for a mixture of 36 CF_3-containing molecules in the absence (upper trace) and in the presence (lower trace) of 9 μM bovine trypsin. The concentration of the molecules was 18 μM. The length of the R_2 filter was 0.4 s, with a τ period of 40 ms. A total of 160 scans were acquired with a repetition time of 3.8 s. A close-up of a narrow spectral region indicated by the arrow is shown for both spectra. The expanded regions are plotted at half the vertical scale with respect to the full spectra. The signal at −61.79 ppm, which originates from the CF_3 signal of 5-(trifluoromethyl)-2-pyridylamidine and that is visible in the close-up of the spectrum recorded in the absence of protein, is the only signal of the mixture which is significantly reduced in intensity in the presence of the protein. This signal disappears in the presence of the protein, as shown in the close-up of the lower trace.

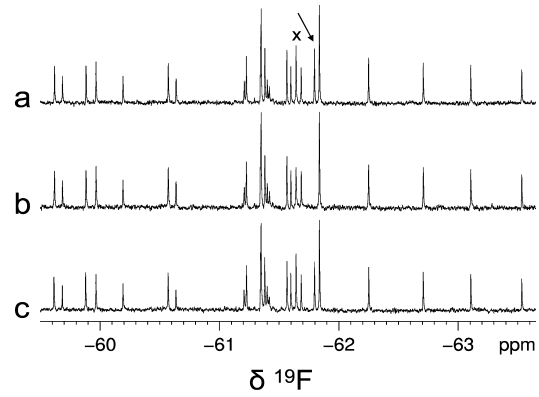


Figure 8. Expanded region of the ^{19}F NMR (564 MHz) R_2 filter spectra recorded with proton decoupling for a mixture of 36 CF_3-containing molecules in the absence of the protein (a), in the presence of 9 μM bovine trypsin (b), and in the presence of 9 μM bovine trypsin and a strong inhibitor (K_D in the nanomolar range) (c). The concentration of the molecules was 18 μM. The length of the R_2 filter was 0.4 s, with a τ period of 40 ms. A total of 160 scans were acquired with a repetition time of 3.8 s. The arrow indicates the CF_3 signal of **1**. In the presence of the protein, the signal is not visible, and it reappears at its full intensity in the presence of the inhibitor (c). In addition, the signal indicated by an X is slightly perturbed in the presence of the protein (b), and its intensity is completely restored in the presence of the inhibitor (c).

in the presence of the protein (Figure 8b), reappears in the spectrum upon addition of the known inihibitor, as shown in Figure 8c. This observation strongly suggests that **1** is a ligand for trypsin and furthermore that the molecule binds to the same site occupied by the inhibitor. The addition of a competitive compound also facilitates the detection of very weak ligands, as depicted in Figure 8. The resonance at −61.64 ppm, labeled with an X in Figure 8a, displays a minute signal intensity reduction in the presence of the protein (Figure 8b). Often this effect would be discarded; however, the intensity change between spectra a and b is meaningful because the original signal

(58) Kovacs, H.; Moskau, D.; Spraul, M. *Prog. NMR Spectrosc.* **2005**, *46*, 131–155.

intensity of the molecule is restored in the presence of the inhibitor (Figure 8c). This means that the peak labeled with an X must also be a binder.

The ^{19}F NMR screening of the LEF library performed at low concentration has five major beneficial features:

(i) As discussed above, it allows the testing of large compound mixtures, thus reducing measuring time and protein consumption and increasing throughput.

(ii) It permits the screening of fragments with limited solubility in water. Therefore, high water solubility of the fragment is not a requirement for the selection of the molecules of the LEF library. Some fragments have a high *A* log *P*, as depicted in Figure 3, and display an experimental solubility in aqueous solution less than 100 μM. Subsequent characterization of these ^{19}F NMR binding fragments with other biophysical techniques could be hampered by their low solubility. In this situation, the selection of very close analogues to the identified fragment, with or without a fluorine moiety, but with superior water solubility, is an approach that can be pursued.

(iii) The possibility of working at low fragment concentration improves the relative sensitivity to protein binding of the NMR-based assay. The transverse relaxation of the ^{19}F resonance of a fragment interacting weakly with the receptor is given by the equation[36,59,60]

$$R_{2,\mathrm{obs}} = p_\mathrm{f}R_{2,\mathrm{f}} + p_\mathrm{b}R_{2,\mathrm{b}} + p_\mathrm{b}p_\mathrm{f}^2\tau_\mathrm{res}4\pi^2(\delta_\mathrm{f} - \delta_\mathrm{b})^2 \qquad (2)$$

where $R_{2,\mathrm{f}}$ and $R_{2,\mathrm{b}}$ correspond to the ^{19}F transverse relaxation rates in the free and bound states, respectively. The term p_f is the fraction of free ligand, defined as $p_\mathrm{f} = 1 - ([\mathrm{EL}]/[\mathrm{L_T}])$, and p_b is the fraction of bound ligand, defined as $p_\mathrm{b} = [\mathrm{EL}]/[\mathrm{L_T}]$. $[\mathrm{L_T}]$ and [EL] are the total and protein-bound ligand concentrations, respectively. The last term in eq 2 is the exchange contribution, originating from the difference between the isotropic chemical shifts δ_f and δ_b of the ^{19}F fragment resonance in the free and bound states, respectively. The term is directly proportional to the residence time of the fragment on the protein, τ_res.

Note that the exchange contribution is weighted by $p_\mathrm{b}p_\mathrm{f}^2$ when formulated with τ_res, and by $p_\mathrm{b}p_\mathrm{f}$ when formulated with τ_ex. τ_ex is related to τ_res according to the expression $\tau_\mathrm{ex} = 1/([\mathrm{E}]k_\mathrm{on} + 1/\tau_\mathrm{res})$, where [E] is the concentration of free protein and k_on is the on-rate constant. It should be pointed out that the term $p_\mathrm{b}p_\mathrm{f}^2$ in eq 2 is not a monotonic function of p_b. However, the approximation $p_\mathrm{b}p_\mathrm{f}^2 \sim p_\mathrm{b}$ can be safely applied in these experiments due to the small value of p_b. Consequently, eq 2 can be rewritten in a simpler expression:

$$R_{2,\mathrm{obs}} \approx p_\mathrm{f}R_{2,\mathrm{f}} + p_\mathrm{b}(R_{2,\mathrm{b}} + R_\mathrm{ex}) \qquad (3)$$

where

$$R_\mathrm{ex} = 4\pi^2\tau_\mathrm{res}(\delta_\mathrm{f} - \delta_\mathrm{b})^2 \qquad (4)$$

The relative sensitivity of the experiment according to eq 3 is proportional to p_b. The fraction of bound ligand is provided by the equation

$$p_\mathrm{b} = \frac{[\mathrm{E_T}] + [\mathrm{L_T}] + K_\mathrm{D} - \sqrt{([\mathrm{E_T}] + [\mathrm{L_T}] + K_\mathrm{D})^2 - 4[\mathrm{E_T}][\mathrm{L_T}]}}{2[\mathrm{L_T}]} \qquad (5)$$

where $[\mathrm{E_T}]$ is the total protein concentration and K_D is the dissociation binding constant of the fragment. A graph displaying the dependence of p_b for a weak-affinity ligand L with a K_D for the receptor of 400 μM as a function of its concentration is shown in Figure 9a.

It is worth noting that the possibility of performing the experiments at low concentration results in a larger p_b, thus allowing the detection of very weak-affinity (K_D in the millimolar range) fragments that would otherwise escape detection when working at higher concentrations. The detection threshold can be further improved by reducing the concentration of the tested molecules with the use of cryogenic probe technology optimized for ^{19}F detection.

(iv) Testing the mixtures at low concentration allows the simultaneous detection of multiple ligands present in the same mixture. This is particularly relevant when testing complex mixtures and when the tested mixture contains multiple ligands with very different affinities for the receptor. A molecule with high affinity for the receptor could prevent the detection of a weak-affinity molecule present in the same mixture. The fraction of bound molecule in the presence of another competing molecule within the tested mixture is given by the equation[61]

$$p_\mathrm{b} = \frac{2\sqrt{(a^2 - 3b)}\cos(\theta/3) - a}{3K_\mathrm{D} + 2\sqrt{(a^2 - 3b)}\cos(\theta/3) - a} \qquad (6)$$

with

$$\theta = \arccos\left[\frac{-2a^3 + 9ab - 27c}{2\sqrt{(a^2 - 3b)^3}}\right] \qquad (7)$$

$$a = K_\mathrm{D} + K_\mathrm{I} + [\mathrm{L_T}] + [\mathrm{I_T}] - [\mathrm{E_T}] \qquad (8)$$

$$b = \{[\mathrm{I_T}] - [\mathrm{E_T}]\}K_\mathrm{D} + \{[\mathrm{L_T}] - [\mathrm{E_T}]\}K_\mathrm{I} + K_\mathrm{D}K_\mathrm{I} \qquad (9)$$

$$c = -K_\mathrm{D}K_\mathrm{I}[\mathrm{E_T}] \qquad (10)$$

where $[\mathrm{I_T}]$ and $[\mathrm{L_T}]$ are the concentrations of the two ligands present in the mixture, and K_I and K_D are their dissociation binding constants. Typically, the concentration of the molecules in the mixture is the same, i.e., $[\mathrm{I_T}] = [\mathrm{L_T}]$, but their dissociation binding constants can be very different. A simulation of p_b for the weak-affinity ligand L ($K_\mathrm{D} = 400$ μM) in the presence of a high-affinity ligand I ($K_\mathrm{I} = 10$ or 40 μM) within the same mixture as a function of the tested mixture concentration is also shown in Figure 9a. The graph displaying the ratio of p_b in the absence and in the presence of ligand I is reported in Figure 9b. At low tested mixture concentration, the p_b of weak-affinity ligand L is not significantly reduced by the presence of the high-affinity ligand I. For example, the screening at a mixture concentration of 20 μM results only in a 1.4- and a 2.6-fold reduction in p_b for L in the presence of I with $K_\mathrm{I} = 40$ and 10 μM, respectively. Therefore, both molecules can be detected simultaneously, despite their large difference in binding constants (10- and 40-fold). A considerable decrease in p_b is

(59) Swift, T. J.; Connick, R. E. *J. Chem. Phys.* **1962**, *37*, 307–320.
(60) (a) Martin, M. L.; Martin, G. J.; Delpuech, J.-J. *Practical NMR Spectroscopy*; Heyden: London, 1980. (b) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon Press: Oxford, 1987.
(61) (a) Wang, Z. X. *FEBS Lett.* **1995**, *360*, 111–114. (b) Sigurskjold, B. W. *Anal. Biochem.* **2000**, *277*, 260–266.

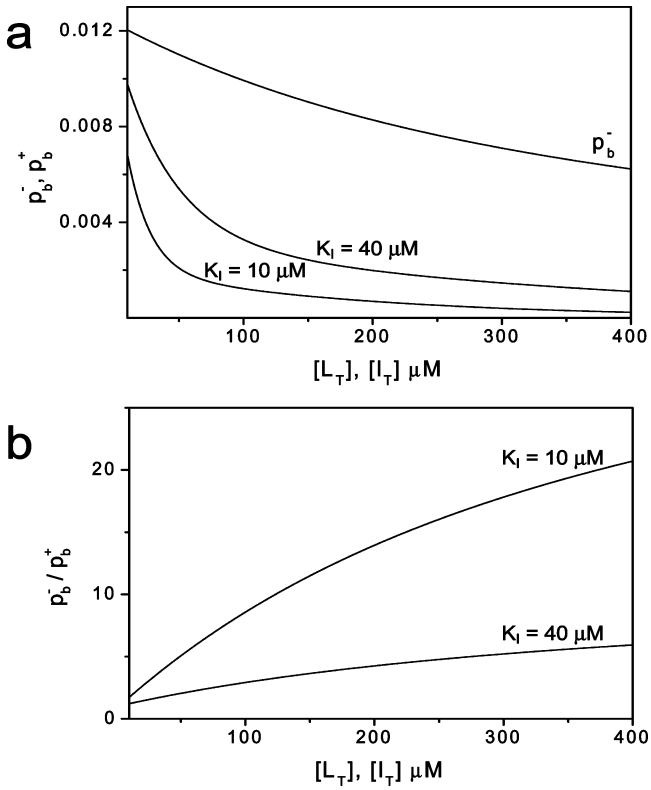


Figure 9. (a) Simulation displaying the fraction of ligand L bound to the receptor in the absence (p_b^-) and in the presence (p_b^+) of another ligand I as a function of the concentration of the molecules ([L_T] and [I_T]) present in the mixture. The *X* axis range is from 10 to 400 μM. Equations 5 and 6 were used for the simulations of p_b^- and p_b^+, respectively, with a protein concentration [E_T] of 5 μM. The concentration of the two molecules in the mixture was assumed to be the same, i.e., [I_T] = [L_T]. The simulation was performed for a ligand L with a K_D of 400 μM in the absence and in the presence of a ligand I with a K_I of 10 or 40 μM, as indicated on the graph. (b) Ratio p_b^-/p_b^+ for the ligand L as a function of the mixture concentration. The value of these two curves gives the fold in attenuation value of p_b of ligand L with the ligand I present in the same mixture.

observed at higher tested concentrations, and the value rapidly approaches zero, thus preventing detection of the weak-affinity ligand. For example, screening at a mixture concentration of 200 μM results in a 4.3- and a 14.0-fold reduction in p_b for L in the presence of I with K_I = 40 and 10 μM, respectively.

When the binding constants of the two molecules are similar, the mutual reduction in p_b is not significant, and both molecules are easily detected. Figure 10 shows the experimental simultaneous detection of two ligands for trypsin present in the same mixture.

(v) Finally testing mixtures at low concentration also allows the detection of very high-affinity ligands and molecules that bind covalently to the receptor. The presence of a strong binder in fragment mixtures is unlikely due to the small size of the chemical fragments. However, mixtures of large fluorinated molecules could contain high-affinity ligands for the receptor. The detection of these strong ligands and, in addition, covalently bound ligands is possible by working at protein concentration comparable to the mixture concentration. In this case, the signal attenuation observed in the ^{19}F spectra without R_2 filter is determined by a reduction in the signal integral rather than by an increase in the signal line width.

Dynamic Range. The dynamic range (DR) of the ^{19}F NMR screening experiments is defined as

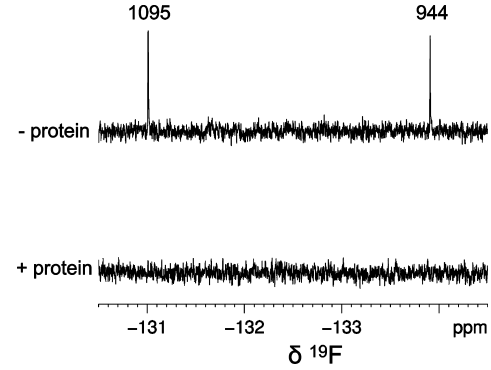


Figure 10. Expanded region of the ^{19}F NMR (564 MHz) R_2 filter spectra recorded with proton decoupling for mixture 103 (CF-containing molecules) in the absence (upper trace) and in the presence (lower trace) of 9 μM bovine trypsin. The length of the R_2 filter was 0.32 s, with a τ period of 40 ms. A total of 96 scans were acquired, with a repetition time of 3.8 s. The concentration of the molecules was 35 μM. The two signals correspond to two molecules of the mixture that bind to the protein, as proved by their disappearance in the spectra in the presence of the protein (lower trace). The label adjacent to the signals indicates the F-cluster number of the two molecules. A significant reduction in the ^{19}F transverse relaxation in the presence of the protein is observed for both molecules, despite their competition for the same binding site.

$$\mathrm{DR} = [(R_{2,b} + R_{ex}) - R_{2,f}] \qquad (11)$$

It has been demonstrated that the DR in these experiments can be very large due to the simultaneous contribution of chemical shift anisotropy and exchange mechanisms, resulting in a sensitive assay.[36] $R_{2,b}$ is large due to the significant role of ^{19}F chemical shift anisotropy at the strong magnetic fields used today. R_{ex} can also be considerable, as shown in the following example.

The dissociation binding constant K_D for the chemical fragment 5-(trifluoromethyl)-2-pyridylamidine, measured with isothermal titration calorimetry (ITC), is 621 ± 60 μM. The fraction of bound fragment p_b in our experimental screening conditions (i.e., tested molecule and protein concentrations of 18 and 9 μM, respectively), calculated with eq 5, is 0.014. The effect on the observed line width, despite the small value of p_b and the small size of the protein investigated (MW of bovine trypsin is 23.8 kDa), is very pronounced and can be observed even without the application of the R_2 filter, as shown in Figure 11b.

Figure 11 shows the spectrum of **1** in the absence of protein and in the presence of protein without and with a strong competitor. In addition to the broadening of the signal, a downfield shift $\Delta_{obs} = (\delta_{obs} - \delta_f)$ of ~6 Hz is observed in the spectrum in the presence of protein, where δ_{obs} is the observed isotropic chemical shift of the ^{19}F resonance. (The Δ_{obs} value is 54 Hz in a sample with a p_b value of 0.12.) The presence of the competitor completely displaces the fragment from the protein, so the signal intensity is restored and Δ_{obs} becomes zero.

A large dispersion in the ^{19}F chemical shift has been observed for the same type of fluorinated amino acid inserted into proteins. For example, a range of up to 16.8 ppm (9475 Hz with a 600 MHz spectrometer) was observed for the ^{19}F resonances of the aromatic amino acid [4-F]-Trp incorporated into hen egg white lysozyme/$(NAG)_3$[62] and 11.0 ppm (6204 Hz with a 600 MHz spectrometer) for the ^{19}F resonances of the aromatic amino acid

(62) Lian, C.; Montez, H.; Le, B.; Patterson, J.; Harrell, S.; Laws, D.; Matsumura, I.; Pearson, J.; Oldfield, E. *Biochemistry* **1994**, *33*, 5238–5245.

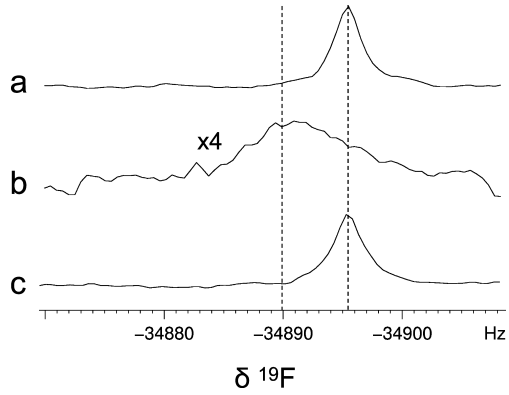


Figure 11. Expanded region of the ^{19}F NMR (564 MHz) spectra recorded with proton decoupling for the molecule **1** in the absence of the protein (a), in the presence of 9 μM bovine trypsin (b), and in the presence of 9 μM bovine trypsin and a strong inhibitor (K_D in the nanomolar range) (c). The spectrum in (b) is plotted at 4 times the vertical scale of (a) and (c) for better visual inspection. The concentration of the molecule was 18 μM. A total of 80 scans were acquired, with a repetition time of 3.8 s. The two dotted vertical lines indicate the position of the resonance in the different spectra.

[3-F]-Tyr incorporated into alkaline phosphatase.[63] These values are an order of magnitude larger than the total shielding expected from ring current and CONH susceptibility anisotropy. Oldfield and co-workers claim that fluorine shielding in proteins appears to be dominated by weak electrostatic field effects.[64] They concluded that electric field-induced shifts in proteins could be very large, on the order of 10 ppm for fluoroaromatic groups, owing to the low dielectric constants expected in protein interiors and the large polarizability of the C−F bond, and consequently these effects might dominate the ^{19}F shift nonequivalencies seen experimentally.[65] However, Gerig and co-workers argue that the conclusion that electrostatic fields are the only factors defining the dependence of fluorine shielding on the tertiary structure of proteins is not sufficient, and that van der Waals interactions should also be considered in this context.[66]

The large electrostatic field in the interior of a protein, resulting from the small dielectric constant value, van der Waals interactions between protein and ligand, and in some cases the involvement of the fluorine atoms in intermolecular hydrogen bonds can also result in a significant ^{19}F isotropic chemical shift difference between the free and bound states of a F or CF_3-containing molecule that interacts with the receptor.

The Δ_{obs} value measured in the spectrum of Figure 11b, the knowledge of the fragment binding constant, and the protein and fragment concentrations used in the experiment allow the determination of the lowest difference $\Delta_{max} = (\delta_b - \delta_f)$ value. For a weak-affinity ligand, this is given by the equation[36b,58]

$$\Delta_{max} \approx \frac{\Delta_{obs}}{p_b} \qquad (12)$$

According to eq 12, a value of 421 Hz (0.75 ppm with a 600 MHz spectrometer) is obtained for Δ_{max} for the ^{19}F resonance of **1**. This value is significant despite the small size of the fragment. It is worth noting that the Δ_{max} value for **1** derived with eq 12 represents only the lowest limit. The Δ_{max} value could be even larger than the value calculated here if this assumption of very fast exchange does not apply completely to **1** (see simulation contained in Figure 12 of ref 36b). Therefore, according to eq 4, it is evident that the exchange term R_{ex} is large and contributes significantly to the observed transverse relaxation rate $R_{2,obs}$ of the fragment ^{19}F resonance in the presence of the protein.

Follow-up Screening. The identified ligand(s) of the LEF library can then be used as a reporter (also known as spy molecule) in the FAXS experiments for the optimization of the active fragment(s) and for the detection of novel ligands by screening other compound collections.

During the optimization phase, it is essential to have a robust and reliable method to assess SAR, i.e., to relate structural changes in newly tested molecules with their effects on binding affinity to the target of interest. The SAR investigation can be either qualitative (ranking) or quantitative (based on K_I determination). In both cases, the structurally similar molecules are screened with the FAXS experiments in the presence of the original NMR hit. Only the ^{19}F signal intensity of the original NMR hit is used for ranking the selected molecules according to their binding strength. With the knowledge of the dissociation binding constant K_D of the original NMR hit, it is also possible to derive the K_I of the selected molecules.[32,36] These values are then used in the calculation of the binding efficiency index (BEI)[67] or the ligand efficiency (LE) index[68] for each molecule.

Analogues to the hits can be selected from an in-house archive or commercial vendors (e.g., fluorinated fragments from ACD, see above). Molecules belonging to the same fluorine clusters of the NMR hits, isomers (or very close analogues) which differ only in the fluorine moiety position, and molecules (with or without fluorine) that are structurally similar are screened with the FAXS experiments[32,36] in the presence of the LEF NMR hit. As described previously, the NMR-based F-scan of the hits (i.e., the testing of close analogues differing only in the position of the fluorine moiety) allows the identification of the possible presence of a fluorophilic protein environment.

In our limited experience, we found that some of the hits belong to the same F-cluster despite their global structural diversity. However, we do not know if this finding *per se* is a clear indication of the presence of a fluorophilic protein environment. Additional experimental data from the LEF library applied to different proteins, together with structural data of these NMR-hits bound to the protein, are required to gain a better understanding of these experimental observations. This work is currently in progress in our laboratory.

The FAXS experiments with the LEF NMR hit as reporter can also be used to screen additional compound libraries, such as proprietary compound collections or focused libraries designed for a specific target class. In this process, novel chemical scaffolds are identified that compete with the fluorinated reporter molecule for the same binding site on the protein. The setup of this assay with **1** as reporter (S) and another fluorinated control

(63) Hull, W. E.; Sykes, B. D. *Biochemistry* **1974**, *13*, 3431–3437.
(64) (a) Oldfield, E. *Annu. Rev. Phys. Chem.* **2002**, *53*, 349–78. (b) Oldfield, E. *Phil. Trans. R. Soc. B* **2005**, *360*, 1347–1361.
(65) (a) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *260*, 1491–1496. (b) Pearson, J. G.; Oldfield, E.; Lee, F. S.; Warshel, A. *J. Am. Chem. Soc.* **1993**, *115*, 6851–6862. (c) de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1994**, *116*, 7453–7454.
(66) (a) Chambers, S. E.; Lau, E. Y.; Gerig, J. T. *J. Am. Chem. Soc.* **1994**, *116*, 3603–3604. (b) Gerig, J. T. *Prog. NMR Spectrosc.* **1994**, *26*, 293–370. (c) Lau, E. Y.; Gerig, J. T. *J. Am. Chem. Soc.* **1996**, *118*, 1194–1200.
(67) Abad-Zapatero, C.; Metz, J. T. *Drug Discovery Today* **2005**, *10*, 464–469.
(68) Hopkins, L. A.; Groom, C. R.; Alex, A. *Drug Discovery Today* **2004**, *9*, 430–431.

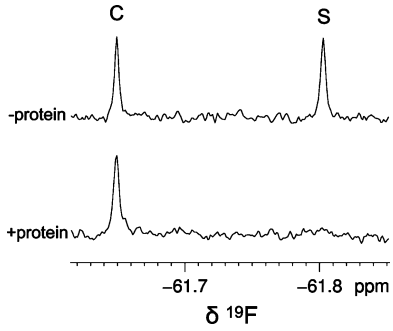


Figure 12. Set-up of the FAXS screening experiments with molecule **1** as reporter (S) and molecule C as control molecule. ^{19}F NMR (564 MHz) R_2 filter spectra recorded with proton decoupling containing the two molecules in the absence (upper trace) and in the presence (lower trace) of 360 nM bovine trypsin. The concentration of the molecules was 18 μM. The length of the R_2 filter was 0.8 s, with a τ period of 40 ms. A total of 96 scans were acquired with a repetition time of 3.8 s.

molecule (C) that does not interact with the receptor is shown in Figure 12.

The control molecule provides an internal reference that allows direct comparison of the data, even when they have been acquired with different experimental conditions (e.g., different number of scans). The setup of the FAXS is straightforward and requires only the optimization of the experimental conditions to be used in the assay. For example, the R_2 filter length in Figure 12 was set to 0.8 s to increase the sensitivity of the assay. This is possible due to the small $R_{2,f}$ value (0.93 s^{-1}) of the reporter fragment **1**. The concentration of the reporter is 18 μM, whereas the protein concentration is only 360 nM. These concentrations are respectively 34.5 and 1725 times lower than the dissociation binding constant of the reporter. This corresponds to a p_b for the reporter of only 0.000563 or, in other words, to 1 molecule bound and 1775 free. Despite the small p_b value and the small size of the protein, a significant effect is observed on the reporter resonance in the presence of the protein. Another parameter that is optimized is the temperature at which the FAXS is performed.[36b] For most protein−ligand interactions (i.e., with exothermic enthalpies), the dissociation rate k_{off}, where $k_{off} = 1/\tau_{res}$, is directly proportional to the temperature. A larger than 10-fold increase in τ_{res} was observed in surface plasmon resonance experiments[69] performed with different ligands by decreasing the temperature from 36 to 6 °C. This change in τ_{res} can result in a dramatic variation of R_{ex}, as previously presented with theoretical simulations.[36b] Therefore, it is possible to select a temperature, within the temperature range of protein stability, for maximizing the exchange contribution R_{ex} and thus further increasing the DR of the assay.

The detection threshold in the FAXS experiments is directly proportional to the dissociation binding constant and inversely proportional to the concentration of the fluorinated spy molecule.[70] Therefore, the use of a weak-affinity spy molecule at low concentration, as possible in these experiments, allows the reliable detection of fluorinated and non-fluorinated molecules that bind only weakly to the receptor, thus capturing the broadest chemical diversity for potential ligands. In addition, the low protein concentration required for these experiments results in a limited protein consumption, thus allowing screening of protein biotargets that can be expressed only in low amounts (e.g., membrane proteins).

(69) Navratilova, I.; et al. *Anal. Biochem.* **2007**, *364*, 67–77.

Conclusion

In summary, we have presented a novel strategy, based on a fluorine fingerprints descriptor, for the design of a fluorinated fragment library that takes into account the local environment of fluorine. The LEF library combined with ^{19}F NMR-based screening represents an efficient and sensitive approach for the initial fragment identification within a fragment drug discovery project. The presence of fluorophilic protein environments is probed with the proposed approach, and this information can then be exploited for the optimization of the initial fragment. The low concentration of the tested mixtures, together with the large dynamic range of the ^{19}F NMR experiments, allows the reliable identification of very weak-affinity ligands and the simultaneous detection of multiple binders present within the same mixture.

Material and Methods

Bovine trypsin was purchased from United States Biological (catalog no. T8672). The NMR samples were in 50 mM Tris, 100 mM NaCl, and 5 mM $CaCl_2$, pH 8.0, and contained 9% D_2O for the lock signal. The 5 mm NMR tubes were used for the experiments. The small molecules were prepared in concentrated stock solutions in deuterated DMSO and stored at 4 and −20 °C. All the ^{19}F NMR experiments were recorded at 27 °C with a Bruker DRX-600 spectrometer operating at a ^{19}F Larmor frequency of 564 MHz and equipped with a SampleJet robot for sample tube automation. The R_2 filter experiments were recorded with the Carr−Purcell−Meibom−Gill scheme with a time interval of 40 ms between the 180° pulses and with different total lengths. The spectra were acquired with proton decoupling using the Waltz-16 composite pulse sequence with a 90° pulse of 150 μs. The data were collected with a spectral width of 42.17 and 29.92 ppm for the CF and CF_3 mixtures, respectively. The acquisition and repetition times were 0.8 and 3.8 s, respectively. The data were multiplied with an exponential multiplication window with a line width of 1 Hz prior to Fourier transformation. Typically 80 and 96 scans, for the CF and CF_3 mixtures, respectively, were recorded for each spectrum.

Chemical shifts are referenced to the $CFCl_3$ signal in water. The theoretical simulations were performed using the Origin 7.0 and Microsoft Excel software packages.

Acknowledgment. We thank A. D'Arcy and F. Villard for helpful assistance with the preparation of the mixtures and P. Erbel and S. Rüdisser for enlightening discussions.

Supporting Information Available: Complete refs 1b, 16, and 69. This material is available free of charge via the Internet at http://pubs.acs.org.

JA905207T

(70) Dalvit, C.; Mongelli, N.; Papeo, G.; Giordano, P.; Veronesi, M.; Moskau, D.; Kümmerle, R. *J. Am. Chem. Soc.* **2005**, *127*, 13380–13385.