Fragment-based drug design facilitated by dynamic combinatorial chemistry and NMR spectroscopy

Peter Kroon, s1687735

18/03/2015

Abstract

Drug development is a challenging process, which often fails. To ameliorate this, we attempted to develop an efficient methodology for fragment-based drug design based on dynamic combinatorial chemistry and NMR spectroscopy. To this end, two libraries were designed based on crystallographic data from the group of Klebe for the model protein endothiapepsin using computer aided modelling. The designed compounds all feature an acylhdyrazone linker and at least one fluorine atom. The first library was discarded due to problems with the synthesis. The second library was analysed using quantitative ^{19}F-NMR and ^{1}H-STD-NMR spectroscopy, using trifluoroacetic acid as an internal reference for the fluorine NMR spectra. No binding compounds could be identified, even though one of the compounds was a known binder. It is hypothesised this is due to the trifluoroacetic acid since the ^{1}H-STD-NMR spectrometry also produced no results. However, it could be concluded that ^{19}F-NMR spectrometry provides better peak resolution with little loss of sensitivity compared to ^{1}H-NMR spectrometry when analysing mixtures; and that STD-NMR requires far less protein than regular NMR spectroscopy for analysing templated DCLs.

Table of Contents

Introduction

Drug development nowadays is plagued by long development times, stemming mainly from the number of compounds that need to be considered, and the amount of time required to synthesise and analyse each of these compounds.[1,2] We hope to ameliorate this situation by employing dynamic combinatorial chemistry based on already known inhibitors to greatly reduce the time required in synthesis, and nuclear magnetic resonance (NMR) spectroscopy to facilitate analysis.

Structure-based drug design

Drug development is a long and costly process, progressing through several discrete steps. First, the enzymes involved in the disease need to be identified and it has to be confirmed that knocking out these enzymes kills only the pathogen (Target identification and validation). Once a target enzyme is selected an inhibitor can be designed based on known inhibitors, enzyme function, or ideally a three-dimensional structure of the enzyme. Commonly, around 100,000 separate compounds are considered at this stage.[1] If a potent inhibitor is found, this lead is further optimised (Lead discovery and optimisation) and finally submitted for clinical trials (Figure 1).[1,2]

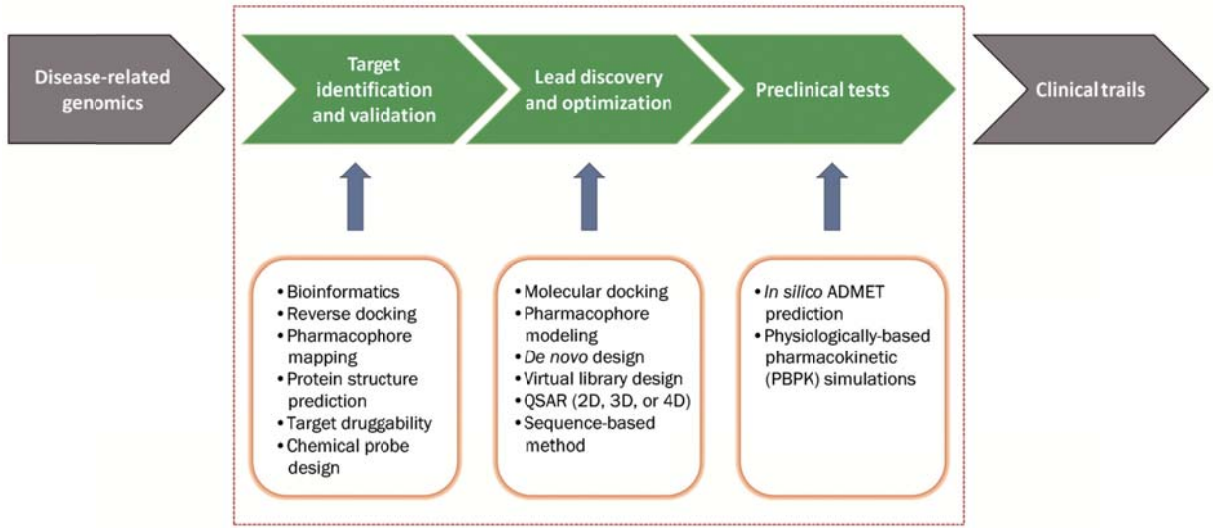


Figure 1: The process from disease to new drug. Figure taken from Jiang *et al*.[2]

Medicinal chemistry is involved mainly in lead optimisation.[1] Once a three-dimensional structure of the enzyme is obtained (usually *via* crystallography), an iterative cycle, depicted in Figure 2, is started. By using computer-aided modelling techniques, a potential inhibitor is designed; this relies heavily on the intuition and skill of the medicinal chemist involved.[3] This compound is then synthesised, and its efficacy is tested *in vitro*. If it is confirmed that the compound inhibits the enzyme, it is attempted to obtain a co-crystal structure of the enzyme with this inhibitor. This new crystal structure is then used *in silico* to identify empty pockets in the enzyme, or moieties of the inhibitor that do not bind efficiently. The process is repeated until a sufficiently potent inhibitor is found.[1,2] This means that before a sufficiently potent inhibitor is found a multitude of (often intricate) compounds have to be synthesised, purified and analysed. This is usually very time-consuming.

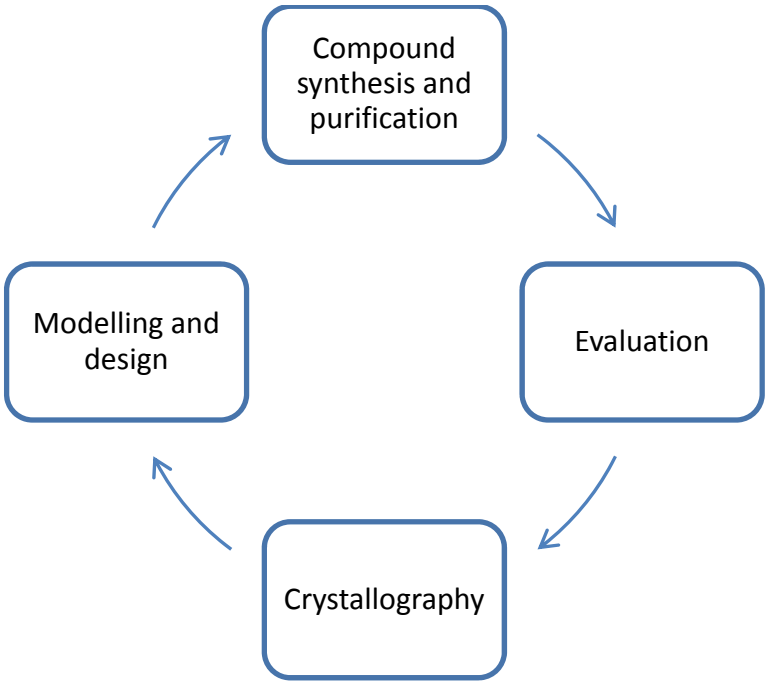


Figure 2: The process of structure-based drug design.

Fragment-based drug design

Fragment-based drug design is a methodology that can be used to facilitate structure-based drug design, and in particular the lead-identification and optimisation steps. Classically, many large compounds are screened against a target, in the hope one or more will bind to the protein (high-throughput screening, HTS). However, because the compounds screened are usually rather large, and enzymes bind compounds very selectively, many compounds often need to be screened to find a successful candidate.[1] Two alternative approaches can be used when applying fragment-based drug design: fragment linking, and fragment growing.

In fragment linking and growing, smaller compounds are screened instead. These will usually bind the target enzyme only weakly, but chemical space can be sampled far more efficiently. If multiple, small compounds bind the protein in different pockets, these can be thought of as fragments of the ultimate inhibitor. In this case, if the binding modes of these fragments can be elucidated, it may be possible to link the fragments together in order to obtain a highly potent inhibitor in an efficient manner. However, great care must be exercised: if there is too much strain in the linker between the fragments, or if it slightly too long or too short, this will prevent the inhibitor from binding like the fragments, leading to a loss of potency. Because of this strict geometrical requirement on the linker, fragment linking is usually considered hard, but there are examples in which this methodology has been successfully applied, resulting in very potent inhibitors.[4–6]

If instead there is only binding information available for one fragment, this fragment may be "grown" in order to make more beneficial interactions with the protein. This process is generally more elaborate than fragment linking, but it is easier. There are still pitfalls however. It is often very easy to identify empty hydrophobic pockets in a three-dimensional model of an enzyme. Growing the inhibitor in this direction would probably result in an insoluble compound however. This

methodology has also been used successfully, resulting in commercially available drugs.[7–10] See Figure 3.

In practice, the distinction between fragment linking and fragment growing is not so clear-cut. It is not uncommon to grow one fragment in order to mimic a second.[11] In all these cases, computer modelling is used to facilitate the process.

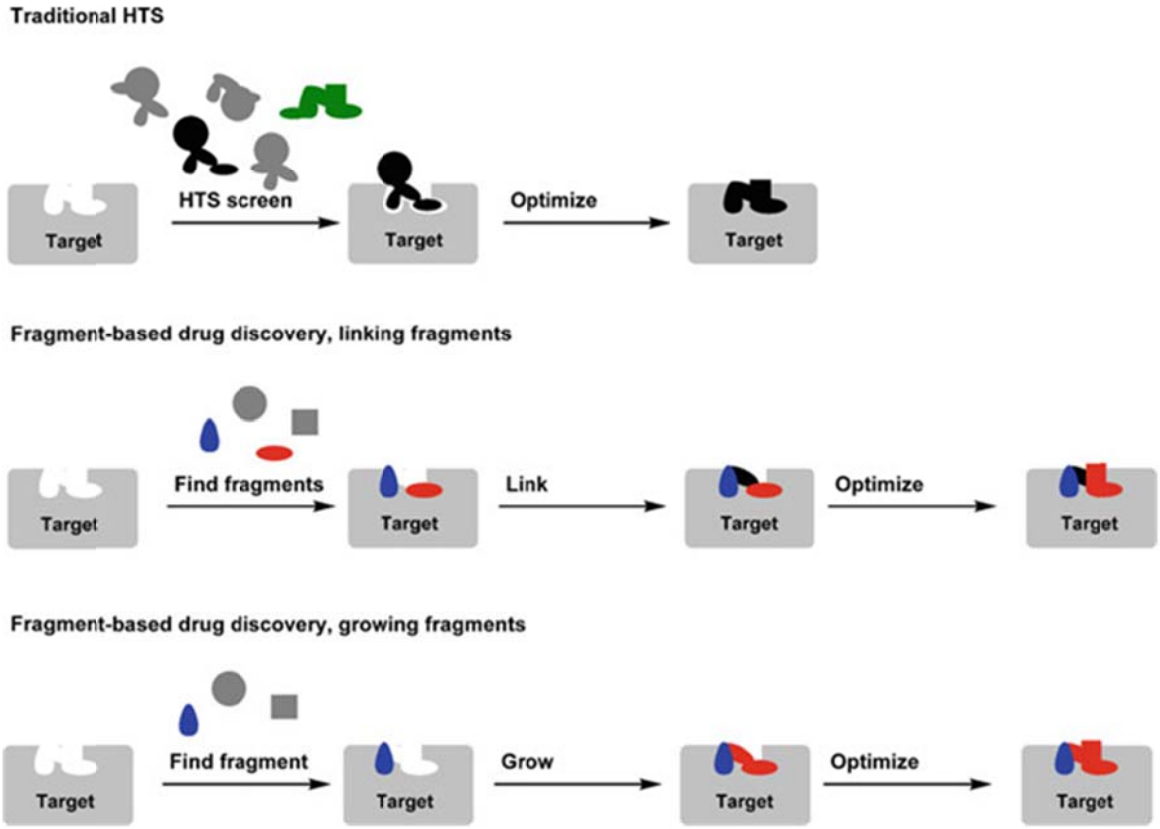


Figure 3: Fragment-based drug design. Top panel: the classical approach to drug design is high-throughput screening; many large compounds are screened against the target enzyme, but only a few — if any — will bind. Middle panel: if two or more small organic molecules bind the target protein these may be linked together to form a potent inhibitor in an elegant manner. Bottom panel: If only one small organic molecule binds the protein, this fragment may be grown into an empty pocket in order to form a potent inhibitor. Figure taken from Erlanson.[1]

Dynamic combinatorial chemistry

Dynamic combinatorial chemistry is an approach in which a library of dynamically interchanging compounds is formed.[12–14] Next, if a template is added, some library members can bind this template, and will in this way be depleted from their equilibrium. The library will then re-equilibrate according to Le Châtelier's principle, making more of the compounds that bind (Figure 4). The degree of change in concentration of individual library members can be used as a measure for binding affinity;* this degree of change in concentration is called an amplification factor.[12–14] It should be noted that the amplification factor is a function of: the affinity of the binder, the concentration of template, and the total concentration of building blocks. This means, that in order to see a significant change in concentration, a lot of template may be required.

* Some care must be exercised since it is possible that compounds which do not bind will also be amplified due to the changes in building block concentrations.

DCC has the advantage that relatively few building blocks need to be synthesised, and that these building blocks can be synthesised in parallel, or may even be commercially available. This allows for efficient screening of large libraries with relatively little effort,[12,14] which means that this technique significantly accelerates both the synthesis and analysis of potential inhibitors, since not every library member has to be synthesised, purified and analysed separately; particularly if a dynamic library can be designed around one or more fragments. For example, if two fragments can link together in a (few) different ways the system will show which linker has the best orientation and properties for the target protein.

Because of this, dynamic combinatorial chemistry is increasingly being used in drug design,[12,14] and there are numerous examples in which this technique has been used successfully.[15–18]

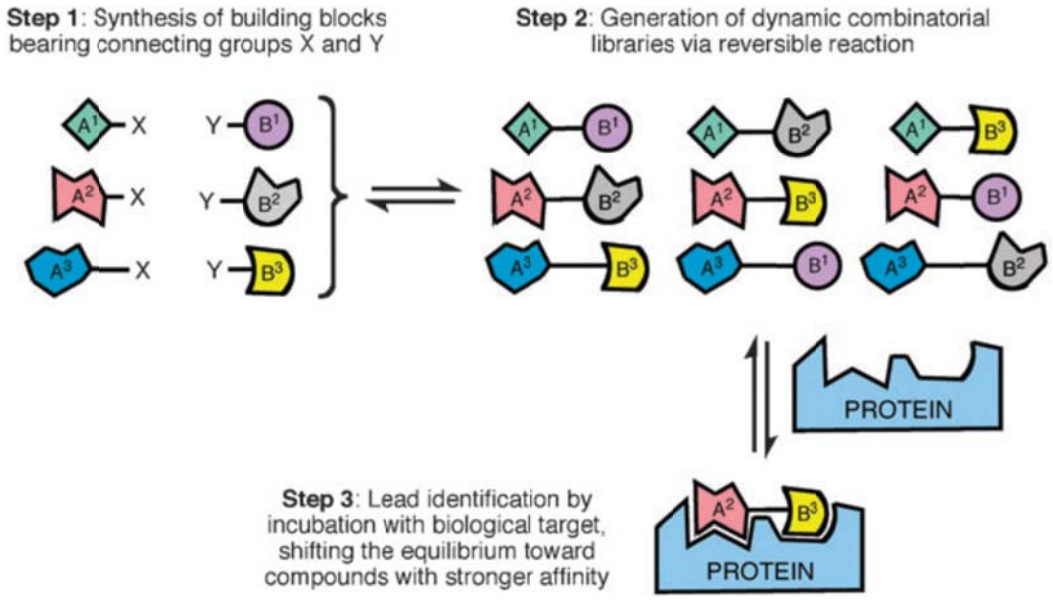
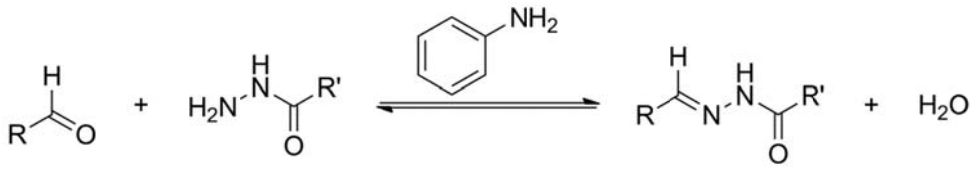


Figure 4: A schematic representation of dynamic combinatorial chemistry. Since the compound A2-B3 binds the template protein best, it is selected from the library. Figure taken from Mamidyala and Finn.[13]

Since the only requirement for DCC is that the library constituents can interchange, many different connections can be used between the building blocks. In principle, all reversible, covalent bond-forming reactions can be used, such as imine formation, transacylation, disulfide formation, the Diels-Alder reaction, *etc*. The only prerequisite is that the reaction runs at an acceptable rate in conditions which can be tolerated by the template.[12,14,19] In case of a protein this is usually an aqueous environment that is slightly acidic or basic. Lastly, it is desirable that the equilibria can be stopped at will by changing the reaction conditions, since this facilitates analysis.[12,14]

To these ends we have opted for acylhydrazones as linking group. Here an aldehyde and hydrazide can reversibly combine to form an acylhydrazone. Acylhydrazones are rather stable, and equilibration occurs at an acidic pH. Additionally, acylhydrazone exchange can be catalysed by a nucleophilic catalyst such as aniline,[20] allowing this reaction to also be used under a broader range of conditions (Scheme 1).



Scheme 1: The formation of an acylhydrazone from an aldehyde and a hydrazide under influence of catalytic aniline.

Classically, DCLs are analysed using methods such as liquid chromatography mass spectrometry (LC-MS). These techniques have some drawbacks since they are destructive, the equilibrium has to be stopped, separation may be hard to achieve, all library members need to have a different mass, and a lot of protein is required since the equilibrium is measured directly. We hope to circumvent some of these drawbacks by using NMR spectroscopy.

NMR techniques

Quantitative NMR

NMR spectroscopy is a common analytical method in synthetic organic chemistry, and in particular ^{1}H-NMR spectroscopy. There are good reasons why ^{1}H is so popular in NMR since it is the most sensitive nucleus, and ^{1}H atoms are abundant in most organic compounds, which results in a lot of structural information. Unfortunately, ^{1}H-NMR has some drawbacks. Firstly, because ^{1}H atoms are so abundant, many solvents also contain them, demanding the use of (expensive) deuterated solvents.* Secondly, ^{1}H-NMR signals span only a narrow spectral range: roughly from –2 to 12 ppm, which may result in overlapping signals.

An alternative to ^{1}H-NMR may be ^{19}F-NMR spectroscopy. ^{19}F is 100% naturally abundant, and in addition, its gyromagnetic ratio is 83% as large as that of ^{1}H. This means that ^{19}F-NMR is almost as sensitive as ^{1}H-NMR. Furthermore, ^{19}F-NMR signals span a much broader spectral range: roughly from –200 to 20 ppm. This means that ^{19}F signals are often well separated. Also, since many common solvents do not contain ^{19}F atoms, no isotopically enriched solvent is needed;* although an internal reference might be desirable.

In order to analyse a DCL with these techniques, a change in equilibrium is measured directly. To do so, two spectra are recorded, one with template and one without, after which the signal intensities of relevant peaks are normalised against an internal reference and compared between the two spectra. Since the change in equilibrium is measured directly, large amounts of template are usually required.

Saturation-transfer difference NMR

An alternative to measuring the change in equilibrium caused by addition of a template is to directly measure the binding of ligands to the template. To this end, saturation-transfer difference NMR (STD-NMR) was developed. The template, in this case a protein, is saturated in a spectral region, which is not of interest, and this saturation spreads through the protein. Once the active site has been saturated, this saturation can transfer to a bound inhibitor by means of spin diffusion through the nuclear Oberhauser effect (NOE).[21] As a result, inhibitors that bind to the protein will also be saturated, and no longer give a signal in NMR. This means that when a regular ^{1}H-NMR spectrum is

* Note that all NMR samples must contain at least some deuterium for locking and shimming.

compared to one where the protein was saturated, the signals corresponding to bound inhibitors will decrease in intensity; compounds that do not bind the template will not be affected (Figure 5).[21]

Because of the dynamic binding and unbinding of compounds, one protein molecule may saturate many ligand molecules, meaning that this technique requires far less protein than quantitative NMR methods. For this to work however, extensive measurement times are required when mixtures are analysed.[19] Also, rapid ligand exchange is required,[21] but for weak binders this is usually not a problem.

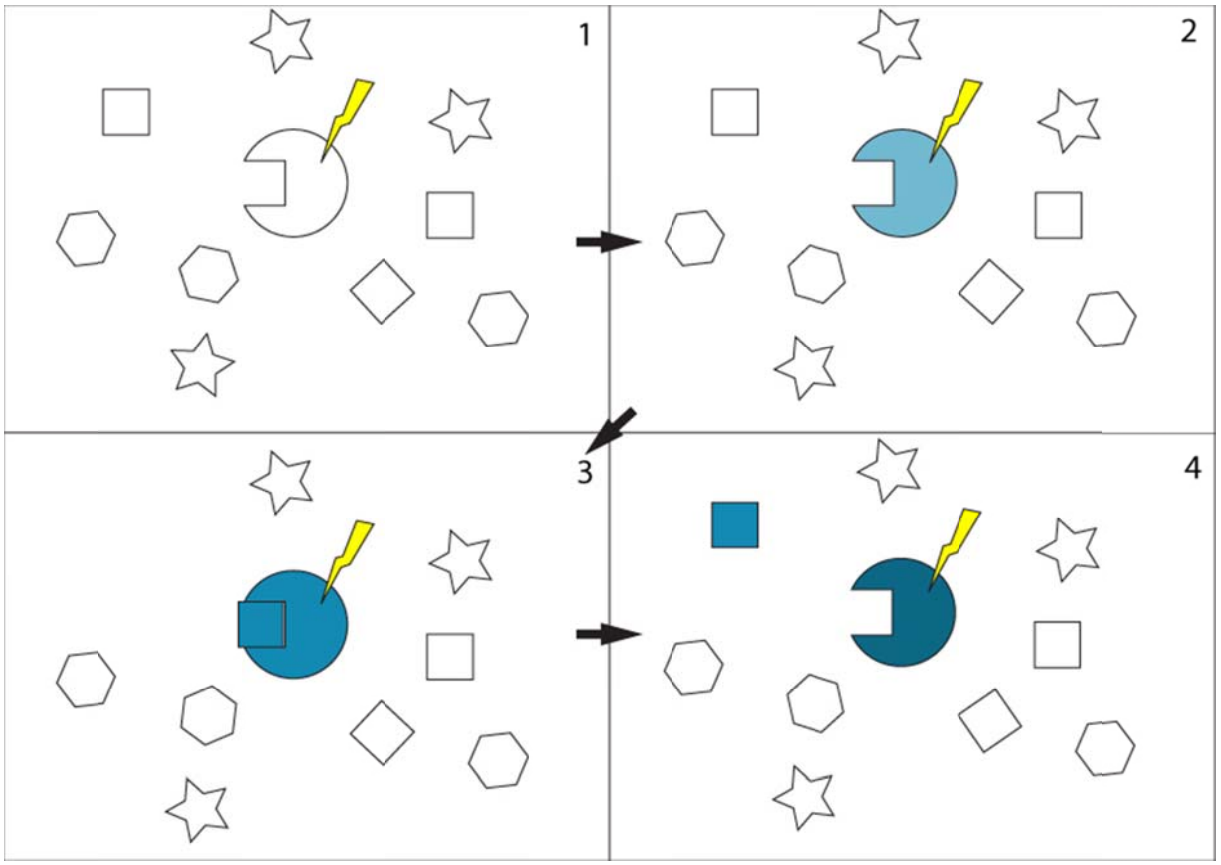


Figure 5: a schematic depiction of STD-NMR spectroscopy. Saturation is depicted in blue. Saturation is transferred from the protein to ligands that bind. Due to dynamic binding and unbinding, one protein molecule can transfer saturation to multiple ligand molecules.

Unfortunately, ^{1}H-STD-NMR spectroscopy suffers from the same problems as normal ^{1}H-NMR spectroscopy: a small spectral range and the requirement for isotopically enriched solvents. To circumvent these hurdles, the technique has been modified to ^{19}F-STD-NMR, in which the template is saturated as before, and the saturation is transferred to the ligand. However, instead of observing the ^{1}H channel, it is decoupled, and ^{19}F is observed. Additional sensitivity can be obtained by explicitly transferring the saturation from hydrogens nuclei in the inhibitor to the fluorine nuclei using an INEPT-like sequence.[22] This technique however, does require an NMR instrument that is capable of measuring ^{19}F whilst saturating ^{1}H.

Of course, when analysing a DCL using any ^{19}F-NMR technique all library members need to contain at least one ^{19}F atom.

Endothiapepsin

In order to demonstrate the advantages DCC can offer in fragment-based drug design, we will attempt to design an inhibitor for the enzyme endothiapepsin. This protein is originally found in the fungus *Endothia parasitica*, and it is a model enzyme for aspartic proteases.[23]

This class of enzymes is widespread, and has varied biological functions and plays a causative role in several serious diseases such as malaria and AIDS.[23–25] Its function in nature is to hydrolyse amide bonds,[23–25] and the inhibitors known today block the active site.[23,26] This active site features a catalytic dyad: two aspartic acid residues (Asp-32 and Asp-215) of which one is protonated.[23,25] These aspartic acids are tightly bound to a water molecule through hydrogen bonds (Figure 6).[23,25] Some of the known inhibitors displace this catalytic water molecule, but others require it for binding.[19]

Figure 6: X-ray crystal structure of endothiapepsin (PDB code: 1ER8[27]). The protein backbone is depicted as a green cartoon. The catalytic dyad (aspartic acid residues 32 and 215) is depicted as a stick model (Colour code: C: green, O: red, N: blue). The water molecule closely bound to the catalytic dyad is shown as a red sphere.

As a model enzyme, endothiapepsin is readily available and crystallises easily. Additionally, the research group of Klebe *et al.*[28] has previously published co-crystal structures of this enzyme in complex with a series of fragments; its mechanism is fully elucidated;[23,25] its pH optimum is 4.5, which is beneficial for acylhydrazone exchange;[23,26] it has been shown in the past that acylhydrazone chemistry is compatible with endothiapepsin;[19] and lastly, our group is currently also working with this enzyme.[19]

Based on the requirements that all potential inhibitors should contain at least one fluorine atom for analysis and contain at least one acylhydrazone moiety to enable the dynamic exchange, we designed a library of compounds based on computer models in order to demonstrate that drug development can be greatly accelerated using DCC and STD-NMR spectroscopy.

Library design

A library of compounds was designed based on fragments in complex with endothiapepsin reported by the group of G. Klebe.[28] The fragments used will be referred to as **F109** and **F284**, in line with their original paper. The binding modes of these fragments are depicted in Figure 7 and their structures in Scheme 2. Since the fragments overlap when binding, they had to be made smaller in order to accommodate an acylhydrazone linker, whilst preserving key interactions with the protein. Fragment **F284** was reduced to a fluorophenyl moiety, and the hydrazide in fragment **F109** was substituted with an ethylamine moiety to prevent interference with the acylhydrazone exchange.

Since it is hard to design the ideal linker and predict exactly in what way the acylhydrazone moiety would have to be oriented, a total of eight inhibitors were designed: four featuring a "forward" acylhydrazone moiety, and four featuring a "backward" acylhydrazone moiety. These were constructed by systematically including methylene spacers at either end of the acylhydrazone moiety (Scheme 3).

Predicted binding modes for all these compounds, along with calculated binding energies can be found in the supporting information.

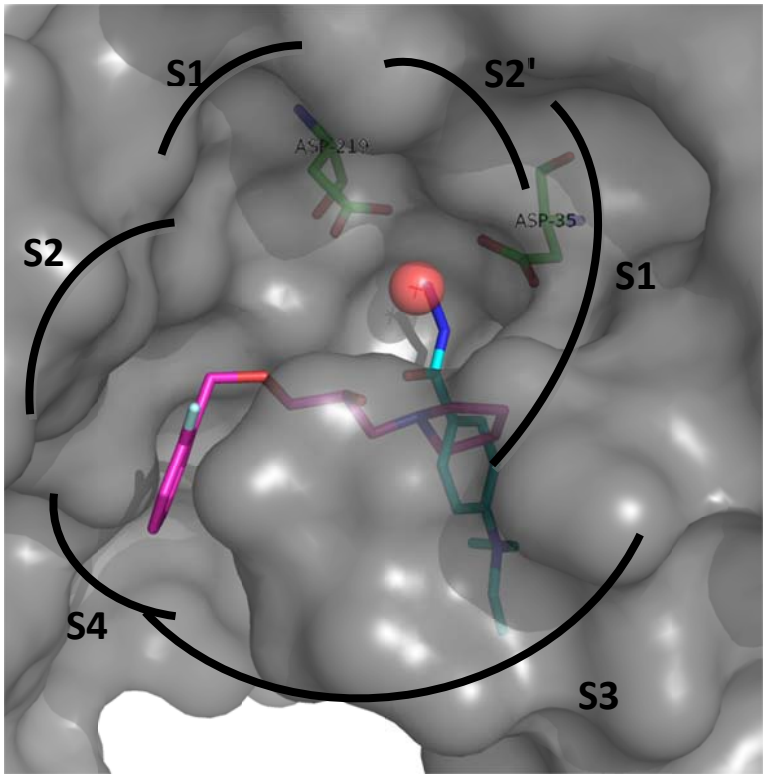
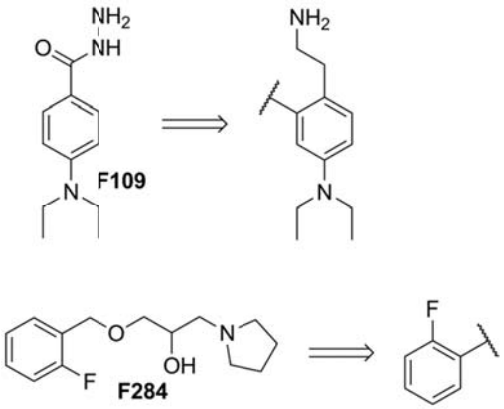
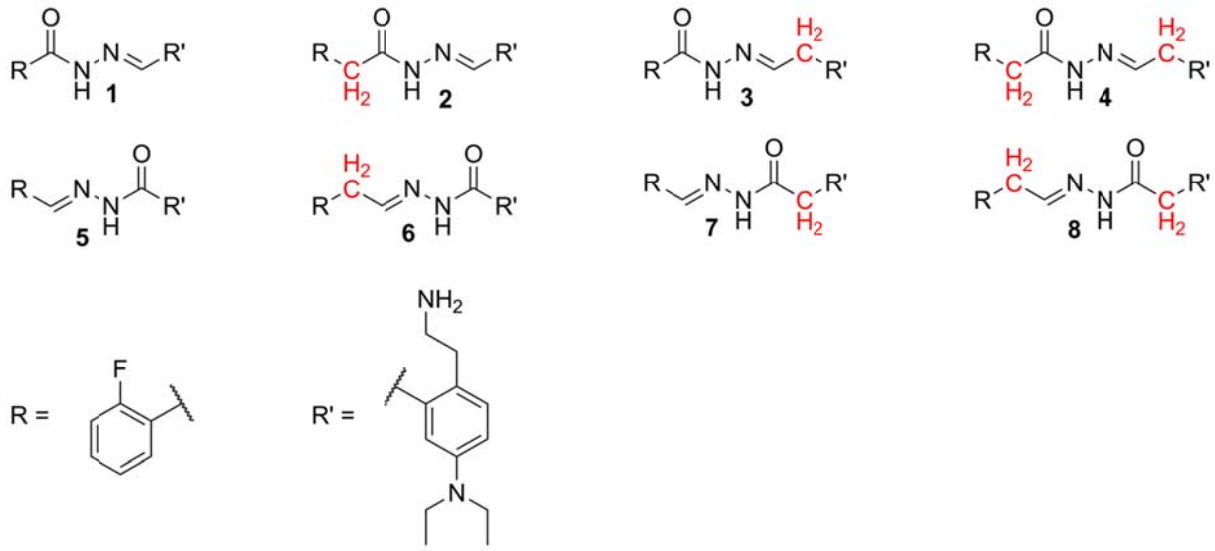


Figure 7: The binding modes of the two fragments co-crystallised with endothiapepsin (PDB codes: 3PBZ and 3PMU).[28] Colour code: O: red, N: blue, F: pale cyan, $C_{aspartic\ dyad}$: green, C_{F109}: teal, C_{F284}: hot pink. Endothiapepsin's solvent-accessible surface is depicted in grey. It can be seen that the fragments bind in different pockets, and that fragment F109 displaces the water molecule bound to the active site upon binding (translucent red sphere).



Scheme 2: After obtaining binding modes for fragments **F109** and **F284**, they were modified to make room for an acylhydrazone linker; reducing **F284** to a fluorophenyl moiety, and exchanging the hydrazide of **F109** for an ethylamine moiety.



Scheme 3: The designed inhibitors. They were formed from fragments **F109** and **F284** (Scheme 2) by adding either a "forward" (**1**–**4**) or "backward" (**5**–**8**) acylhydrazone moiety. A methylene spacer was systematically included at either end of the acylhydrazone moiety, and is depicted in red.

Due to problems with the synthesis of the R' fragment in Scheme 3 (see Supporting Information) a second library was designed, this time consisting of commercially available compounds.

The R' fragment was replaced for imidazole aldehyde (**9**), which was co-crystallised with endothiapepsin in unpublished work by M. Mondal, N. Radeva, A. Hirsch and G. Klebe. Based on **9** and **F284**, a new library was designed. Since there was ample space around **F284** inside the protein, the position of the fluorine was varied (**10**–**15**), and some compounds with trifluoromethyl substituents were also designed (Figure 8 and Scheme 4, **16**–**18**).

These nine building blocks were obtained commercially and used to construct a DCL based on co-crystal structures of endothiapepsin in order to demonstrate that the use of DCC can significantly accelerate the drug-development process since both the synthesis and biological analysis can be done for many compounds in one experiment. The analysis was done using NMR spectroscopy, enabling non-destructive analysis of the DCL.

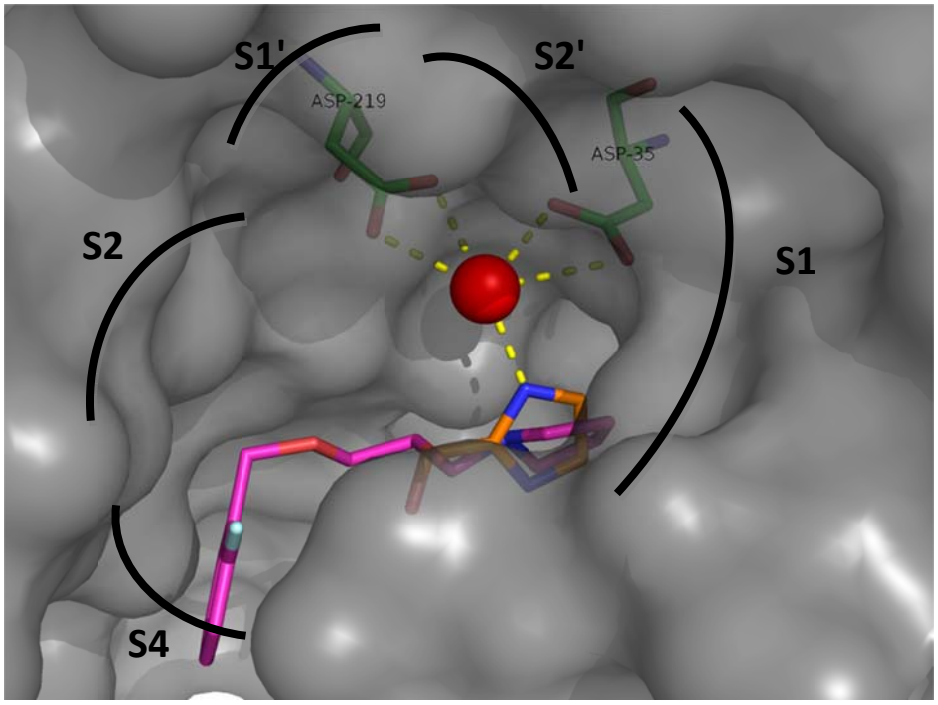
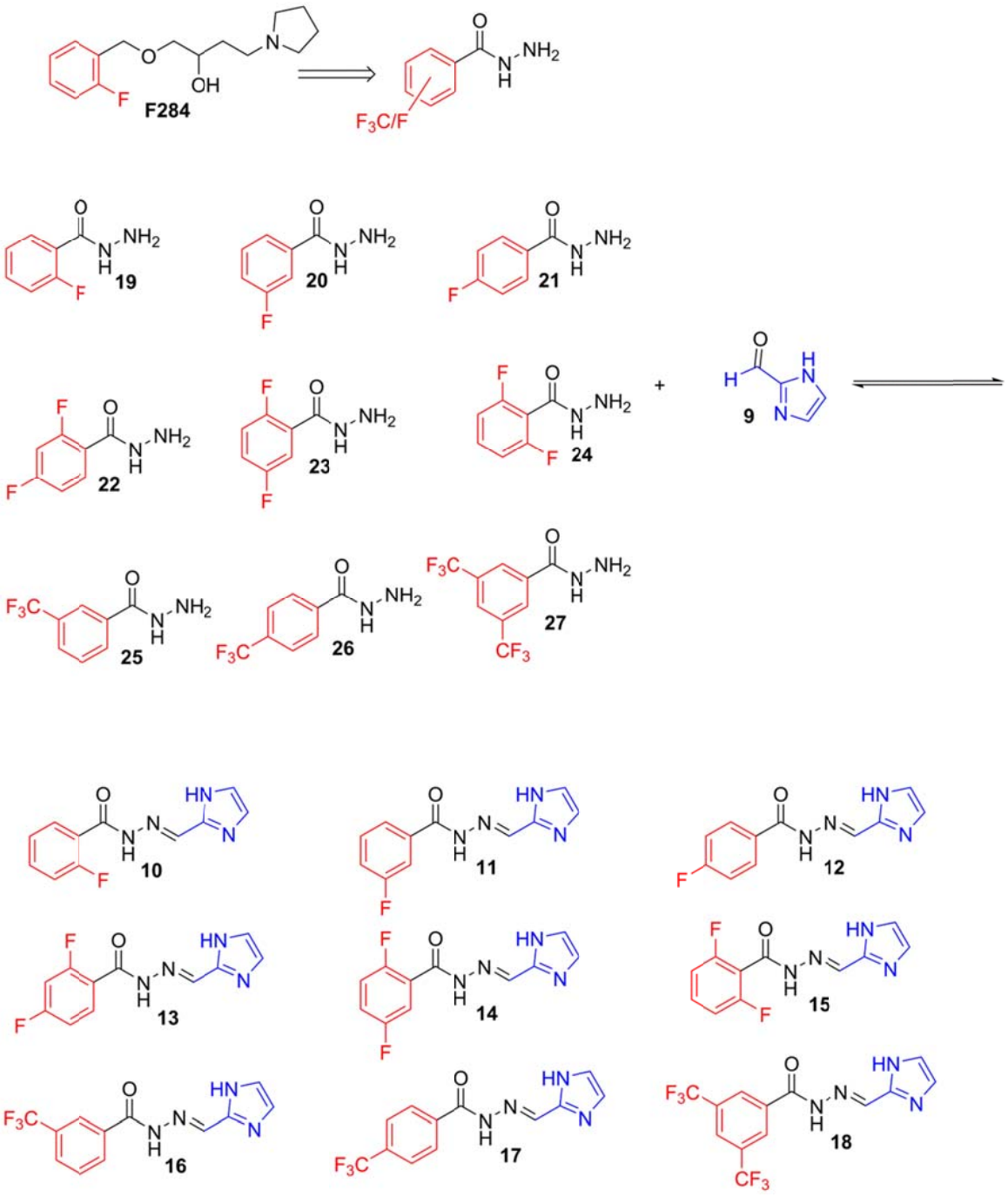


Figure 8: Co-crystal structures of **F284** (PDB code: 3PMU)[28] and imidazole aldehyde.* Colour code: O: red, N: blue, F: pale cyan, $C_{aspartic\ dyad}$: green, $C_{imidazol\ aldehyde}$: orange, C_{F284}: hot pink. Endothiapepsin's solvent-accessible surface is depicted in grey. It can be seen that imidazole aldehyde requires a water molecule (red sphere) in order to bind to the protein's active site *via* hydrogen bonds (dashed yellow lines).

* Work of M. Mondal, N. Radeva, A. Hirsch and G. Klebe. To be published.



Scheme 4: The building blocks used. The compounds are based on **F284** and **9**, but the position of **t**he fluorine substituent in **F284** was varied, and potentially substituted for a **t**rifluoromethyl group in order to yield acylhydrazones **10**–**18** from hydrazides **19**–**27**.

Modelling

The designed libraries were first studied *in silico*, to see if the designed compounds were viable inhibitors and would bind to the enzyme like the fragments upon which they were based.

Modelling was performed using the software program Moloc[29] by first generating the desired acylhydrazone inside the protein, and then optimising its position whilst keeping the enzyme stationary. Water molecules were reset to their positions according to the crystal structure for each acylhydrazone.

According to Moloc, all acylhydrazones should bind to the catalytic dyad through a hydrogen bond mediated by a water molecule; and occasionally form additional hydrogen bonds. Furthermore, Van der Waals interactions were generally very favourable. As an example, the binding modes of **13** and **15** are shown in Figure 9 and 10, respectively. Binding modes for **10**–**18** are provided in the Supporting Information.

Each acylhydrazone was docked inside the enzyme using the FlexX docking module in the LeadIT suite.[30] Aspartic acid residue 219 was protonated and key water molecules were unrestrained; *E*/*Z* isomerism was considered. The 30 poses which were most promising according to the software were visually inspected.

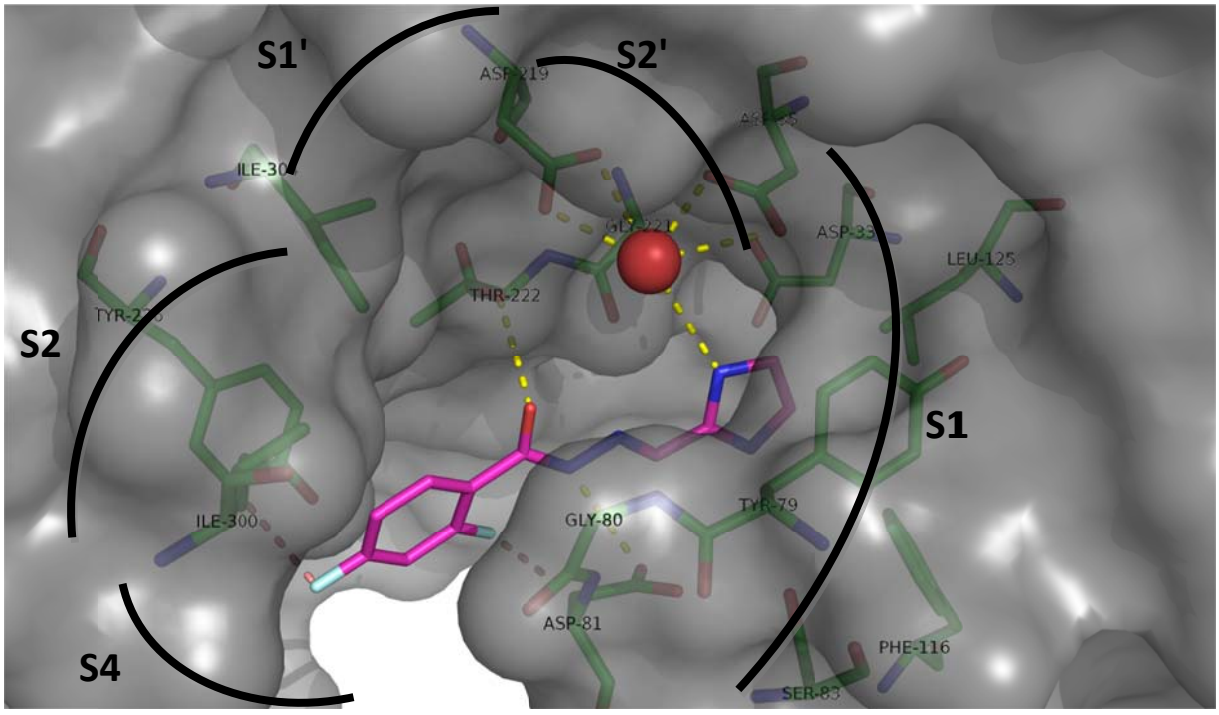


Figure 9: The predicted binding mode of **13**. Colour code: O: red, F: pale cyan, $C_{protein}$: green, C_{13}: purple, H-bonds: dashed, yellow lines. Favourable Van der Waals interactions involving fluorine atoms are depicted as dashed, salmon lines. The water molecule bound closely to the catalytic dyad is shown as a red sphere.

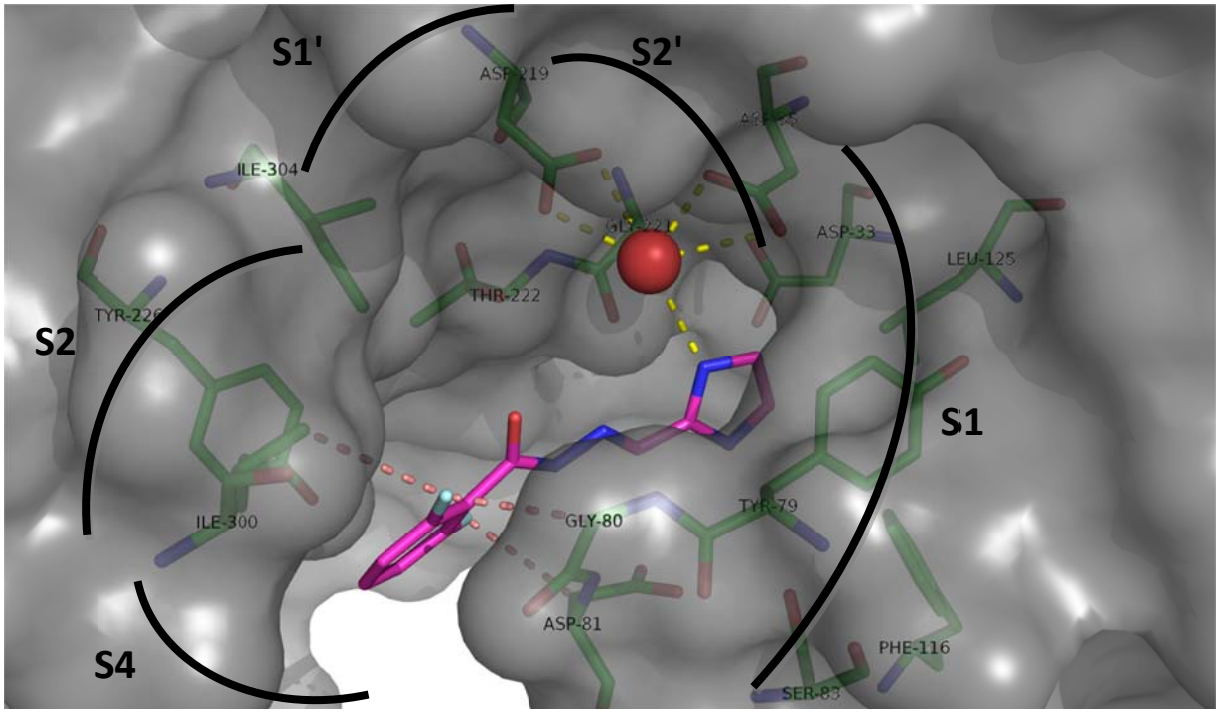


Figure 10: The predicted binding mode of **15**. Colour code: O: red, F: pale cyan, $C_{protein}$: green, C_{15}: purple, H-bonds: dashed, yellow lines. Favourable Van der Waals interactions involving fluorine atoms are depicted as dashed, salmon lines. The water molecule bound closely to the catalytic dyad is shown as a red sphere.

Unfortunately, FlexX produced no poses for any of the acylhydrazones in which the imidazole moiety was bound to the catalytic dyad with a hydrogen bond *via* a water molecule. For this reason the FlexX results were discarded. Some ligands, which were most promising according to Moloc, were scored by Hyde[30] directly; however, the resulting calculated energy of binding was very poor, and more importantly, the ligand was not bound to the catalytic dyad.

To further analyse this anomaly, **9** was docked using FlexX. In this case, the software did produce poses in which the ligand was bound to the catalytic dyad as indicated by the crystal structure. As of yet, it is unknown why the software would not produce poses in which the acylhydrazones were bound to the catalytic dyad.

Discussion

Quantitative ^{19}F-NMR

Two spectra were recorded, one of a sample containing a high protein concentration, and one without protein (experimental details can be found in the Supporting Information). Peaks were well separated, although less so for trifluoromethylated compounds. The integrals of the peaks from the library members were normalised against the integral of peak from trifluoroacetic acid and compared. Since sum of the concentration of acylhydrazone and the concentration the corresponding hydrazide should be constant, their respective peak areas were summed to 100%. The result is plotted in Figure 11. Unfortunately, upon addition of protein, no significant change in concentrations is observed. This means that none of the compounds bind to the protein according to these data. The raw data and spectra are provided in the Supporting Information.

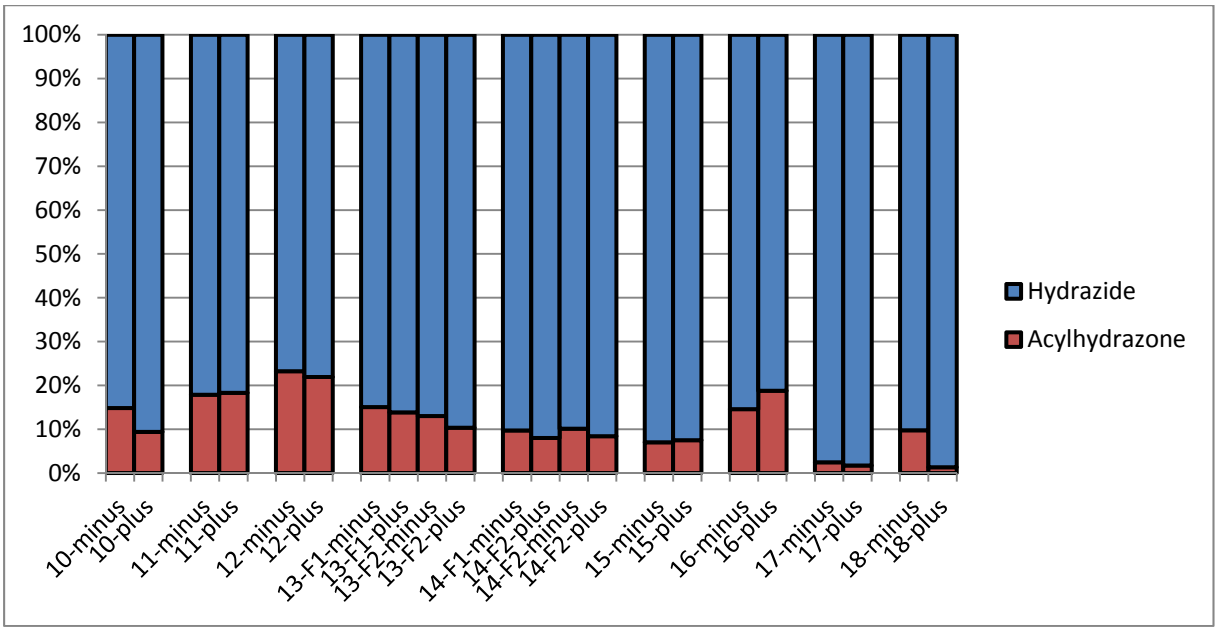


Figure 11:The difference in peak area for separate library members between a sample containing protein (plus), and one without (minus).

^{1}H-STD-NMR

No signal could be observed from the ^{1}H-STD-NMR spectrum, not even one originating from **9** (the spectrum is provided in the Supporting Information). Since **9** is a known inhibitor this could indicate a problem with library, and it is hypothesised that the trifluoroacetic acid interferes with the binding of library members to the protein by either denaturing the protein, blocking the active site, or stopping the acylhydrazone exchange. Therefore, these experiments should be repeated with either a lower concentration of trifluoroacetic acid, a different reference compound, or with a library of known binders in order to validate the methodology.

^{19}F-STD-NMR

Despite our best efforts, no ^{19}F-STD-NMR spectrum could be recorded on the NMR spectrometers available. No machine was available with a probe that could be tuned to ^{19}F and ^{1}H simultaneously. Tuning a probe to ^{19}F and saturating ^{1}H at −1.2 ppm by directly specifying the frequency produced too many artefacts to yield a useful spectrum.

Conclusions

In conclusion, DCC can play a significant role in both fragment linking and fragment growing, and should be part of every medicinal chemist's toolkit. However, analysis of the resulting mixtures is often challenging, but here a variety of NMR techniques can help, providing a way to analyse an equilibrating library in a non-destructive manner. STD-NMR is particularly interesting in this sense, since it requires very little protein.

As demonstrated, ^{19}F-NMR could extend these techniques further, granting less background signals and excellent peak separation. The only additional requirement is that all relevant compounds must contain at least one ^{19}F atom.

Modelling and the NMR studies are in agreement, and no inhibitor was found in this study. It would be worthwhile to repeat this methodology with a library of known inhibitors, as well as repeating it with a different reference compound to make sure trifluoroacetic acid does not influence the measurements. Once it is conclusively proven that ^{19}F-STD-NMR and quantitative ^{19}F-NMR are suitable for analysing DCLs, it should be explored how little protein is required as a function of the free energy of binding of the library members.

Acknowledgements

I would like to particularly thank Pieter van der Meulen for his extensive help with all the NMR measurements, as well as his limitless patience and perseverance in trying to get ^{19}F-STD-NMR to work on a machine that is not designed to do so. Furthermore, I would like to thank Milon Mondal for his daily supervision and help with the syntheses; and I would like to thank Anna Hirsch for supervising this research project in her group.

Bibliography

1. Erlanson, D. A. Introduction to fragment-based drug discovery. *Top. Curr. Chem.* **317,** 1–32 (2012).

2. Ou-Yang, S.-S. *et al.* Computational drug discovery. *Acta Pharmacol. Sin.* **33,** 1131–1140 (2012).

3. Kutchukian, P. S. *et al.* Inside the mind of a medicinal chemist: the role of human bias in compound prioritization during drug discovery. *PLoS One*. **7,** e48476 (2012).

4. Hajduk, P. J. *et al.* NMR-based modification of matrix metalloproteinase inhibitors with improved bioavailability. *J. Med. Chem.* **45,** 5628–5639 (2002).

5. Park, C. M. *et al.* Discovery of an orally bioavailable small molecule inhibitor of prosurvival B-cell lymphoma 2 proteins. *J. Med. Chem.* **51,** 6902–6915 (2008).

6. Wada, C. K. The evolution of the matrix metalloproteinase inhibitor drug discovery program at abbott laboratories. *Curr. Top. Med. Chem.* **4,** 1255–1267 (2004).

7. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363,** 809–819 (2010).

8. Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature*. **467,** 596–599 (2010).

9. Wyatt, P. G. *et al.* Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J. Med. Chem.* **51,** 4986–4999 (2008).

10. Artis, D. R. *et al.* Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 262–267 (2009).

11. Sandanayaka, V. *et al.* Discovery of 4-[(2S)-2-{[4-(4-Chlorophenoxy)phenoxy]methyl}-1-pyrrolidinyl] butanoic acid (DG-051) as a novel leukotriene A4 hydrolase inhibitor of leukotriene B4 biosynthesis. *J. Med. Chem.* **53,** 573–585 (2010).

12. Ramström, O. & Lehn, J.-M. Drug discovery by dynamic combinatorial libraries. *Nat. Rev. Drug Discov.* **1,** 26–36 (2002).

13. Mamidyala, S. K. & Finn, M. G. In situ click chemistry: probing the binding landscapes of biological molecules. *Chem. Soc. Rev.* **39,** 1252–1261 (2010).

14. Mondal, M. & Hirsch, A. K. H. Dynamic combinatorial chemistry: a tool to facilitate the identification of inhibitors for protein targets. *Chem. Soc. Rev.* **44,** 2455–2488 (2015).

15. Erlanson, D. A. *et al.* In situ assembly of enzyme inhibitors using extended tethering. *Nat. Biotechnol.* **21,** 308–314 (2003).

16. Maly, D. J., Choong, I. C. & Ellman, J. A. Combinatorial target-guided ligand assembly: identification of potent subtype-selective c-Src inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2419–2424 (2000).

17. Huc, I. & Lehn, J. M. Virtual combinatorial libraries: dynamic generation of molecular and supramolecular diversity by self-assembly. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 2106–2110 (1997).

18. Bunyapaiboonsri, T. *et al.* Dynamic deconvolution of a pre-equilibrated dynamic combinatorial library of acetylcholinesterase inhibitors. *Chembiochem*. **2,** 438–444 (2001).

19. Mondal, M. *et al.* Structure-based design of inhibitors of the aspartic protease endothiapepsin by exploiting dynamic combinatorial chemistry. *Angew. Chem. Int. Ed. Engl.* **53,** 3259–3263 (2014).

20. Bhat, V. T. *et al.* Nucleophilic catalysis of acylhydrazone equilibration for protein-directed dynamic covalent chemistry. *Nat. Chem.* **2,** 490–497 (2010).

21. Viegas, A., Manso, J., Nobrega, F. L. & Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **88,** 990–994 (2011).

22. Diercks, T. *et al.* Fluorinated carbohydrates as lectin ligands: versatile sensors in 19F-detected saturation transfer difference NMR spectroscopy. *Chemistry*. **15,** 5666–5668 (2009).

23. Coates, L. *et al.* The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray diffraction. *J. Am. Chem. Soc.* **130,** 7235–7237 (2008).

24. Davies, D. R. The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* **19,** 189–215 (1990).

25. Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. A. & Cooper, J. B. A neutron laue diffraction study of endothiapepsin: Implications for the aspartic proteinase mechanism. *Biochemistry*. **40,** 13149–13157 (2001).

26. Coates, L., Erskine, P. T., Crump, M. P., Wood, S. P. & Cooper, J. B. Five Atomic Resolution Structures of Endothiapepsin Inhibitor Complexes: Implications for the Aspartic Proteinase Mechanism. *J. Mol. Biol.* **318,** 1405–1415 (2002).

27. Pearl, L. & Blundell, T. The active site of aspartic proteinases. *FEBS Lett.* **174,** 96–101 (1984).

28. Köster, H. *et al.* A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **54,** 7784–7796 (2011).

29. Gerber, P. R. & Müller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided. Mol. Des.* **9,** 251–268 (1995).

30. Augustin, B. G. S. LeadIT 2.1.3. at <http://www.biosolveit.de>

31. Schrödinger LLC. The PyMOL Molecular Graphics System, Version 1.7.2. (2010). at <http://www.pymol.org>

32. McDonald, C. *et al.* The N-iodosuccinimide-mediated conversion of aldehydes to methyl esters. *J. Org. Chem.* **54,** 1213–1215 (1989).

33. Sharma, A. K., Subramani, A. V. & Gorman, C. B. Efficient synthesis of halo indanones via chlorosulfonic acid mediated Friedel–Crafts cyclization of aryl propionic acids and their use in alkylation reactions. *Tetrahedron*. **63,** 389–395 (2007).

34. Zhou, D. *et al.* 2-(Pyrrolidin-1-yl)ethyl-3,4-dihydroisoquinolin-1(2H)-one derivatives as potent and selective histamine-3 receptor antagonists. *J. Med. Chem.* **55,** 2452–2468 (2012).

35. Surry, D. S. & Buchwald, S. L. Dialkylbiaryl Phosphines in Pd-Catalyzed Amination: A User's Guide. *Chem. Sci.* **2,** 27–50 (2011).

Supporting Information

All figures in this work featuring endothiapepin are generated using PyMOL.[31]

Library A — Compounds 1 to 8

Modelling methods

The target compounds were first designed inside the protein using Moloc,[29] and their position was optimised whilst keeping the protein stationary. Each acylhydrazone was docked inside the enzyme using the FlexX docking module in the LeadIT suite.[30] Aspartic acid residue 219 was protonated and key water molecules were unrestrained; *E*/*Z* isomerism was considered. 30 poses were kept per ligand.

The resulting poses were visually inspected to ensure they bound to the protein like the original fragments, and were otherwise discarded. The most promising pose was used for Hyde scoring[30] in order to obtain a predicted free energy of binding (ΔG_{Hyde}).

Colour code for the figures: O: red, N: blue, F: pale cyan, $C_{protein}$: green, C_{ligand}: orange. The solvent-accessible surface of the protein is shown in grey. Hydrogen bonds are shown as dashed, yellow lines.

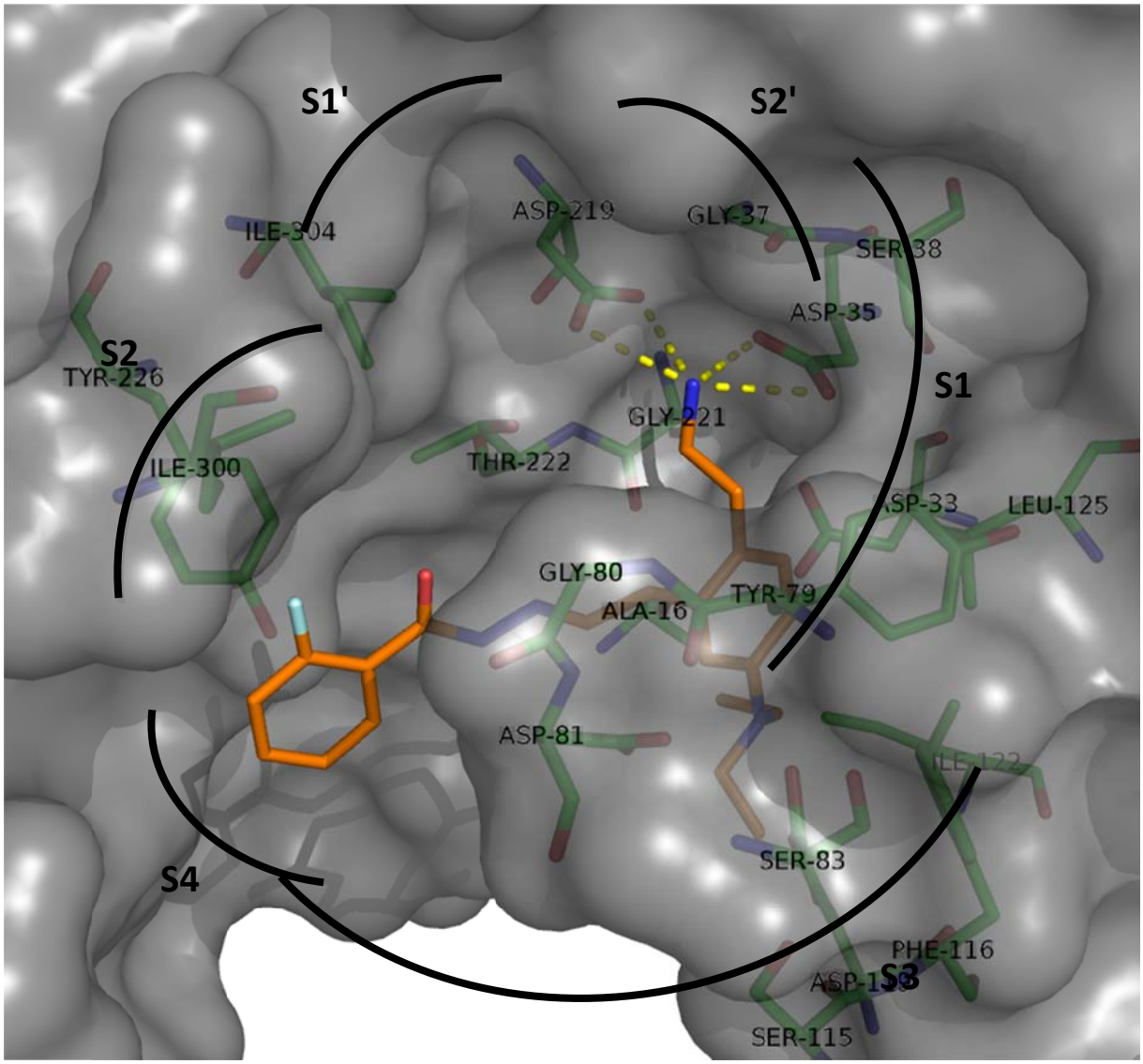
Modelling results

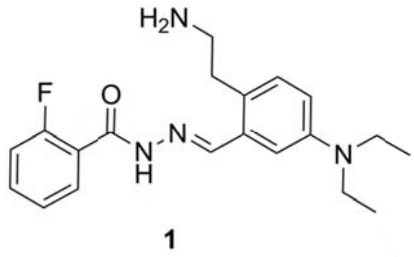
Table S1 depicts the binding energies as predicted by Hyde.

Compound	ΔG_{Hyde} (kJ/mol)
1	−30
2	−39
3	−21
4	−30
5	−32
6	−39
7	−19
8	−17

Table S1: The binding energies for compounds **1**–**8**, predicted by Hyde.

1: (*E*)-*N*'-(2-(2-Aminoethyl)-5-(diethylamino)benzylidene)-2-fluorobenzohydrazide

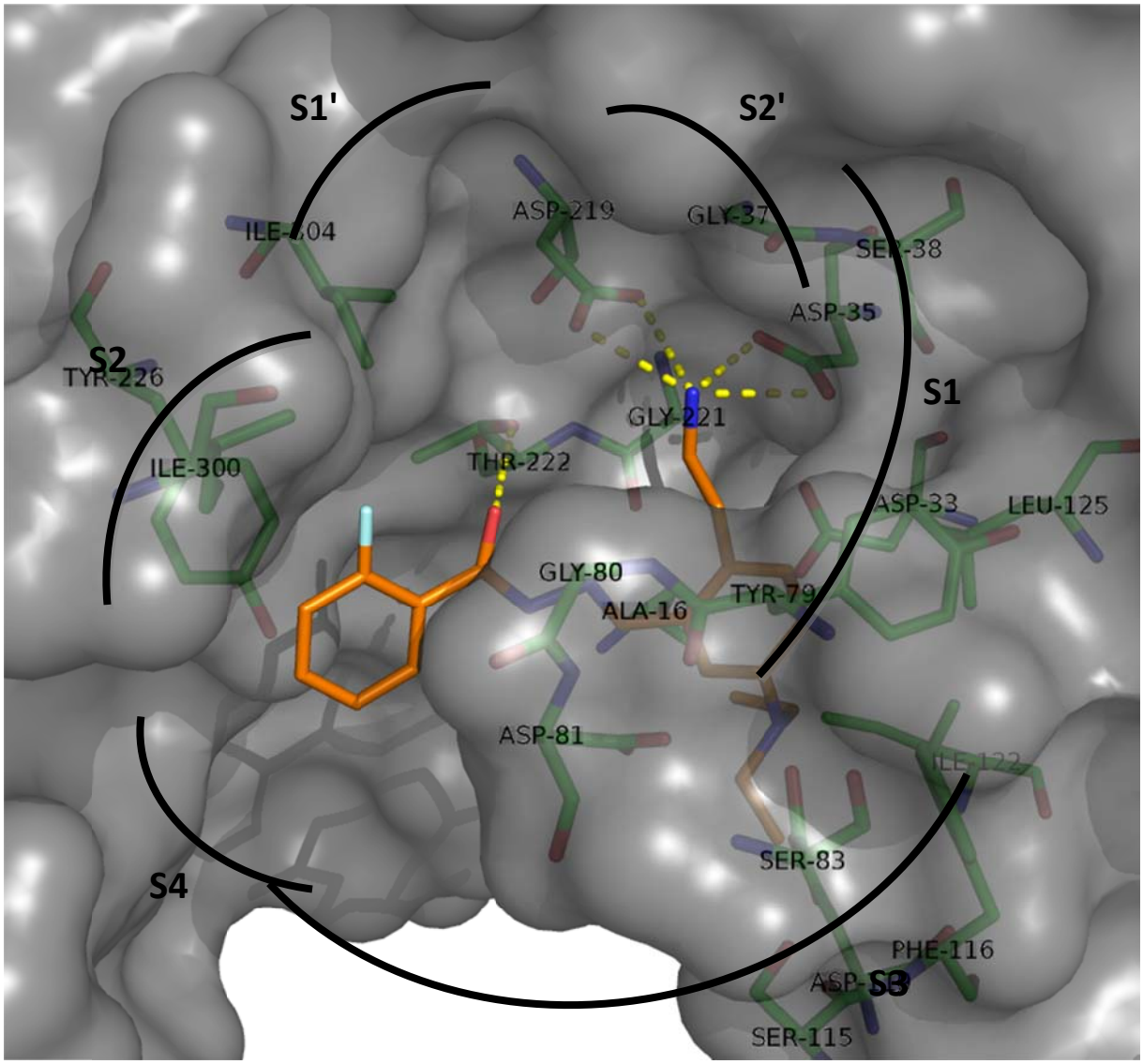


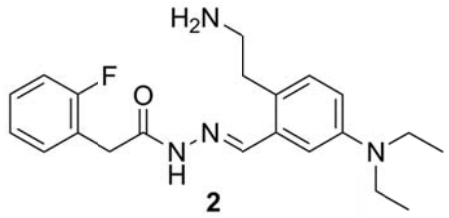


1

ΔG_{Hyde} = −30 kJ/mol

2:(*E*)-*N*'-(2-(2-Aminoethyl)-5-(diethylamino)benzylidene)-2-(2-fluorophenyl)acetohydrazide

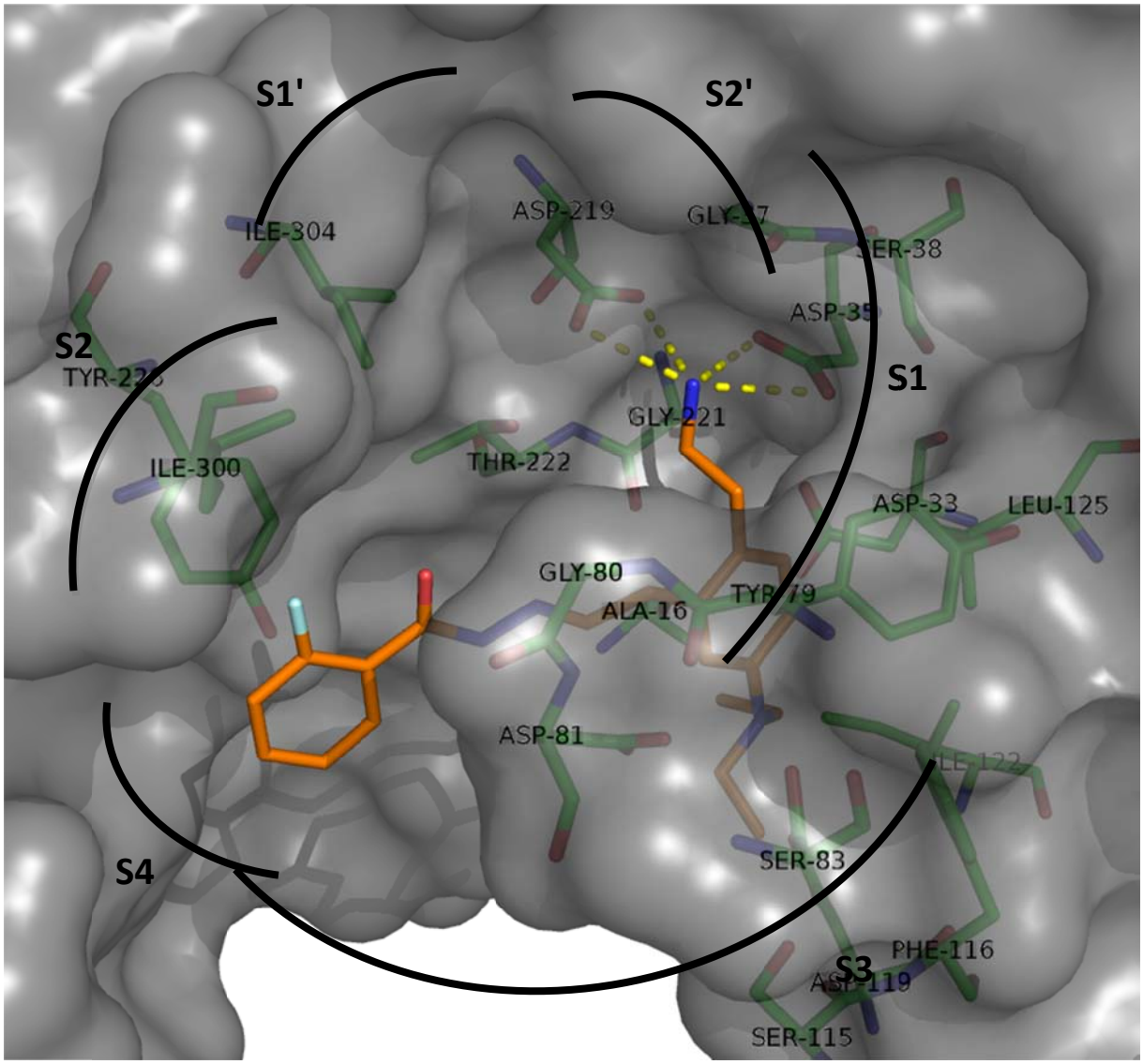


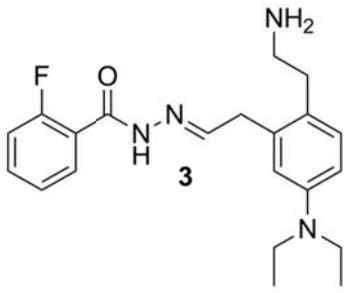


2

ΔG_{Hyde} = −39 kJ/mol

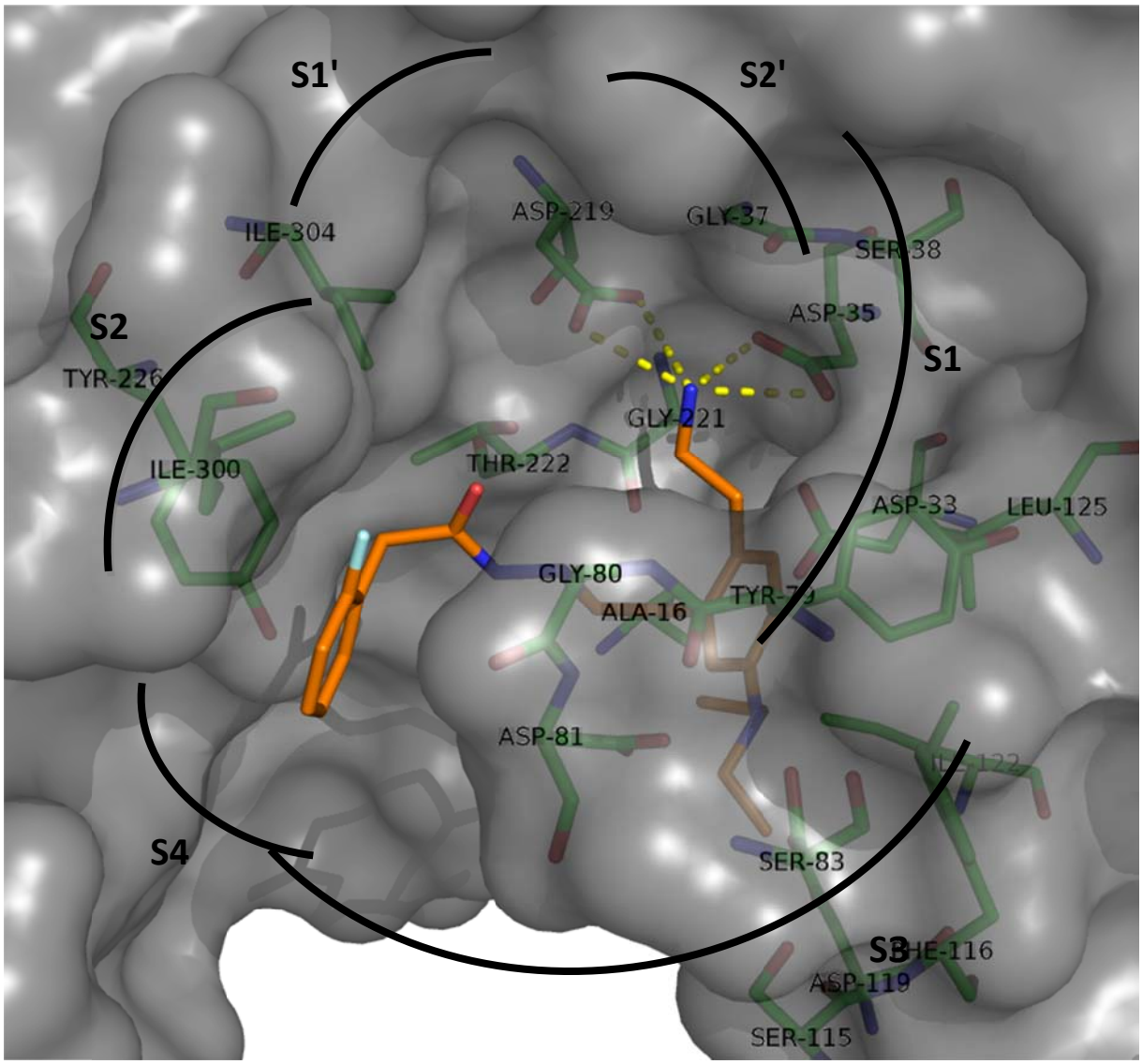
3: (*E*)-*N*'-(2-(2-(2-Aminoethyl)-5-(diethylamino)phenyl)ethylidene)-2-fluorobenzohydrazide

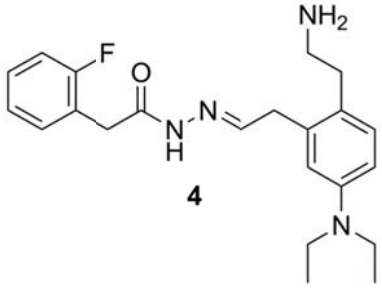




ΔG_{Hyde} = −21 kJ/mol

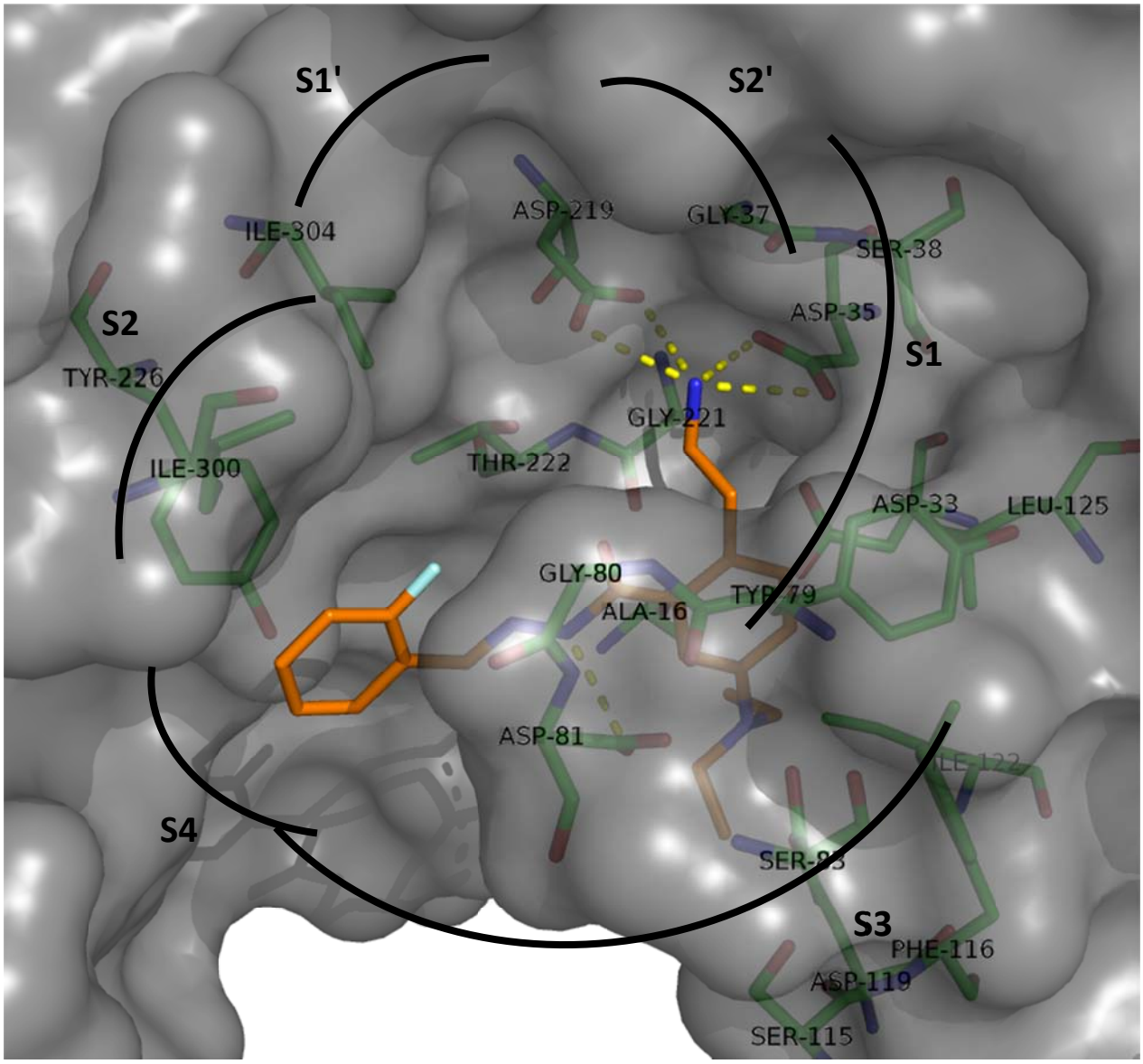
4: (*E*)-*N*'-(2-(2-(2-Aminoethyl)-5-(diethylamino)phenyl)ethylidene)-2-(2-fluorophenyl)acetohydrazide

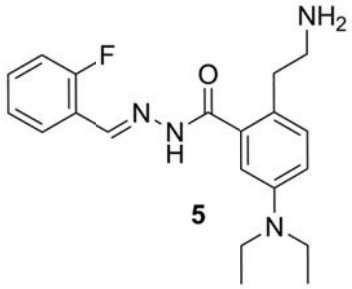




ΔG_{Hyde} = −30 kJ/mol

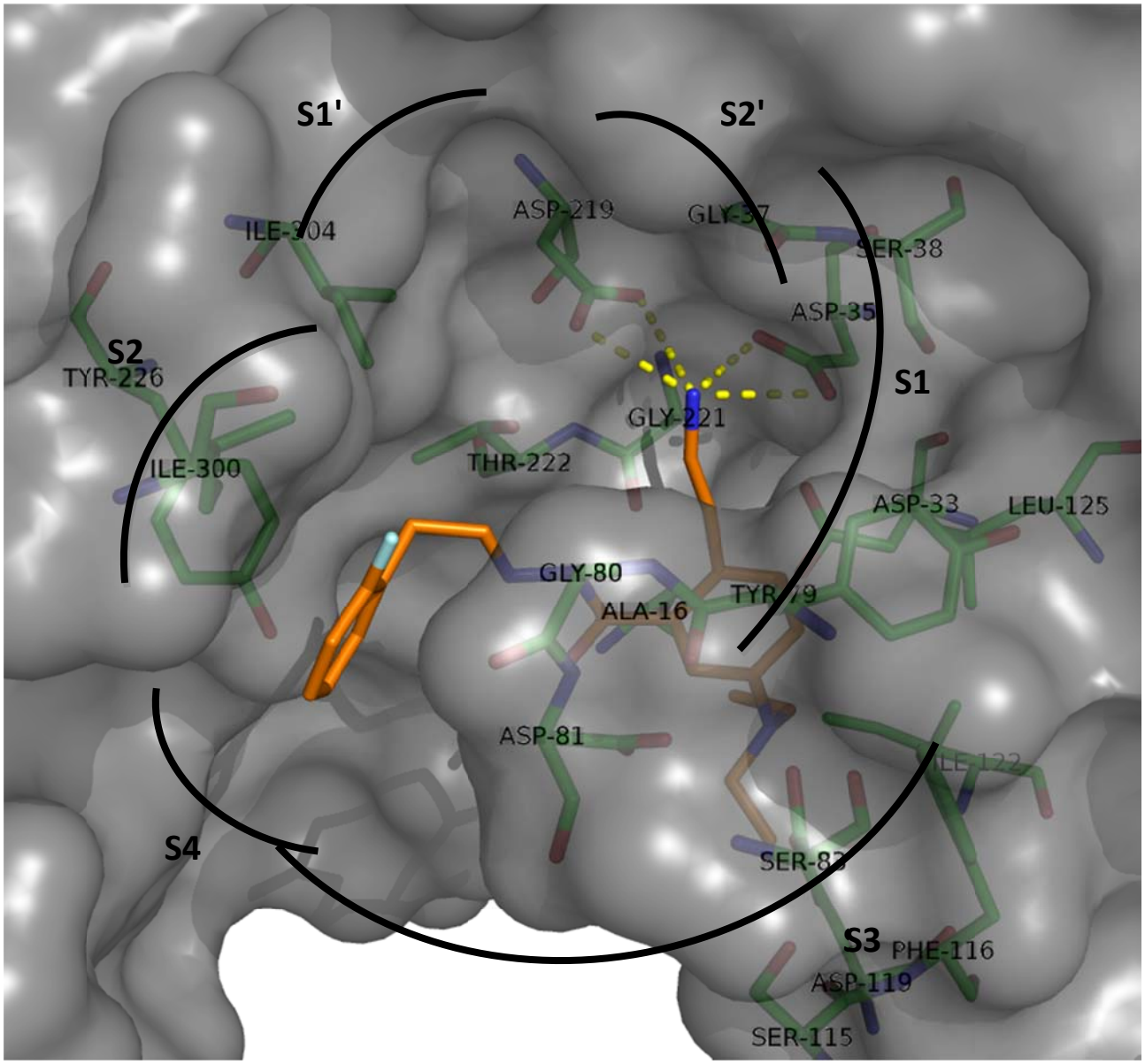
5: (*E*)-2-(2-Aminoethyl)-5-(diethylamino)-*N*'-(2-fluorobenzylidene)benzohydrazide

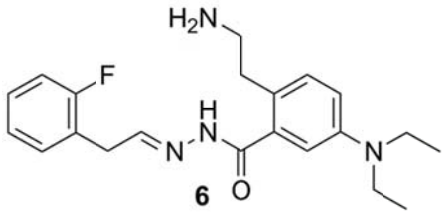




ΔG_{Hyde} = −32 kJ/mol

6: (*E*)-2-(2-Aminoethyl)-5-(diethylamino)-*N*'-(2-(2-fluorophenyl)ethylidene)benzohydrazide

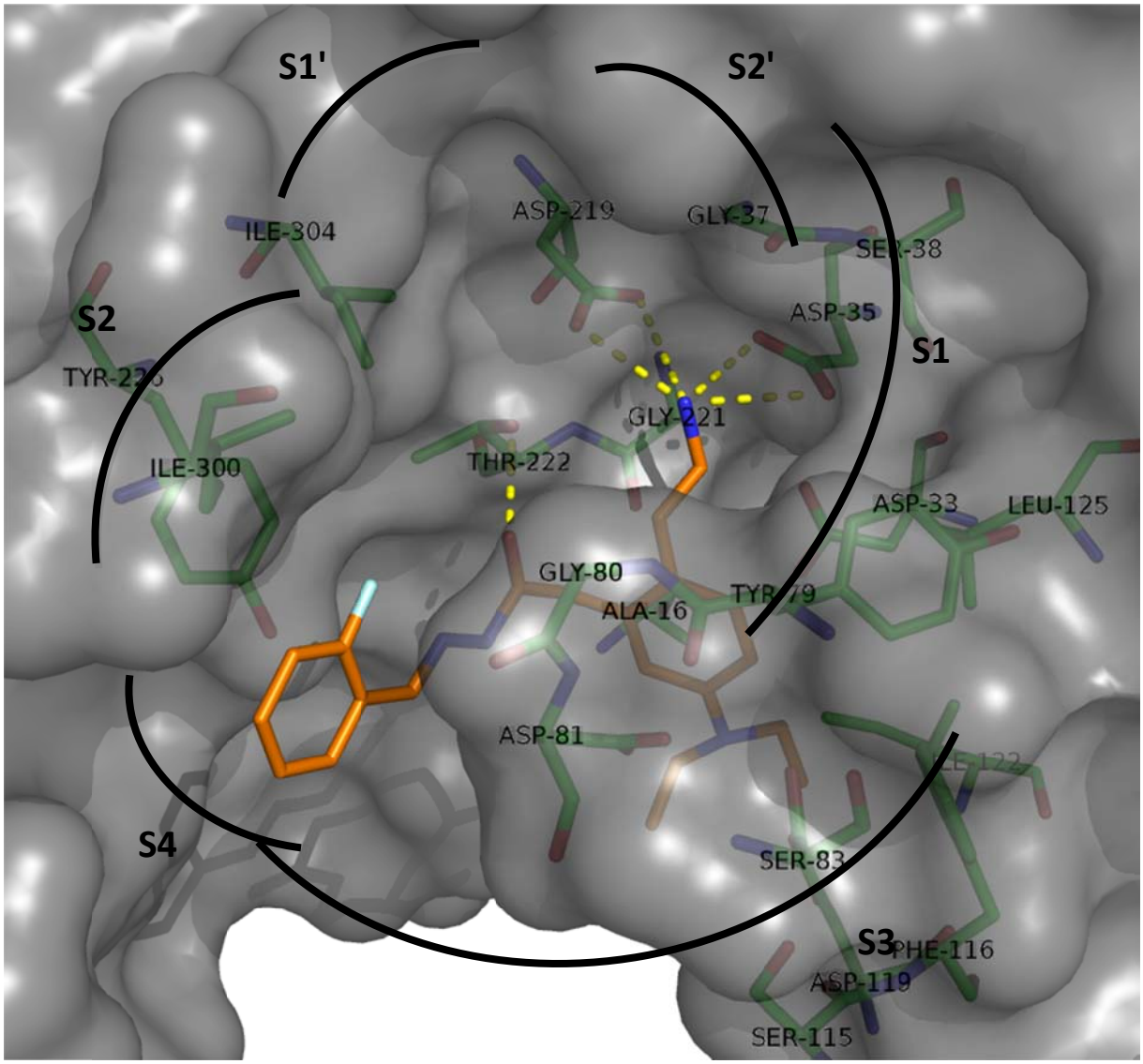


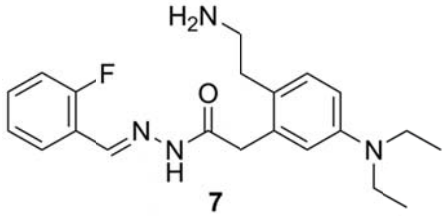


6

ΔG_{Hyde} = −39 kJ/mol

7: (*E*)-2-(2-(2-Aminoethyl)-5-(diethylamino)phenyl)-*N*'-(2-fluorobenzylidene)acetohydrazide

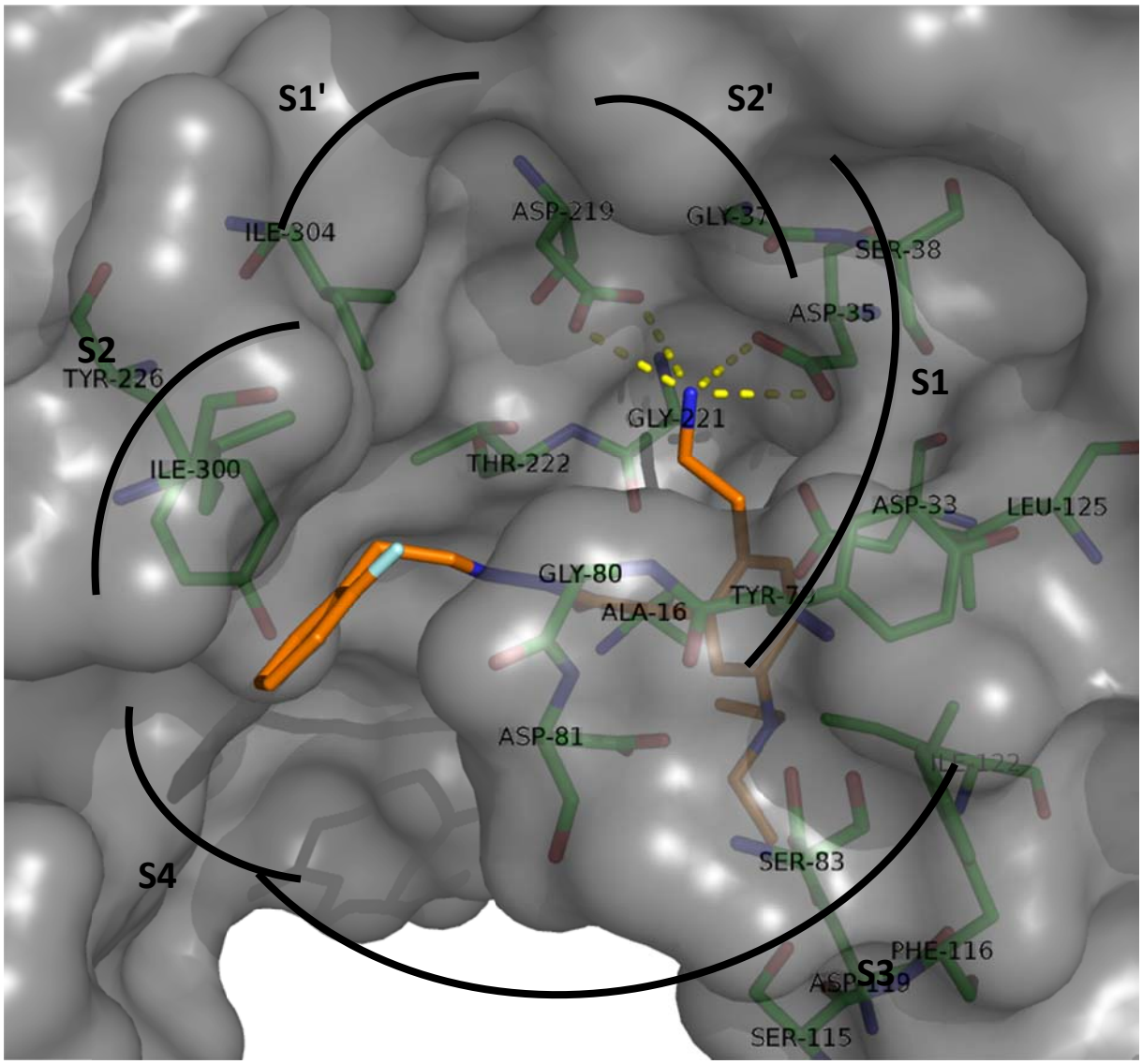


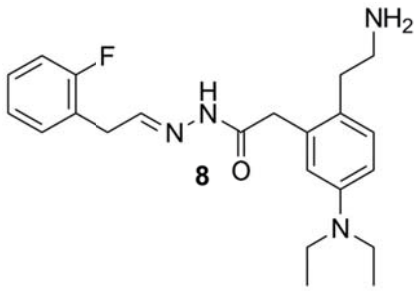


7

ΔG_{Hyde} = −19 kJ/mol

8: (*E*)-2-(2-(2-Aminoethyl)-5-(diethylamino)phenyl)-*N*'-(2-(2-fluorophenyl)ethylidene)acetohydrazide

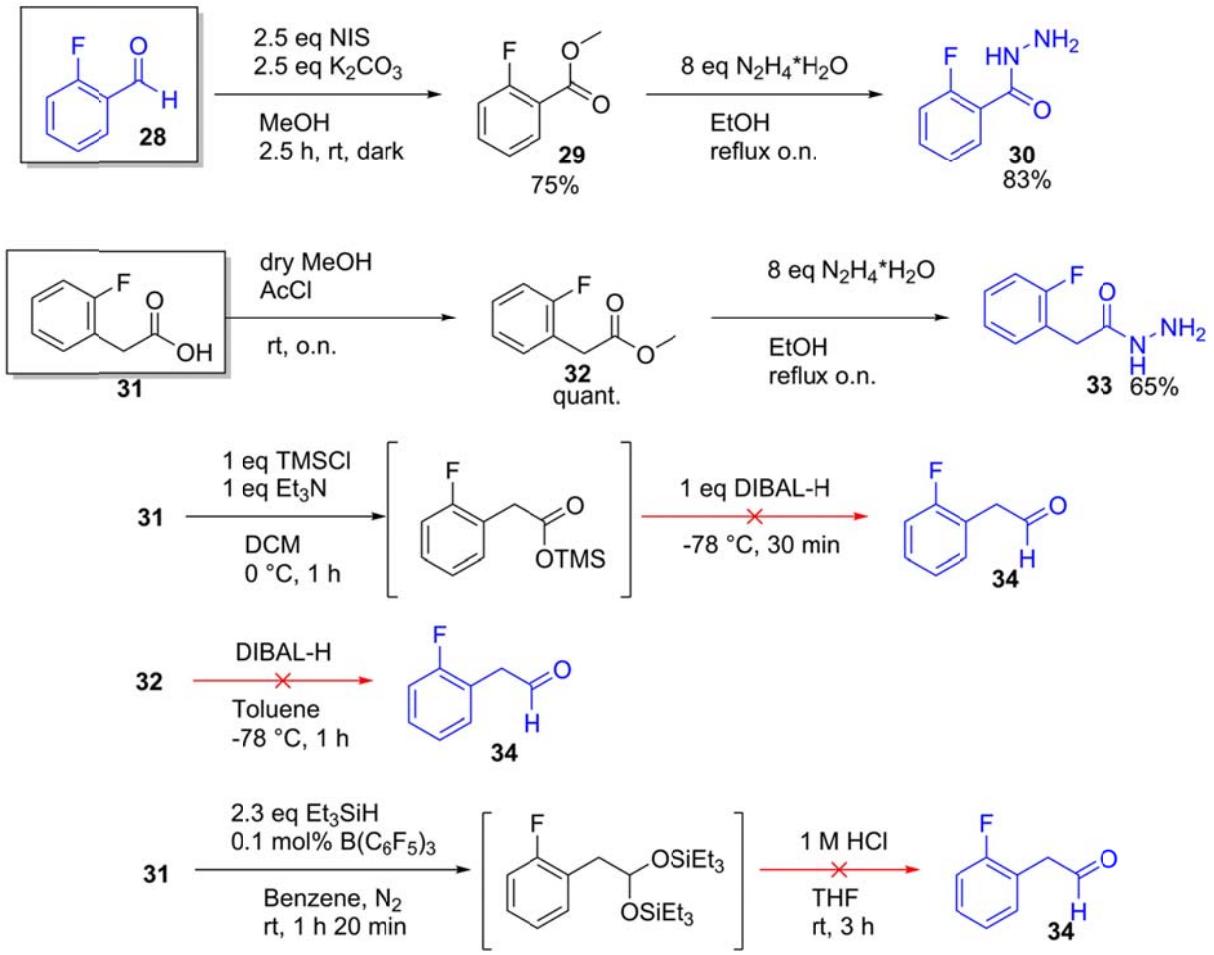




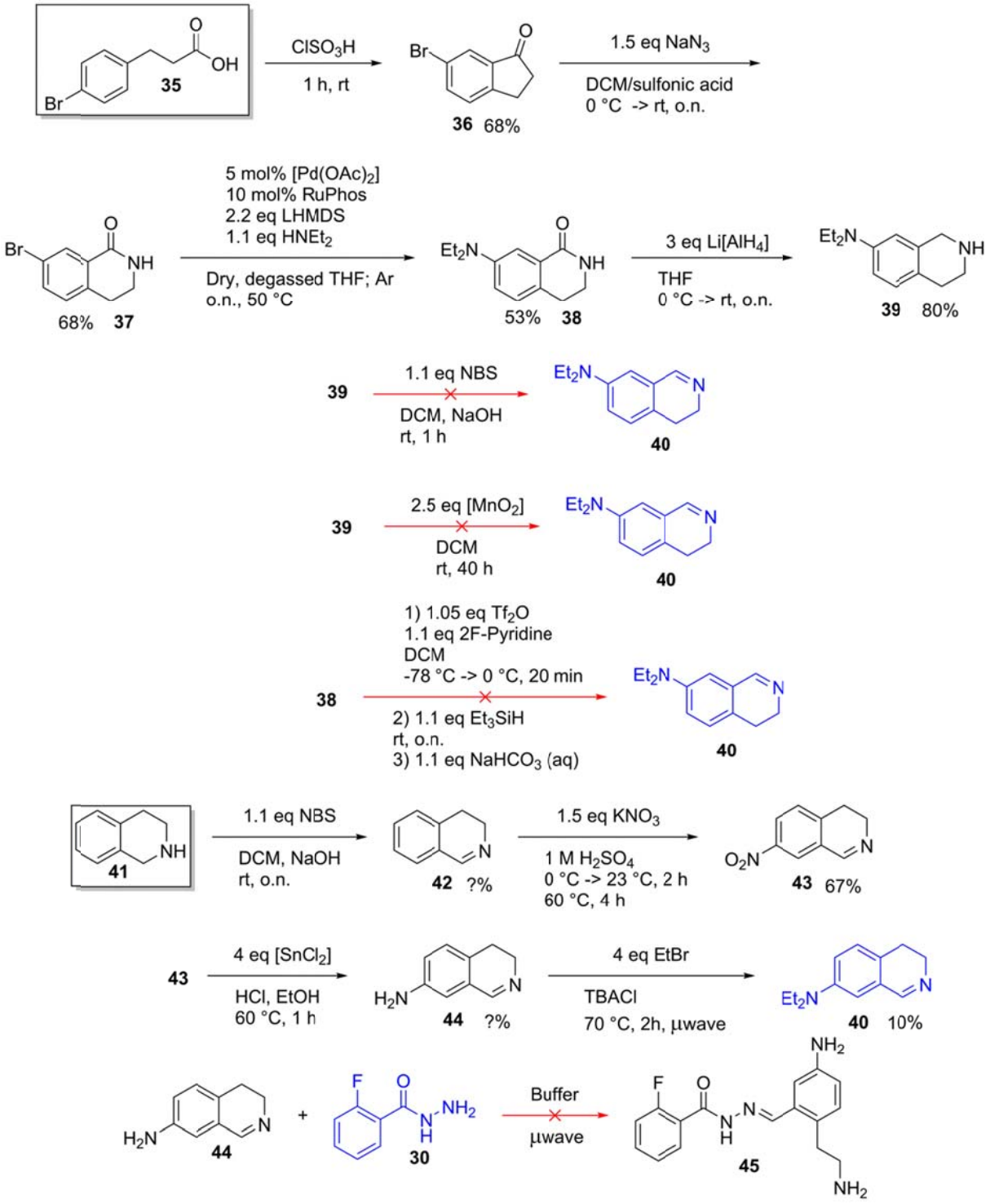
ΔG_{Hyde} = −17 kJ/mol

Synthesis

The planned and attempted synthetic routes are depicted in the following schemes. Commercially obtained starting materials are surrounded with a box, and target compounds are in blue (Scheme S1 and S2). Since not all desired compounds could be obtained, a different library was designed (see Library design).



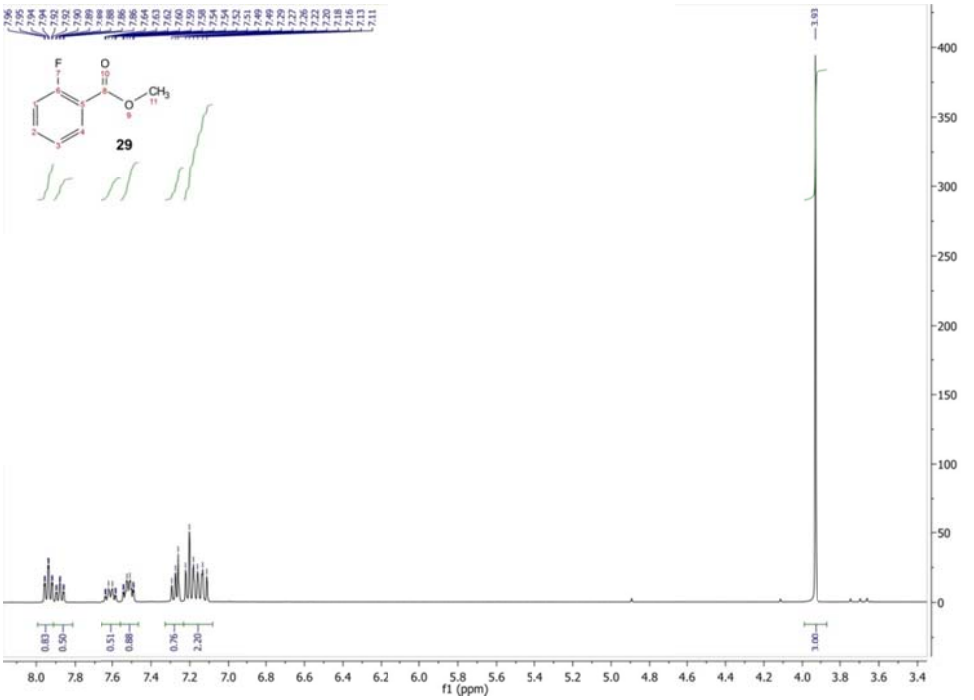
Scheme S1: The used synthetic routes for the fluorophenyl compounds. Commercially obtained compounds are surrounded with a box; target compounds are in blue; failed reactions are shown with a red, crossed arrow. Abbreviations: o.n.: over night; rt: room temperature; quant.: quantitative.

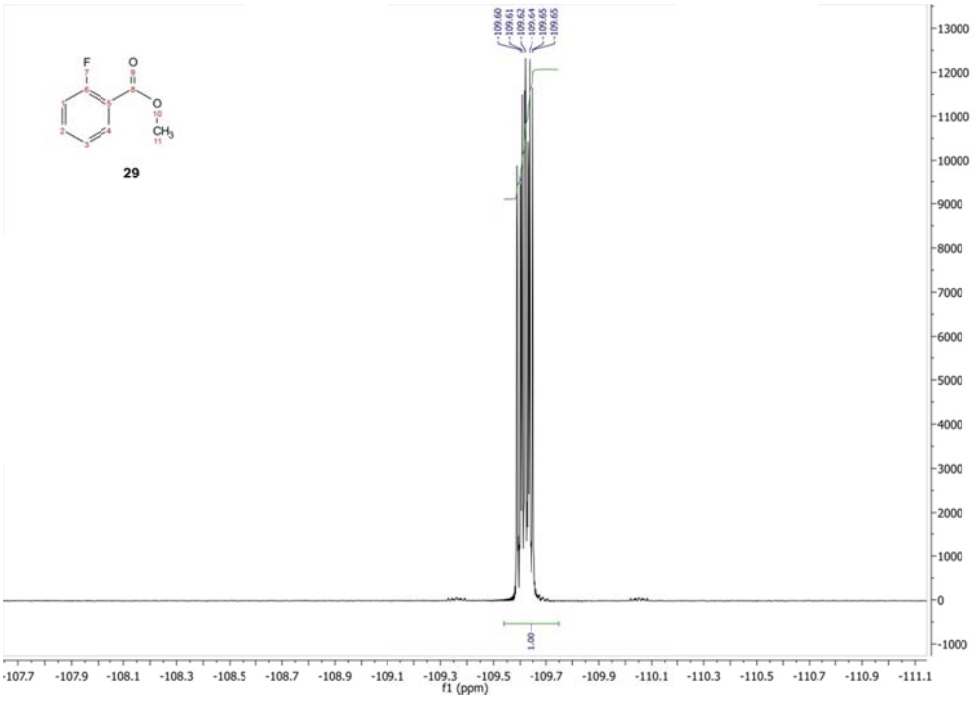


Scheme S2: The used synthetic routes for the 2-(2-aminoethyl)-5-(diethylamino)benzohydrazide compounds. Commercially obtained compounds are surrounded with a box; target compounds are in blue; failed reactions are shown with a red, crossed arrow. Abbreviations: o.n.: over night; rt: room temperature.

Methyl 2-fluorobenzoate (**29**)[32]

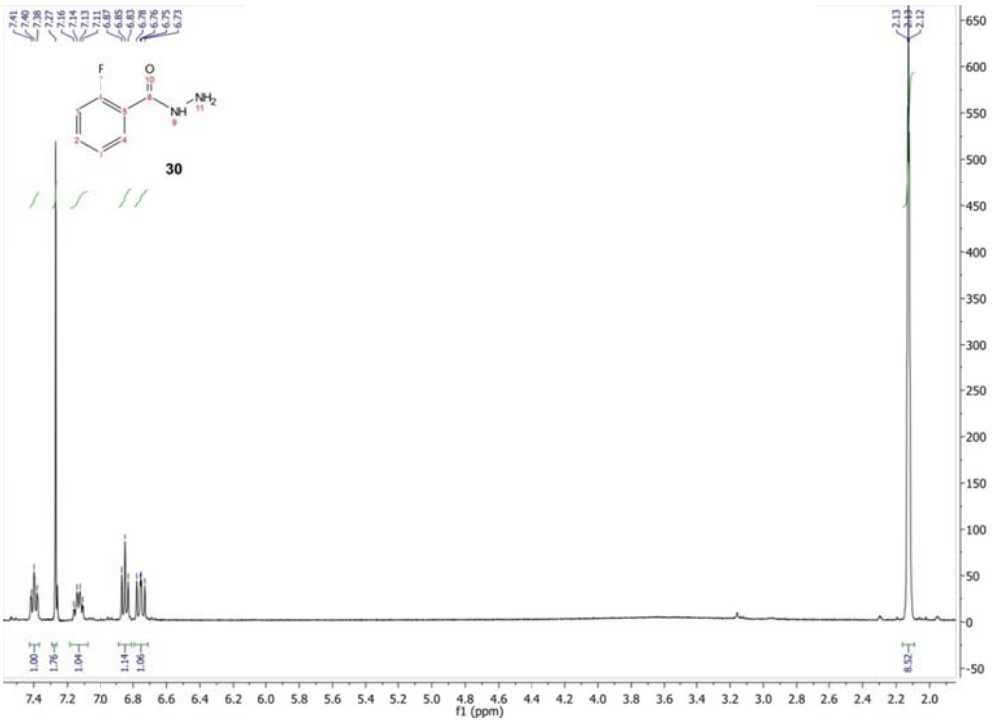
N-Iodosuccinimide (1.4 g, 6.22 mmol, 2.4 eq) and dried K_2CO_3 (1 g, 7.24 mmol, 2.8 eq) were dissolved in dry methanol (25 mL), and **28** (0.41 mL, 0.32 g, 2.56 mmol, 1 eq) was added. The mixture was stirred in the dark for 2.5 h, and the reaction was followed by TLC (SiO_2; 20% ethylacetate in heptane). Upon completion, water (16 mL) and sodium thiosulfate (1.6 g) were added. The resulting mixture was extracted with 50% ether in pentane (4 ×). The organic phase was washed with saturated aqueous NaCl solution, dried over magnesium sulfate and filtered. Evaporation of the solvent *in vacuo* yielded **29** as a yellow liquid (0.44 g, 2.85 mmol, 75%). ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 7.94 (td, 1H), 7.56 – 7.46 (m, 1H), 7.20 (td, 1H), 7.14 (ddd, 1H), 3.93 (s, 3H). ^{19}F-NMR(376 MHz, $CDCl_3$) δ= −109.50 – −109.74 (m, 1F).

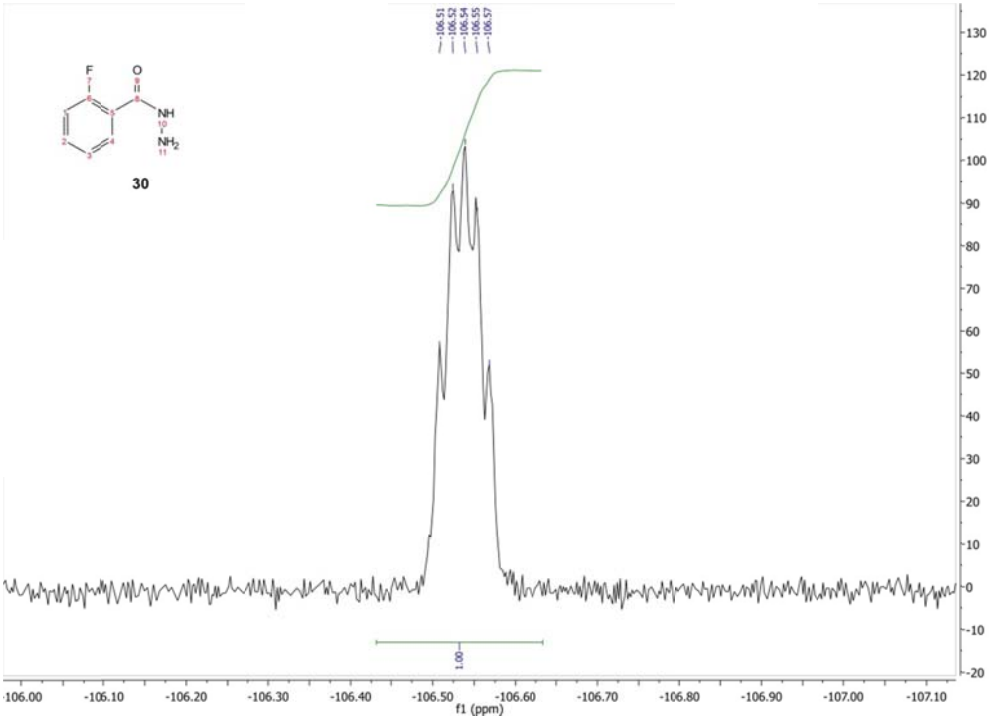




2-Fluorobenzohydrazide (**30**)[19]

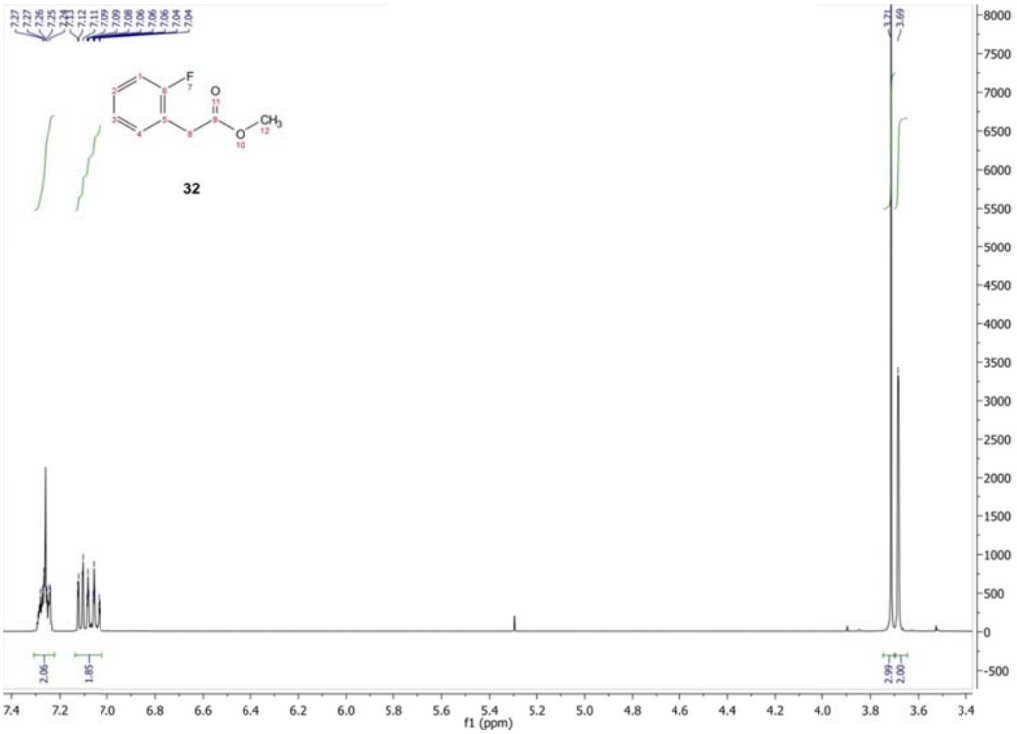
29 (0.4 g, 2.60 mmol, 1 eq) was dissolved in methanol (1.7 mL) and $N_2H_4 \cdot H_2O$ (1.05 g, 21 mmol, 8 eq). Upon addition of hydrazine the solution discolored. The solution was refluxed for 6 h whilst following the reaction with TLC (SiO_2; 30% ethylacetate in heptane). Upon completion a few drops of 2 M HCl (aq) were added and the methanol was removed *in vacuo*. 1.25 M ethanolic HCl (15 mL) was added and the resulting mixture was sonicated. Compounds that did not dissolve were removed by filtration, and the solvent was removed from the resulting solution *in vacuo* yielding **30** (329 mg, 2.13 mmol, 82.5%) as a bright orange solid. ^{1}H-NMR(400MHz, $CDCl_3$): δ= 7.39 (td, 1H), 7.12 (tdd, 1H), 6.84 (td, 1H), 6.79 – 6.69 (m, 1H), 2.12 (p, 2H). ^{19}F-NMR(376 MHz, $CDCl_3$): δ= −106.54 (p, 1F).

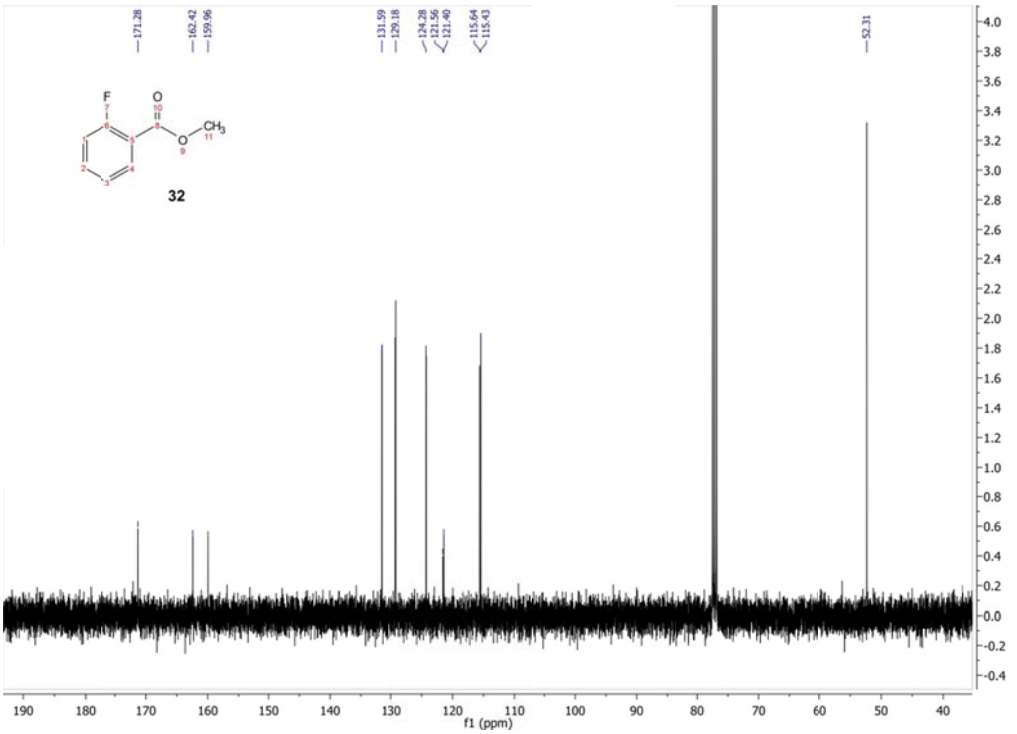


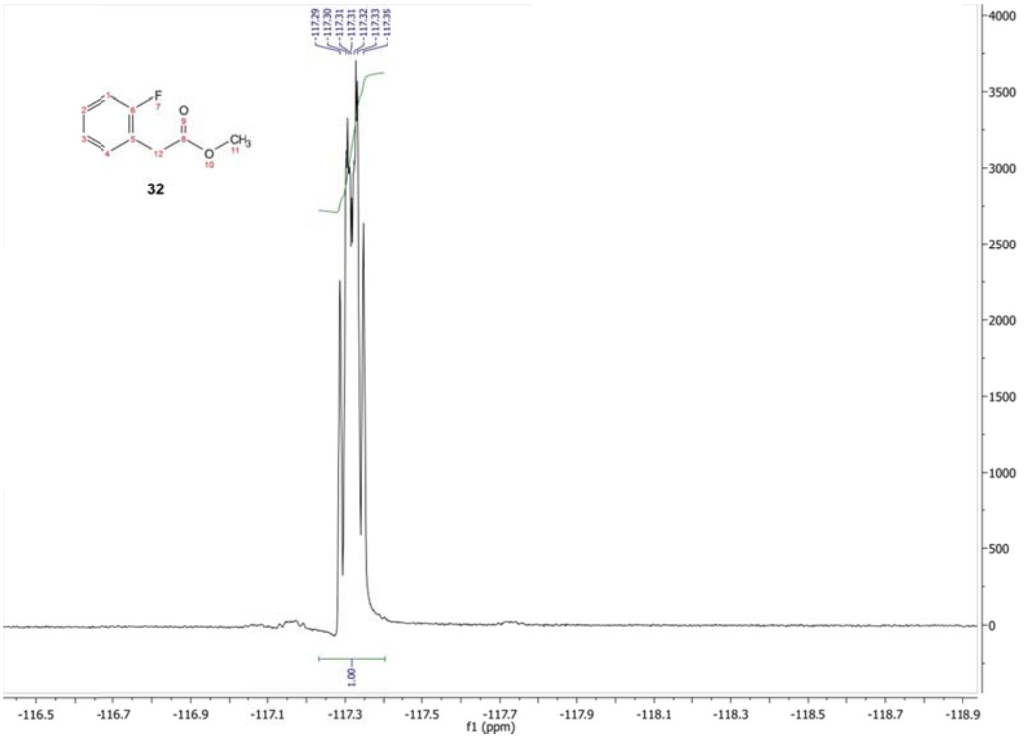


Methyl 2-(2-fluorophenyl)acetate (**32**)*

31 (1 g, 6.49 mmol, 1 eq) was dissolved in dry methanol (30 mL) in a round-bottom flask equipped with a septum and acetyl chloride (2.5 mL, 35.16 mmol, 5.4 eq) was added slowly. The mixture was left to stir overnight after which the conversion was checked with TLC (SiO_2; 20% ethylacetate in heptane). Upon completion the solvent was evaporated *in vacuo* which yielded a quantitative amount of pure **32**. ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 7.21 – 7.12 (m, 2H), 7.05 – 6.90 (m, 2H), 3.61 (s, 3H), 3.58 (s, 2H). ^{13}C-NMR(101 MHz, $CDCl_3$): δ= 171.28, 162.41, 159.96, 131.57, 129.22, 124.26, 121.56, 121.40, 115.64, 115.43, 52.31, 34.44. ^{19}F-NMR(376 MHz, $CDCl_3$): δ= −117.21 – −117.46 (m).

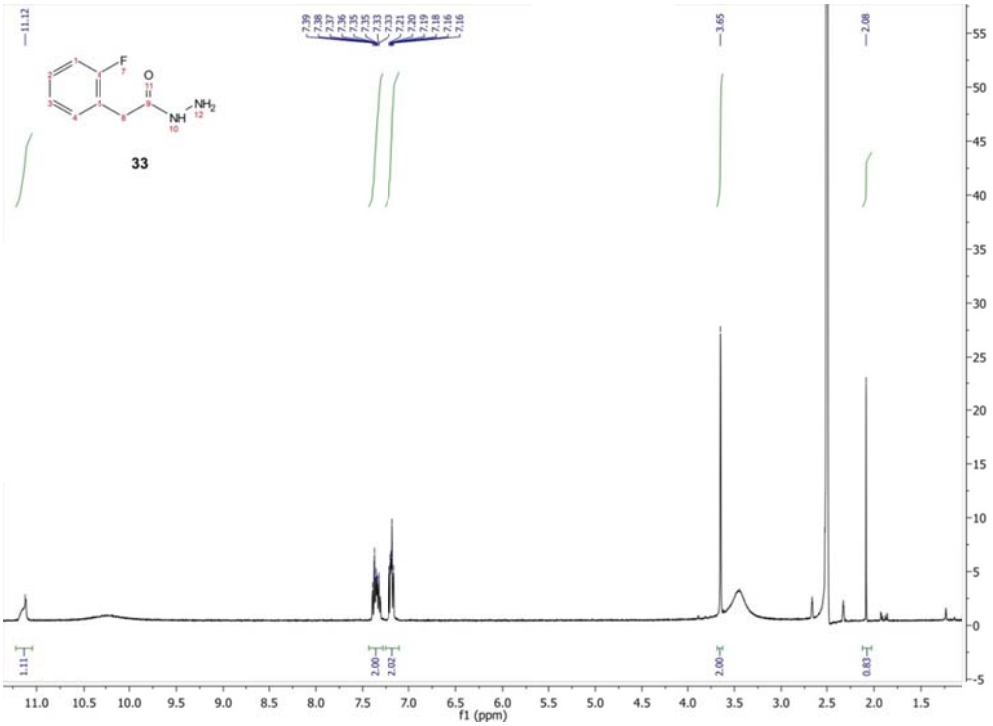


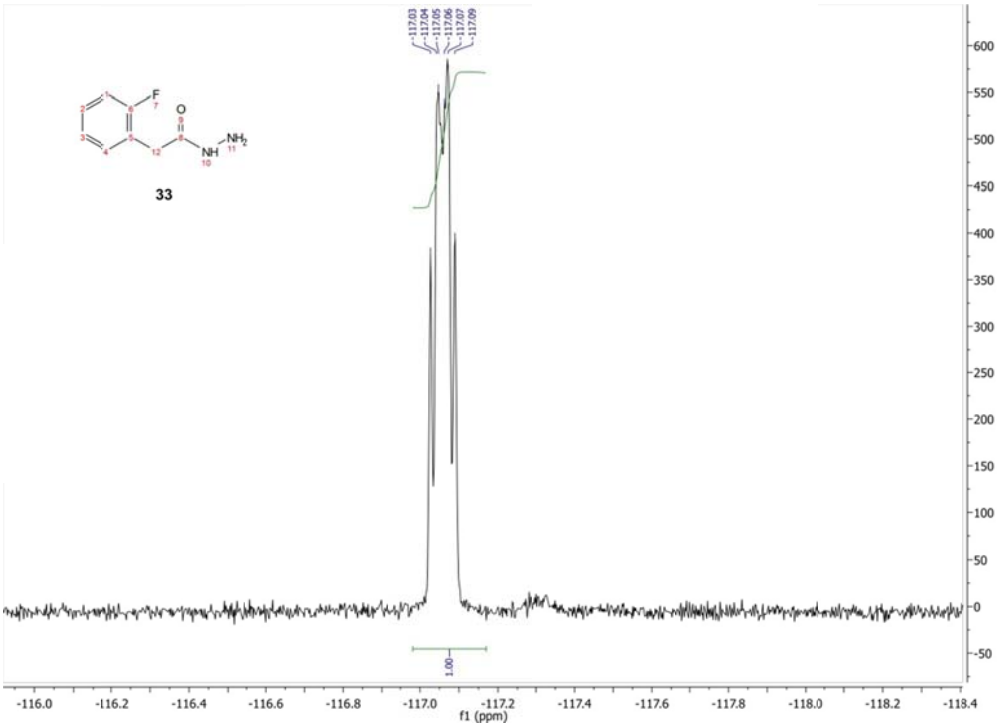



32
-117.29
-117.30
-117.31
-117.31
-117.32
-117.33
-117.35
4000
3500
3000
2500
2000
1500
1000
500
0
1.00
-116.5
-116.7
-116.9
-117.1
-117.3
-117.5
-117.7
-117.9
-118.1
-118.3
-118.5
-118.7
-118.9
f1 (ppm)

2-(2-Fluorophenyl)acetohydrazide (**33**)*

32 (0.7 g, 4.16 mmol, 1 eq) was dissolved in ethanol (1 mL) and $N_2H_4 \cdot H_2O$ (1.61 mL, 1.66 g, 33 mmol, 8 eq). The resulting mixture was refluxed overnight. The conversion was checked with TLC (SiO_2; 30% ethylacetate in heptane). Upon completion a few drops of 2 M HCl (aq) were added and the methanol was removed *in vacuo*. 1.25 M ethanolic HCl (23.3 mL) was added and the resulting mixture was sonicated. Compounds that did not dissolve were removed by filtration, and the solvent was removed from the resulting solution *in vacuo* yielding **30** (454 mg, 2.70 mmol, 65%) as an off-white solid. ^{1}H-NMR(400 MHz, DMSO-d_6): δ= 11.11 (s), 7.42 – 7.29 (m, 2H), 7.22 – 7.14 (m, 2H), 3.65 (s, 2H), 2.08 (s, 1H). ^{19}F-NMR(376 MHz, DMSO-d_6): δ= −117.06 (q, 1F).





2-(2-Fluorophenyl)acetaldehyde (**34**)*

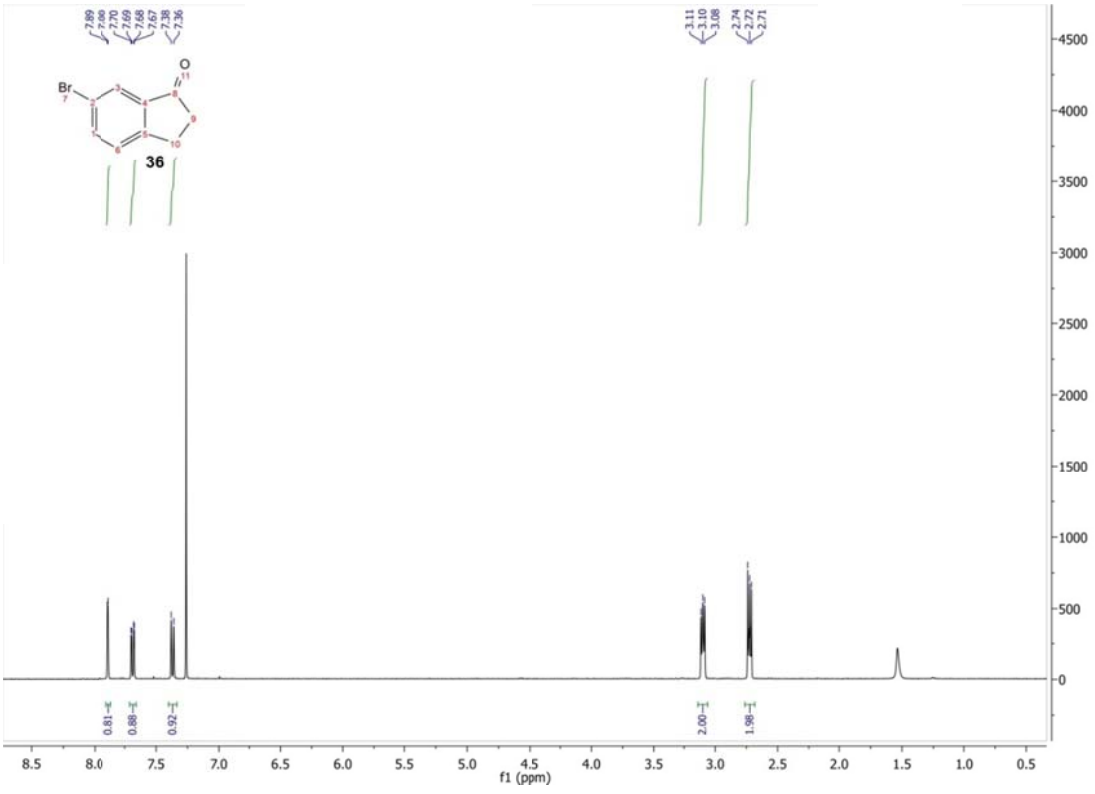
31 (0.2 g, 297 mmol, 1 eq) was dissolved in DCM (2 mL). TMSCl (165 µL, 141 mg, 297 mmol, 1 eq) and trimethylamine (181 µL, 131 mg, 297 mmol, 1 eq) were added at 0°C. The mixture was stirred for 1 h; after which it was cooled to −78°C followed by the addition of 1.2 M DIBAL-H in toluene (1.08 mL). The resulting mixture was stirred for 30 min. No formation of aldehyde **34** was observed by ^{1}H-NMR, instead, only aromatic hydrogen atoms are observed.

34 could not be synthesised by reducing **32** with DIBAL-H in toluene at −78°C. Reduction to the corresponding alcohol was observed.

31 (771 mg, 5 mmol, 1 eq); benzene (3 mL); triethylsilane (1.84 mL, 1.34 g, 11.5 mmol, 2.3 eq); and triperfluorophenylborane (2.56 mg, 5 µmol, 0.001 eq) were added to a dried flask equipped with a nitrogen line. After 1.3 h the crude reaction mixture was dried *in vacuo*. THF (20 mL) was added followed by 1 M HCl (aq, 20 mL) whilst stirring vigorously for 3 h at room temperature. The resulting mixture was extracted three times with ether. The organic layers were combined, dried over magnesium sulfate, filtered and the solvent was removed *in vacuo*. The product was purified using column chromatography (SiO_2; 2% ethylacetate in pentane). No product formation was observed.

5-Bromo-2,3-dihydro-1*H*-inden-1-one (**36**)[33]

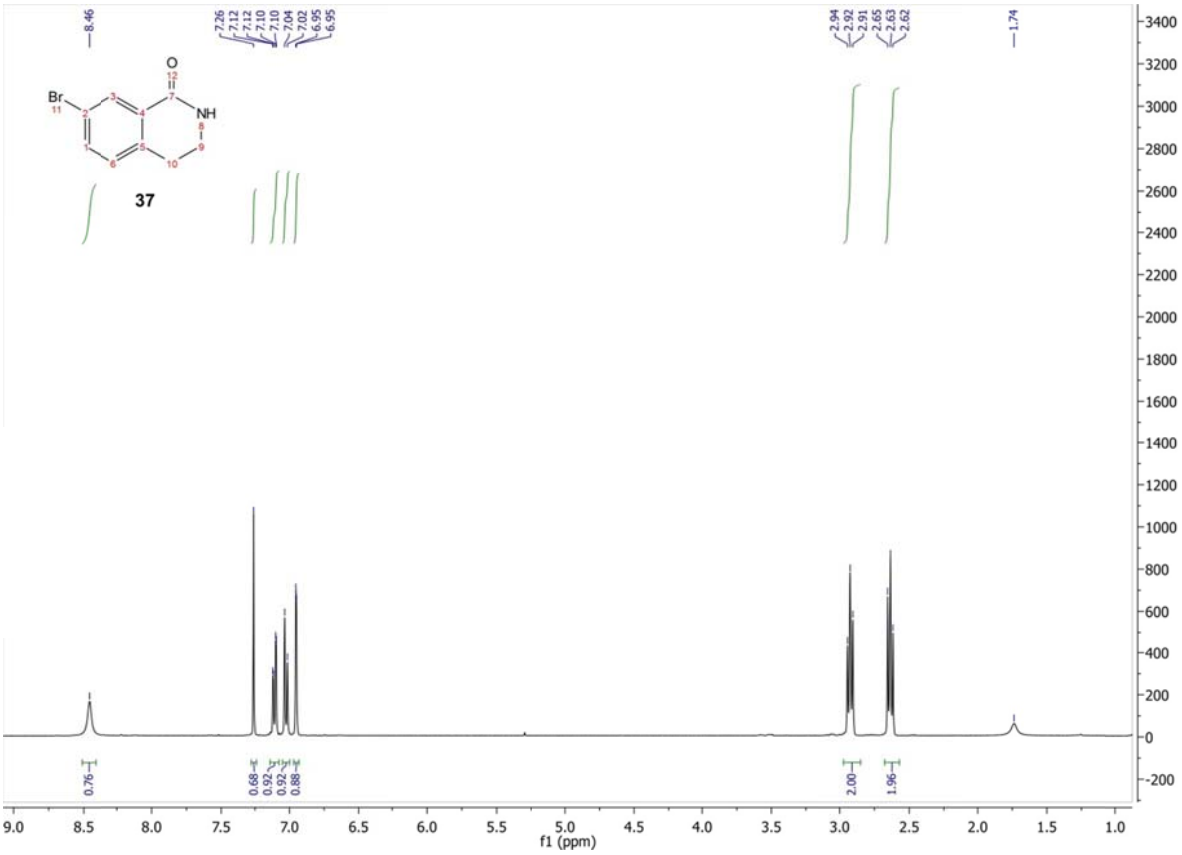
To **35** (5.27 g, 23 mmol, 1 eq) was added chlorosulfonic acid (40 mL, 70.1 g, 602 mmol, 26 eq) and the solution was stirred for 1 h at 0°C after which the reaction was quenched with ice-cold water (70 mL) and extracted with DCM. The organic layer was washed with concentrated sodium bicarbonate (aq), dried over sodium sulfate, filtered and the solvent was removed *in vacuo*, which yielded **36** (3.2 g, 15.3 mmol, 68%) as white needles. ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 7.88 (d, 1H), 7.69 (dd, 1H), 7.37 (d, 1H), 3.10 (t, 2H), 2.77 – 2.69 (m, 2H).



* Reaction performed by Milon Mondal.

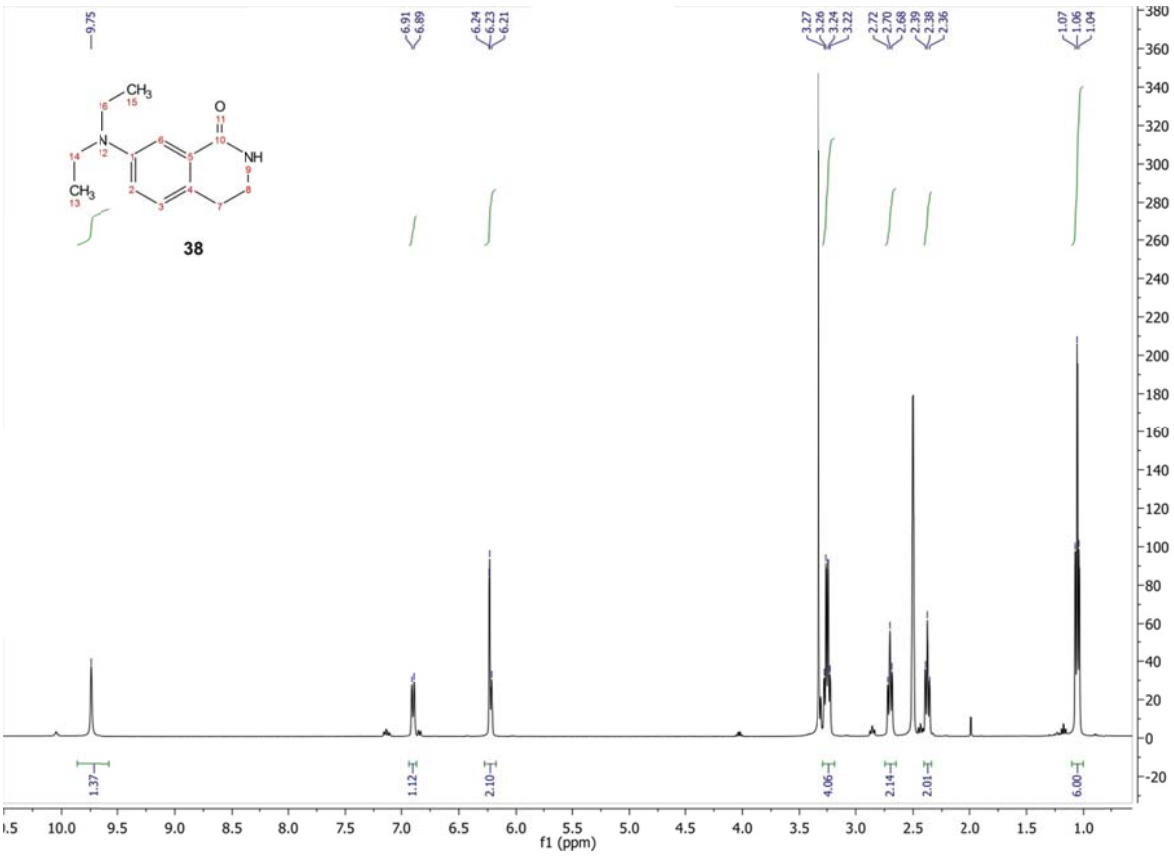
6-Bromo-3,4-dihydroisoquinolin-1(2*H*)-one (**37**)[34]

To a solution of **36** (3 g, 14.2 mmol, 1 eq) in DCM (84 mL) and methansulfonic acid (42 mL) at 0°C sodium azide (1.39 g, 21.32 mmol, 1.5 eq) was added slowly. The resulting mixture was allowed to warm to room temperature and stirred overnight. The mixture was partitioned between DCM and 1 M sodium hydroxide (aq, 50 mL). The aqueous layer was extracted with DCM. The combined organic layers were washed sequentially with water and saturated aqueous NaCl solution, dried over sodium sulfate, filtered, and concentrated *in vacuo*. The resulting solid was recrystallised from ethylacetate or DCM. Four batches yielded a total of 2.16 g of **37** (9.55 mmol, 68%) as white crystals. ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 8.46 (s, 1H), 7.11 (dd, 1H), 7.05 – 6.94 (m, 2H), 2.93 (t, 2H), 2.64 (dd, 2H).



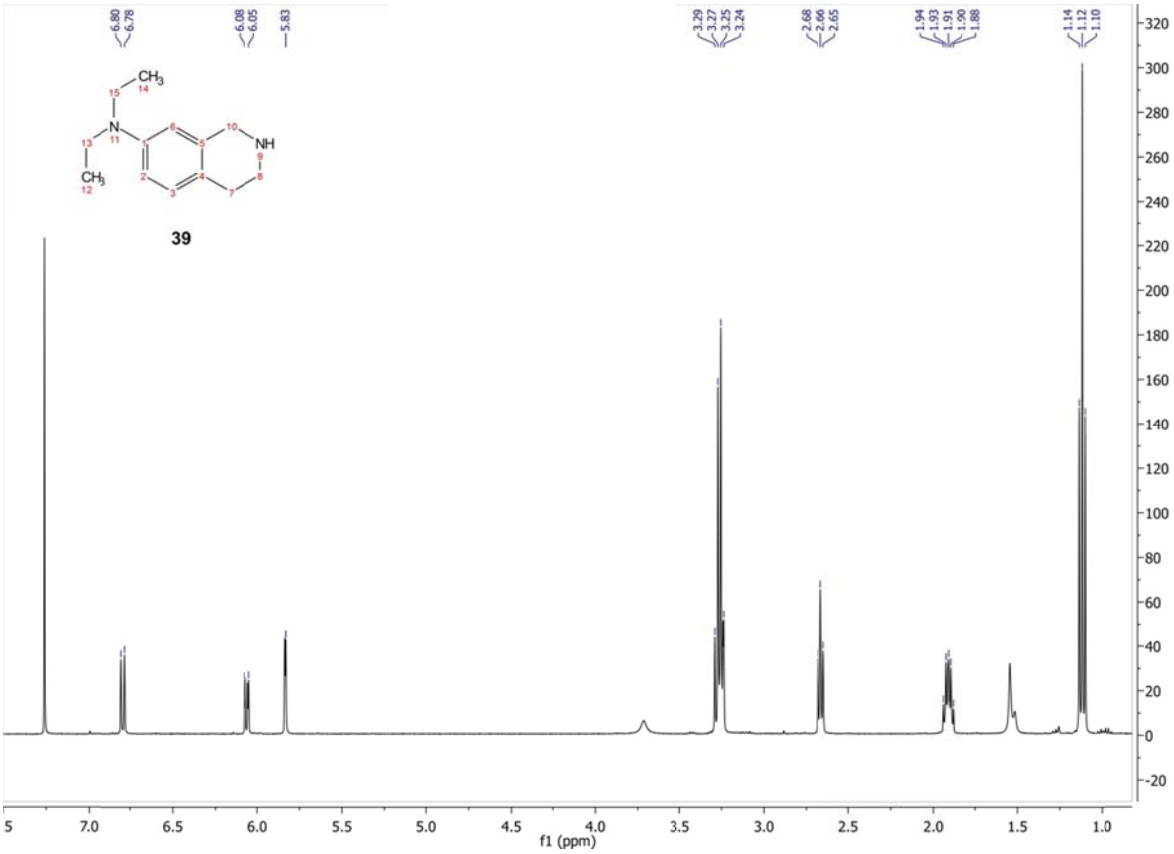
6-(Diethylamino)-3,4-dihydroisoquinolin-1(2*H*)-one (**38**)[35]

An oven-dried Schlenk flask equipped with a magnetic stirring bar and a septum was charged with **37** (900 mg, 3.98 mmol, 1 eq), palladium(II) acetate (45 mg, 0.2 mmol, 0.05 eq) and 2-dicyclohexylphosphino-2′,6′-diisopropoxybiphenyl (RuPhos, 185.4 µg, 0.42 mmol, 0.1 eq). The flask was placed under an Argon atmosphere, and 1 M lithium bis(trimethylsilyl)amide in THF (LHMDS, 8.748 mL, 8.75 mmol, 2.2 eq) and diethylamine (0.453 mL, 17.77 g, 243 mmol, 61 eq) were added *via* a syringe. The solution was heated to 50°C and stirred for 24 h whilst following the reaction with TLC (SiO_2; 50% ethylacetate in pentane). After cooling the mixture to room temperature it was diluted with ethylacetate, washed with water, and concentrated *in vacuo*. This yielded **38** (458 mg, 2.1 mmol, 53%) as an orange solid. The product was used without further purification. ^{1}H-NMR(400 MHz, DMSO-d_6): δ= 9.74 (s, 1H), 6.90 (d, 1H), 6.28 – 6.12 (m, 2H), 3.25 (q, 4H), 2.70 (t, 2H), 2.37 (t, 2H), 1.05 (t, 6H).



N,*N*-Diethyl-1,2,3,4-tetrahydroisoquinolin-6-amine (**39**)*

38 (358 mg, 1.62 mmol, 1 eq) was added to a solution of $LiAlH_4$ (650 mg, 4.86 mmol, 3 eq) in THF (35 mL) over the course of 30 min. The mixture was gently refluxed overnight, cooled in an ice bath. Water (3.6 mL) was added dropwise, followed by 4 M sodium hydroxide (aq**,** 3.6 mL), and water (10.8 mL). The precipitate was removed and washed with THF. The filtrate was evaporated. This yielded **39** (270 mg, 1.32 mmol, 80%). ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 6.80 (d, 1H)**,** 6.07 (d, 1H), 5.84 (s, 1H), 3.29 – 3.19 (m, 4H), 2.67 (t, 2H), 1.97 – 1.85 (m, 2H), 1.12 (t, 6H).



* Reaction performed by Milon Mondal.

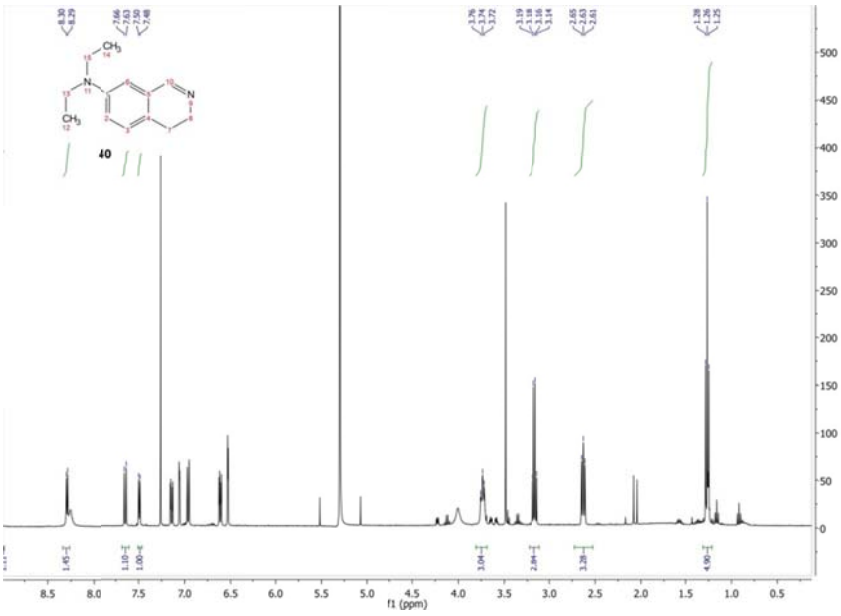
N,*N*-Diethyl-3,4-dihydroisoquinolin-6-amine (**40**)*

39 (10 mg, 0.049 mmol, 1 eq) was dissolved in DCM (100 µL). *N*-Bromosuccinimide (19.6 mg, 0.054 mmol, 1.1 eq) was added and the reaction mixture was stirred for 30 min at room temperature. To this mixture 30% NaOH (aq, 33 µL, 5 eq) was added and the resulting biphasic mixture was stirred vigorously for 1 h. The organic layer was washed with water (131 µL) and 1 M HCl (aq, 131 µL). The acidic extracts were made alkaline to pH 10 by addition of 1 M NaOH (aq) and extracted with DCM. The organic layers were washed with saturated aqueous NaCl solution, dried over magnesium sulfate and filtered. The solvent was removed *in vacuo*. No product was formed according to ^{1}H-NMR.

A mixture of **39** (50 mg, 0.245 mmol, 1 eq) and MnO_2 (53 mg, 0.61 mmol, 2.5 eq) in DCM (5 mL) was stirred at room temperature for 40 h. The mixture was filtered, and the filtrate evaporated. This yielded over-oxidised product. No improvement was observed with shorter reaction times.

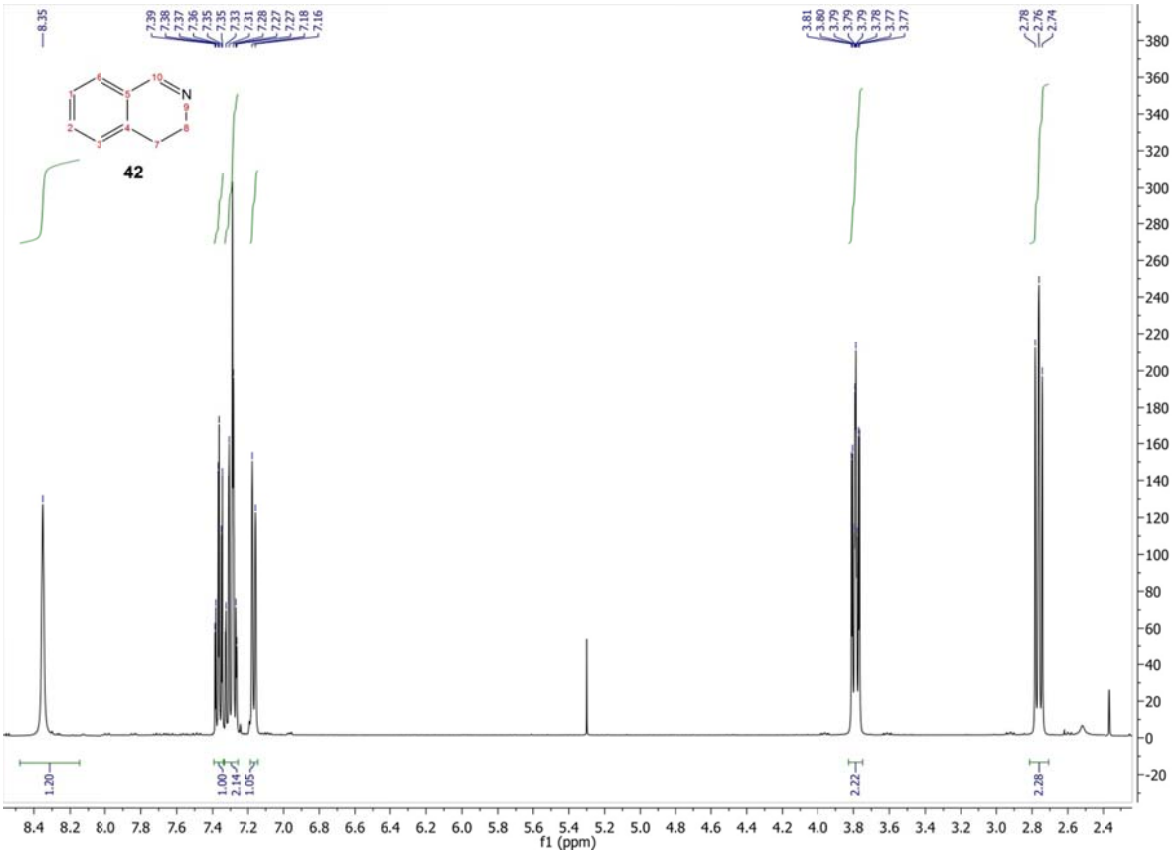
To a flame-dried flask equipped with a septum **38** (44 mg, 0.2 mmol, 1 eq) was added and was dissolved in dry DCL (0.8 mL). 2-Fluoropyridine (19 µL, 21 mg, 0.22 mmol, 1.1 eq) was added. The solution was cooled to −78°C and stirred for 10 min. Triflic anhydride (35 µL, 59 mg, 0.21 mmol, 1.05 eq) was added dropwise *via* a syringe and the reaction was stirred for 10 min. The mixture was then heated to 0°C and stirred for an additional 10 min. Triethylsilane (35 µL, 26 mg, 0.22 mmol, 1.1 eq) was added dropwise and the mixture was stirred for 10 min. The solution was heated to room temperature and stirred for 5 h. The reaction was quenched by addition of saturated sodium bicarbonate (aq, 0.2 mL) and diluted with DCM (0.8 mL). The layers were separated, and the aqueous layer was extracted with DCM (2 ×). The organic layers were combined and dried over sodium sulfate, filtered, and the solvent was removed *in vacuo*. No product was observed by DART-MS.

44 (50 mg, 0.342 mmol, 1 eq), tertbutylammonium chloride (350 mg, 1.26 mmol, 3.7 eq), and ethylbromide (149 mg, 1.369 mmol, 4 eq) were added to a flask and stirred at 75°C overnight. The mixture was diluted with concentrated sodium bicarbonate (aq) and extracted with ethylacetate. This yielded 10% of **40**. ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 9.01 (s, 1H), 8.29 (d, 1H), 7.64 (d, 1H), 7.49 (d, 1H), 3.74 (t, 2H), 3.17 (q, 4H), 2.63 (t, 2H), 1.26 (t, 6H).



3,4-Dihydroisoquinoline (**42**)*

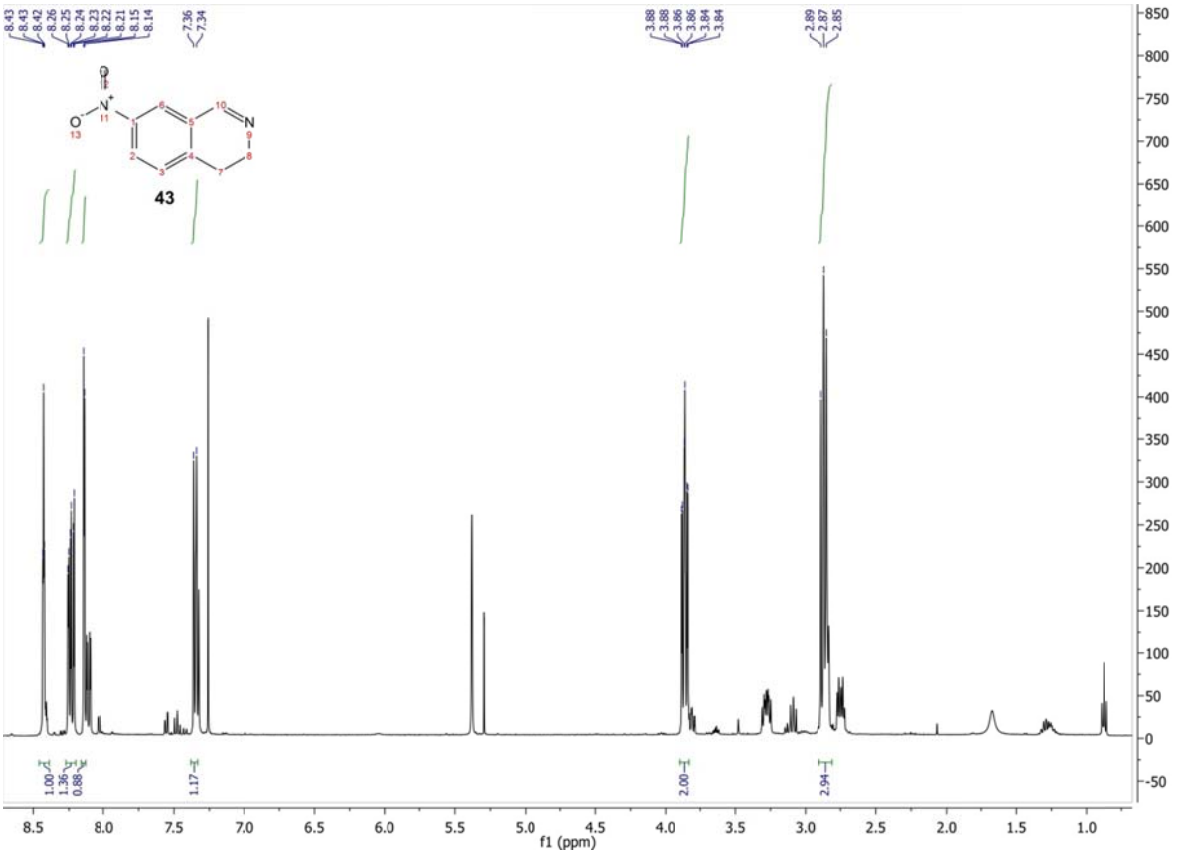
41 (5 g, 37.5 mmol, 1 eq) was dissolved in DCM (75 mL). *N*-Bromosuccinimide (7.5 g, 41.25 mmol, 1.1 eq) was added slowly and the resulting mixture was stirred for 30 min at room temperature. 30% NaOH (aq, 25 mL) was added and the resulting biphasic mixture was stirred vigorously overnight. The organic layer was washed with water and 1 M HCl (aq, 100 mL). The acidic extracts were made alkaline to pH 10 by addition of 1 M NaOH (aq) and extracted with DCM. The combined organic layers were washed with saturated aqueous NaCl solution, dried over magnesium sulfate and filtered. The solvent was removed *in vacuo*, and the product was purified using flash column chromatography (SiO_2; 6.25% methanol in DCM). ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 8.35 (s, 1H), 7.37 (td, 1H), 7.34 – 7.25 (m, 2H), 7.17 (d, 1H), 3.83 – 3.74 (m, 2H), 2.76 (q, 2H).



* Reaction performed by Milon Mondal.

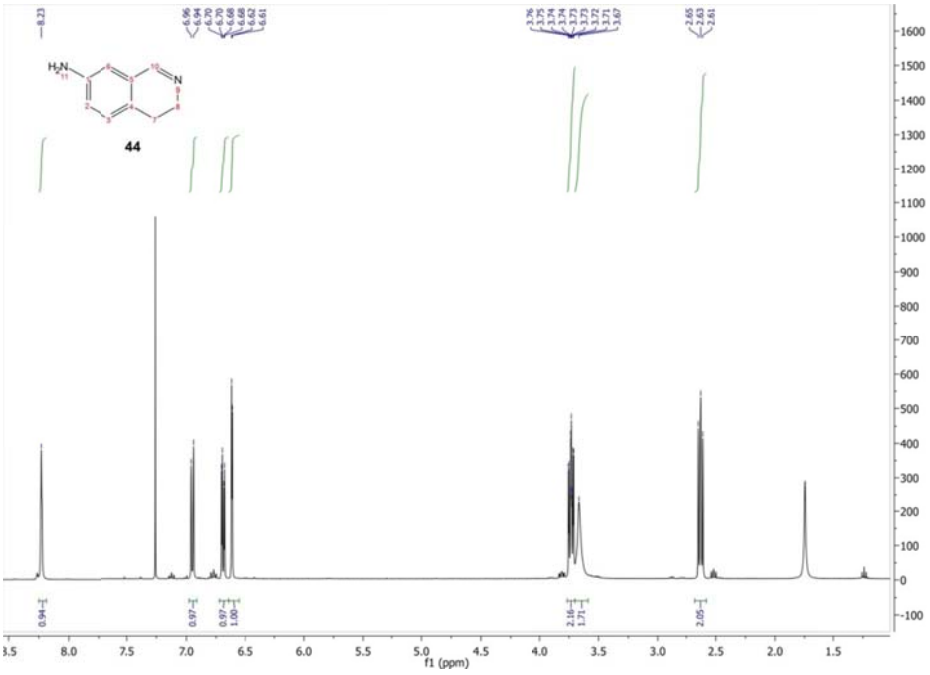
6-Nitro-3,4-dihydroisoquinoline (**43**)*

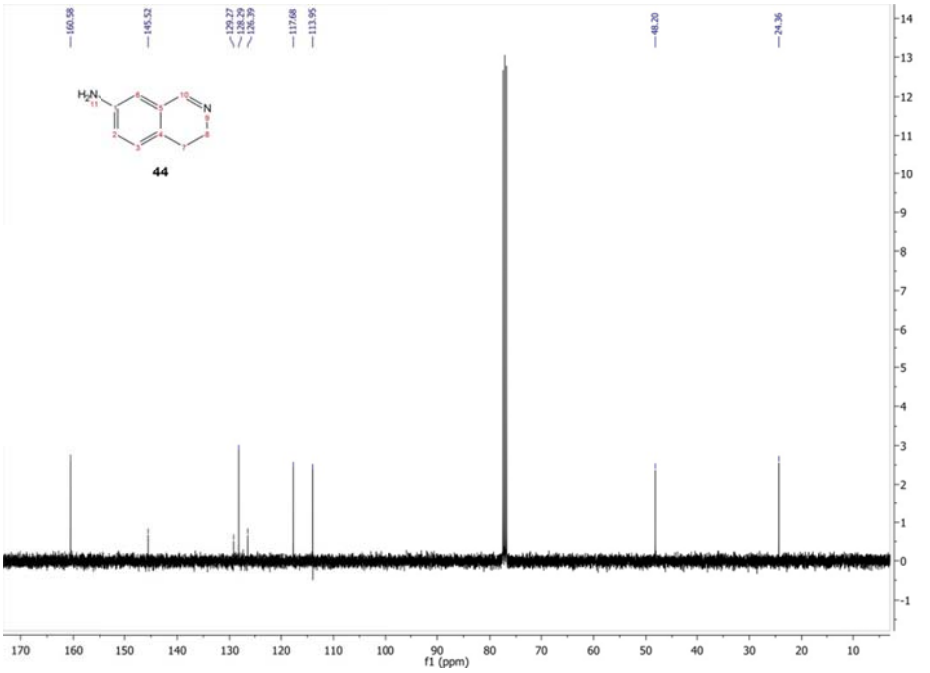
Oven dried KNO_3 (2.02 g, 29.6 mmol, 1.5 eq) was dissolved in 1 M sulfuric acid (aq, 20 mL). To this mixture **42** (2.6 g, 19.82 mmol, 1 eq) was added at 0°C and the resulting mixture was slowly warmed to room temperature over 2 h. It was then immersed in a preheated oil bath at 60°C for 4 h, after which the mixture was made alkaline to pH 10 by addition of 3 M NaOH (aq) and extracted with DCM. The organic layers were washed with saturated aqueous NaCl solution, dried over magnesium sulfate and filtered. The solvent was removed *in vacuo*. The product was purified using column chromatography (SiO_2; 50% ethylacetate in hexane). This yielded 67% product, and 33% hydrated product. ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 8.43 (t, 1H), 8.27 – 8.19 (m, 1H), 8.14 (d, 1H), 7.35 (d, 1H), 3.91 – 3.82 (m, 2H), 2.87 (t, 2H).



3,4-Dihydroisoquinoline-6-amine (**44**)*

43 (0.1 g, 0.56 mmol, 1 eq) was dissolved in ethanol (25 mL) and heated to 60°C. This solution was treated with a solution of tin(II) chloride dehydrate (0.5 g, 2.28 mmol, 4 eq) in concentrated HCl (aq, 2 mL) and heated to 60°C for 1 h. After cooling to room temperature the mixture was poured onto ice-cold water (17 mL) and made alkaline to pH 9 with potassium hydroxide pellets, liberating an oily residue. The residue was extracted into DCM and dried *in vacuo*. The product was purified by column chromatography (SiO_2; 0.5% ammonia and 4.5% methanol in DCM). This yielded **44** as a dark yellow oil. ^{1}H-NMR(400 MHz, $CDCl_3$): δ= 8.23 (t, 1H), 6.95 (d, 1H), 6.69 (dd, 1H), 6.62 (d, 1H), 3.77 – 3.69 (m, 2H), 3.67 (s, 2H), 2.63 (t, 2H). ^{13}C-NMR(101 MHz, $CDCl_3$): δ= 160.59, 145.52, 129.27, 128.29, 126.39, 117.68, 113.95, 48.20, 24.37.





* Reaction perfomed by Milon Mondal.

N'-(4-Amino-2-(2-aminoethyl)benzylidene)-2-fluorobenzohydrazide (**45**)*

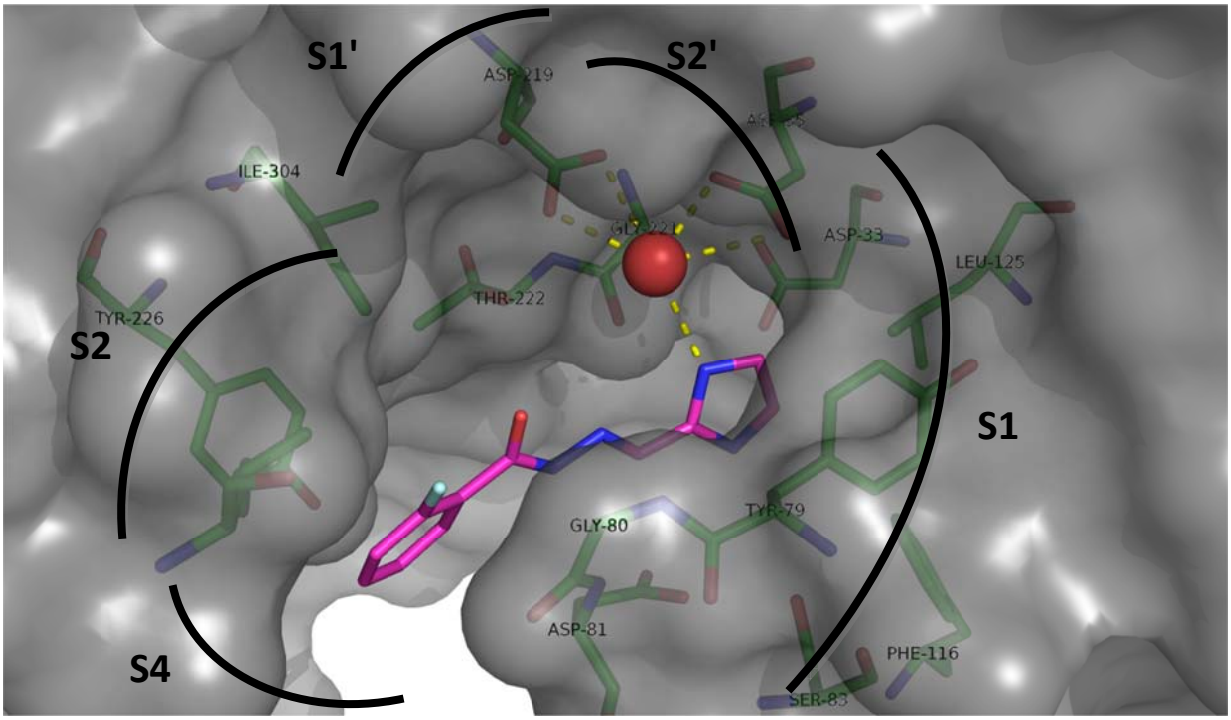
44 (50 mg, 0.324 mmol, 1 eq) and **30** (71 mg, 0.389 mmol, 1.2 eq) were dissolved in sodium acetate buffer (0.6 mL, 0.1 M, pH 4.6). The mixture was heated in a microwave, but no product formation was observed.

Library B — Compounds 10 to 18

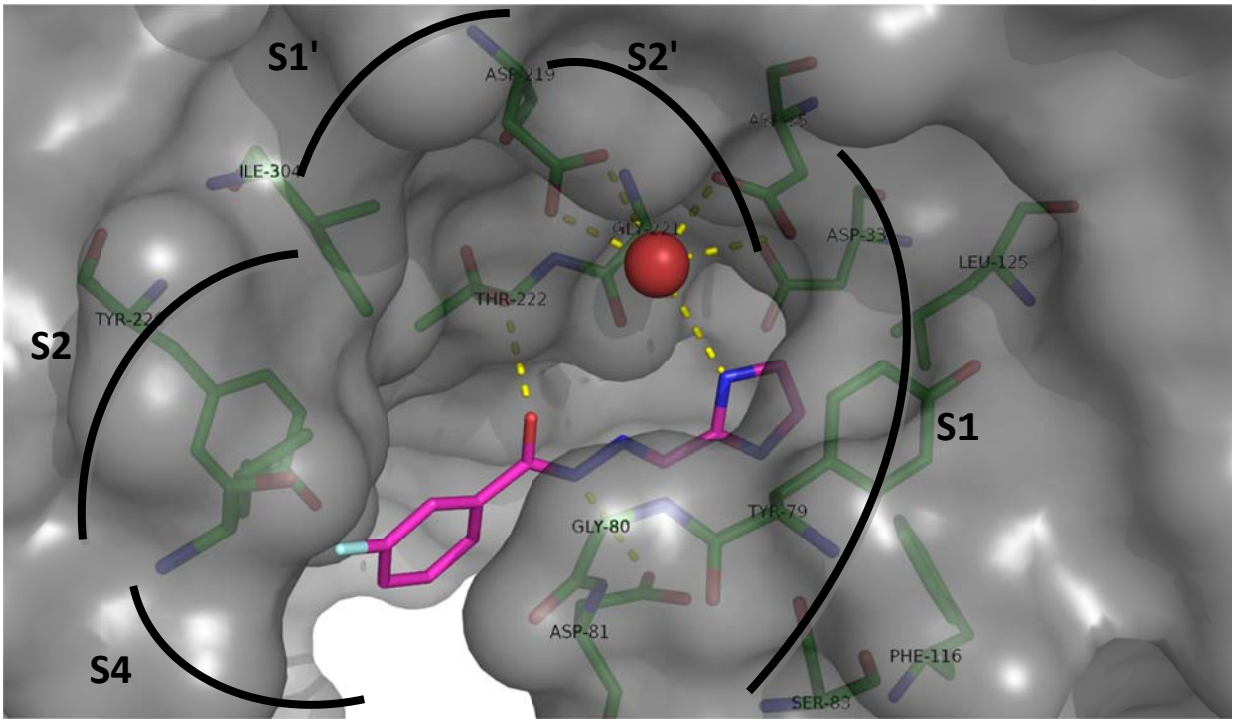
Modelling results

Colour code for the figures: O: red, N: blue, F: pale cyan, $C_{protein}$: green, C_{ligand}: purple. The solvent-accessible surface of the protein is shown in grey. Hydrogen bonds are shown as dashed, yellow lines. The water molecule bound to the catalytic dyad is shown as a red sphere.

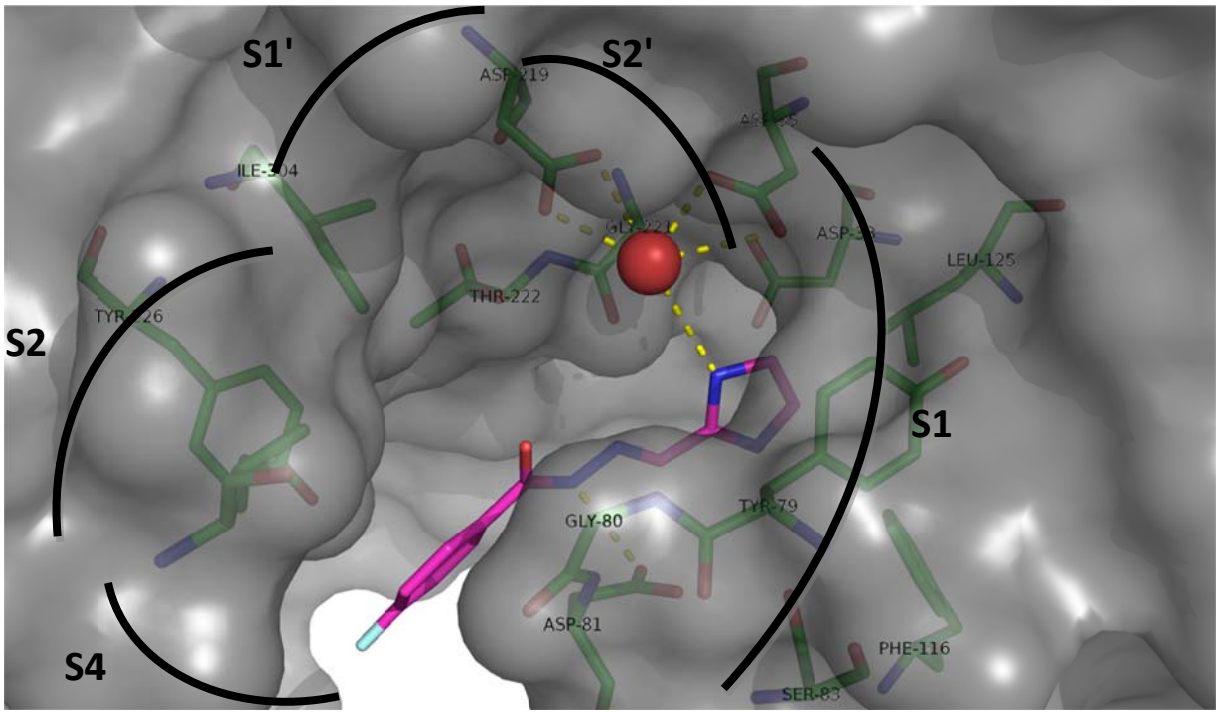
10: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-2-fluorobenzohydrazide



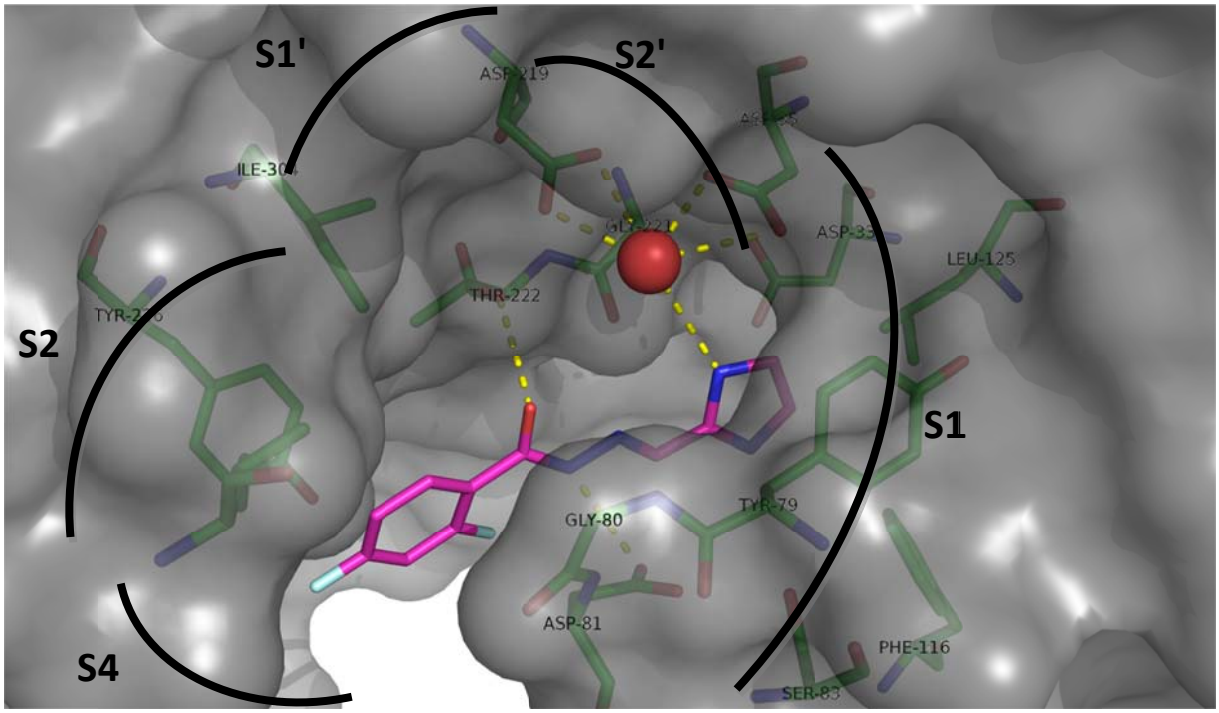
11: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-3-fluorobenzohydrazide



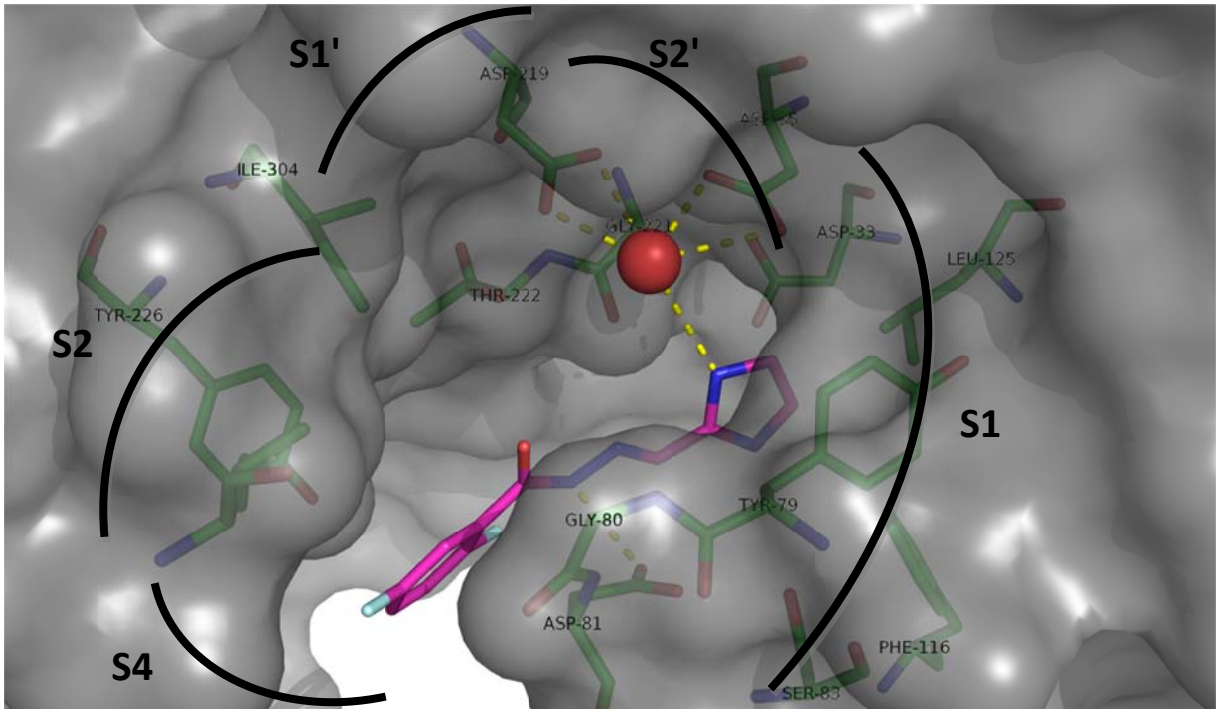
12: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-4-fluorobenzohydrazide



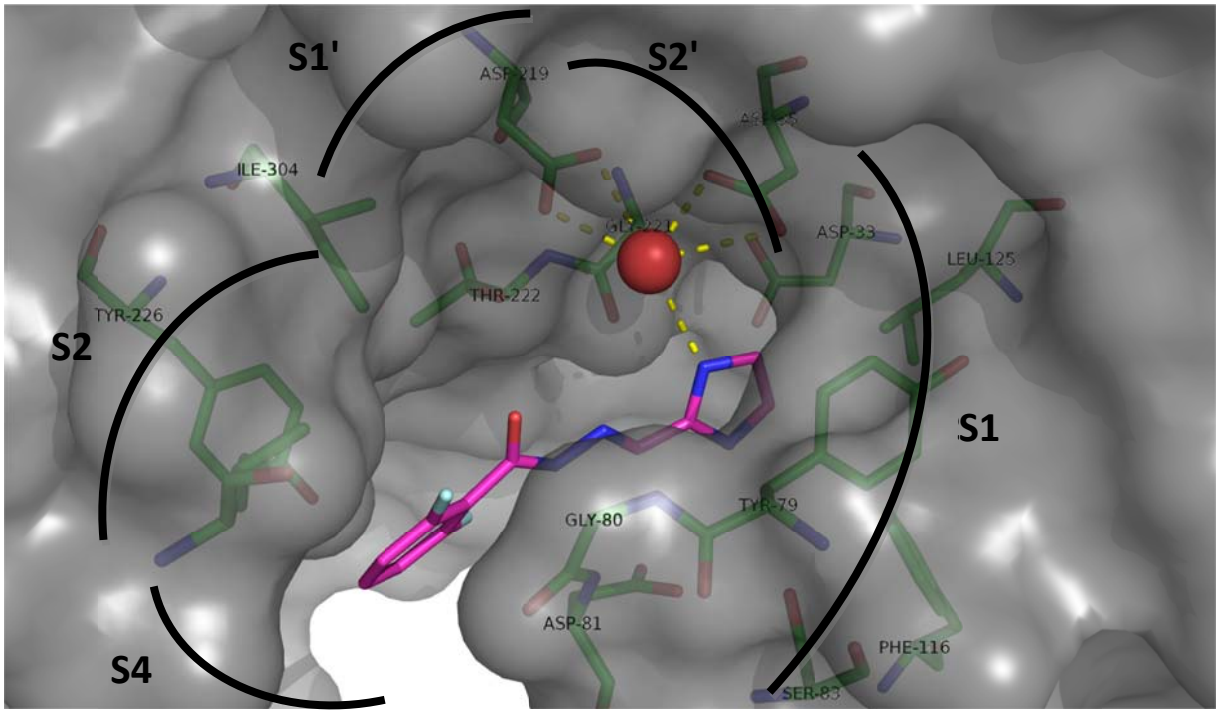
13: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-2,4-difluorobenzohydrazide



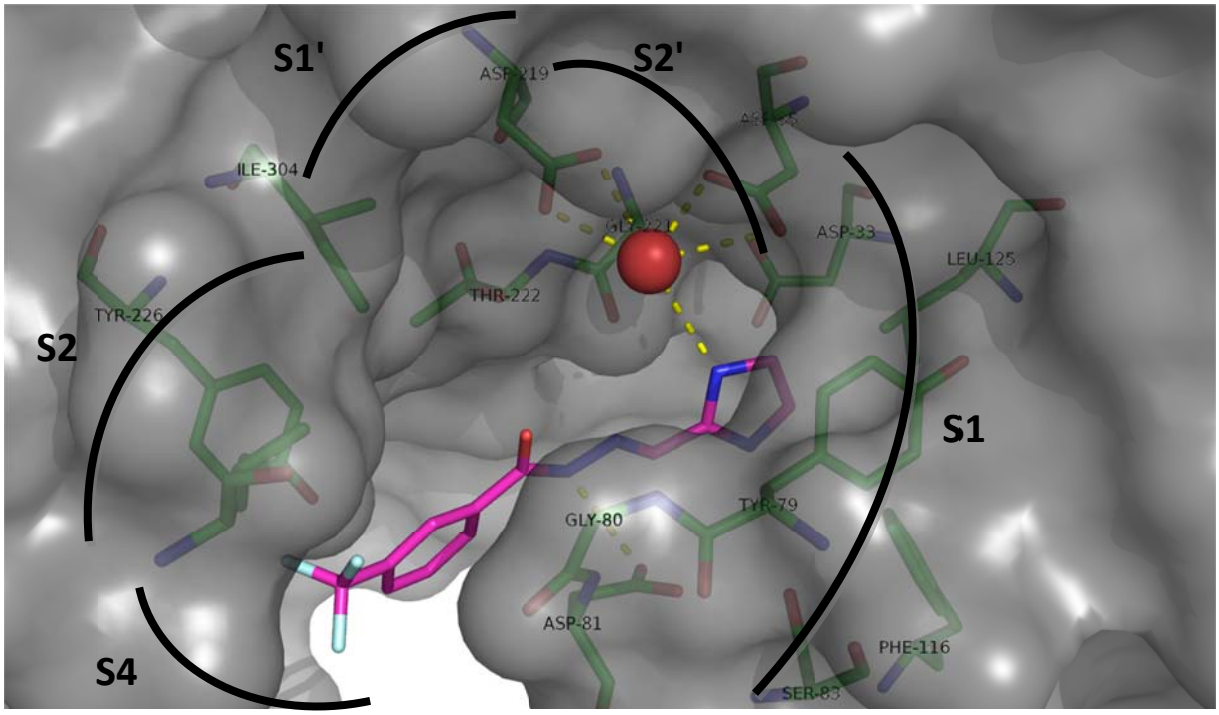
14: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-2,5-difluorobenzohydrazide



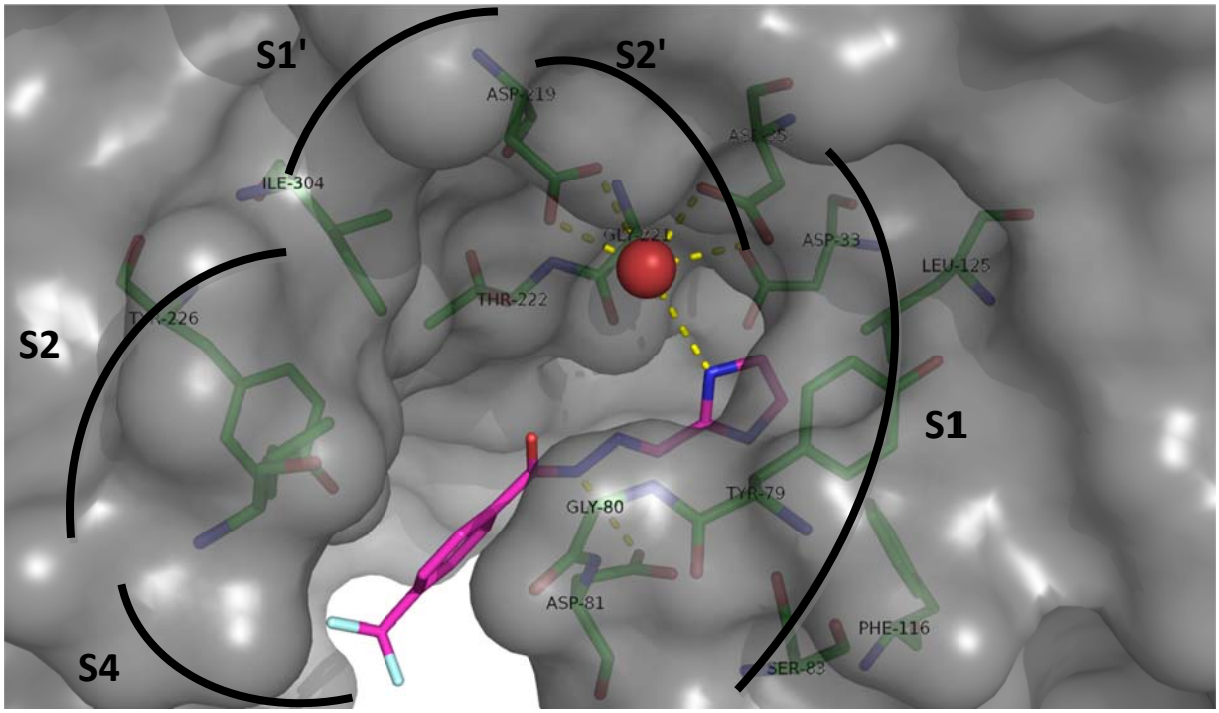
15: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-2,6-difluorobenzohydrazide



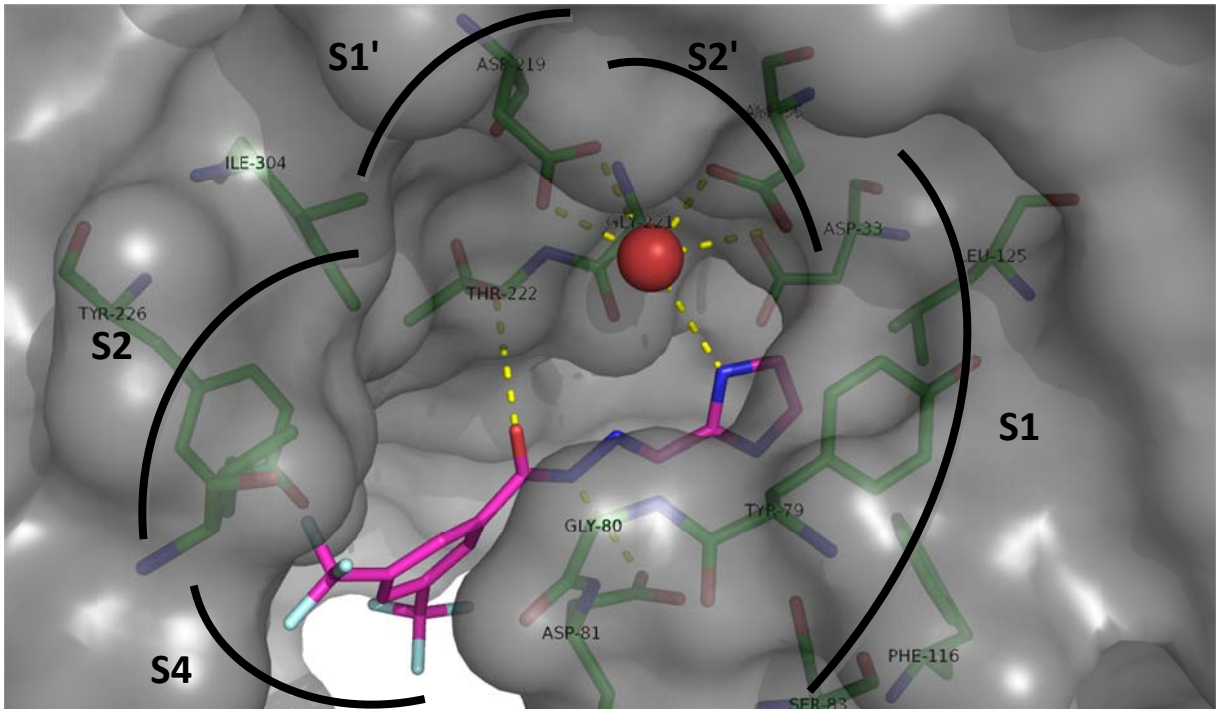
16: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-3-(trifluoromethyl)benzohydrazide



17: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-4-(trifluoromethyl)benzohydrazide



18: (*E*)-*N*'-((1*H*-Imidazol-2-yl)methylene)-3,5-bis(trifluoromethyl)benzohydrazide



NMR experiments

Preparation of NMR samples

High protein concentration

An NMR sample was prepared such that the solution contained each hydrazide (200 µM); imidazole aldehyde (400 µM); protein (100 µM); and DMSO (5% (v/v)) in deuterated sodium acetate buffer (0.1 M, pH 4.6). Trifluoroacetic acid (0.17% (v/v)) was added as a reference. The sample was equilibrated overnight by mild shaking at room temperature.

Low protein concentration

An NMR sample was prepared such that the solution contained each hydrazide (400 µM); imidazole aldehyde (800 µM); protein (4 µM); and DMSO (5% (v/v)) in deuterated sodium acetate buffer (0.1 M, pH 4.6). Trifluoroacetic acid (0.17% (v/v)) was added as a reference. The sample was equilibrated overnight by mild shaking at room temperature.

Background

An NMR sample was prepared such that the solution contained each hydrazide (400 µM); imidazole aldehyde (800 µM); and DMSO (5% (v/v)) in deuterated sodium acetate buffer (0.1 M, pH 4.6). Trifluoroacetic acid (0.17% (v/v)) was added as a reference. The sample was equilibrated overnight by mild shaking at room temperature.

Reference samples

An NMR sample was prepared such that the solution contained one hydrazide (400 µM); imidazole aldehyde (1600 µM); and DMSO (5% (v/v)) in deuterated sodium acetate buffer (0.1 M, pH 4.6). Trifluoroacetic acid (0.17% (v/v)) was added as a reference. The sample was equilibrated overnight by mild shaking at room temperature.

Quantitative ^{19}F-NMR

The spectra for quantitative ^{19}F-NMR were recorded on a 500 MHz Varian Inova NMR spectrometer equipped with an ID probe at 415 MHz; acquiring 2048 scans with a pulse width of 60° and referenced against trifluoroacetic acid at −76.5 ppm. The integral of the trifluoroacetic acid peak was set to 1000, and peaks of the library members were assigned using spectra from the reference samples.

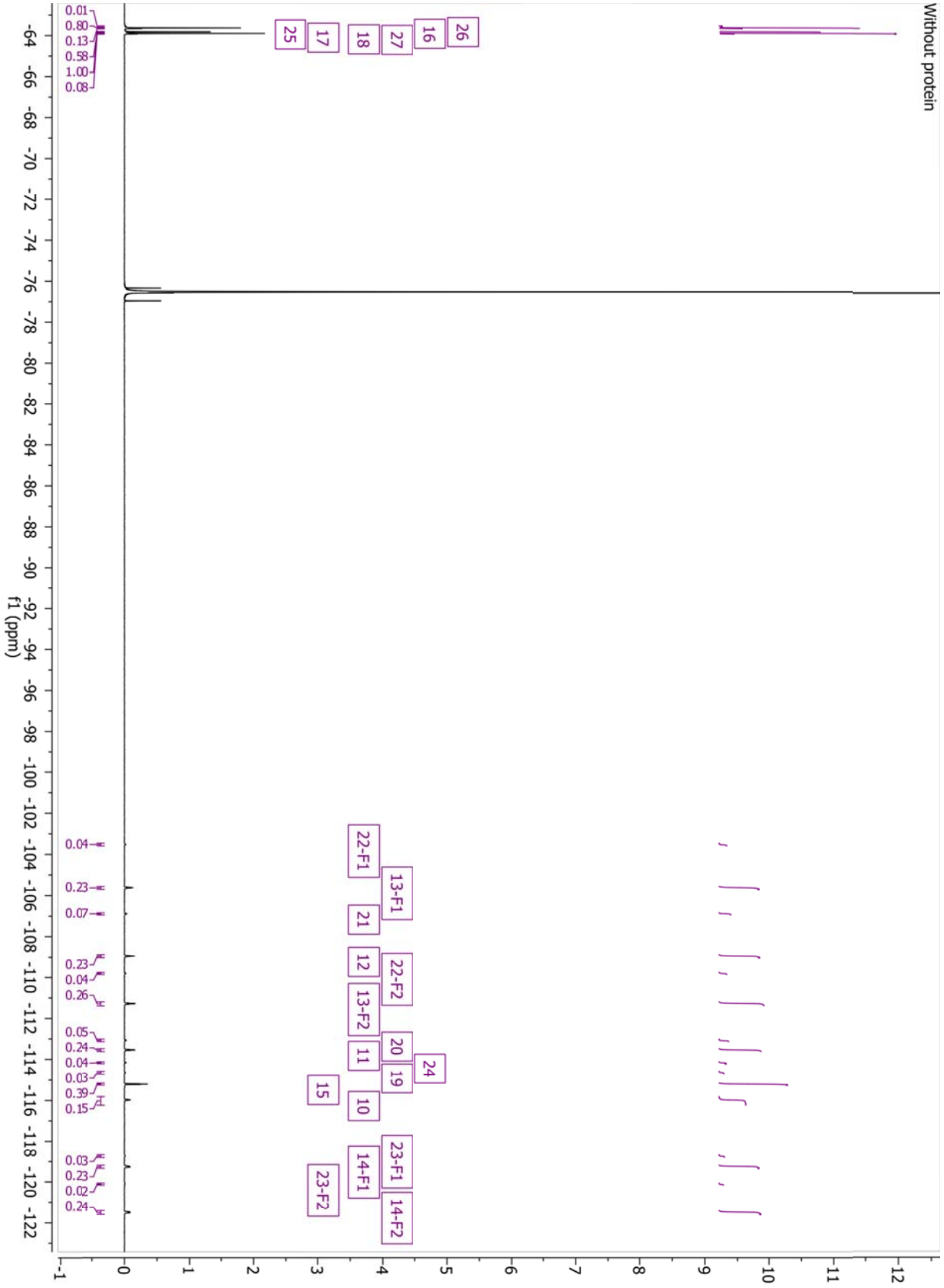
^{1}H-STD-NMR

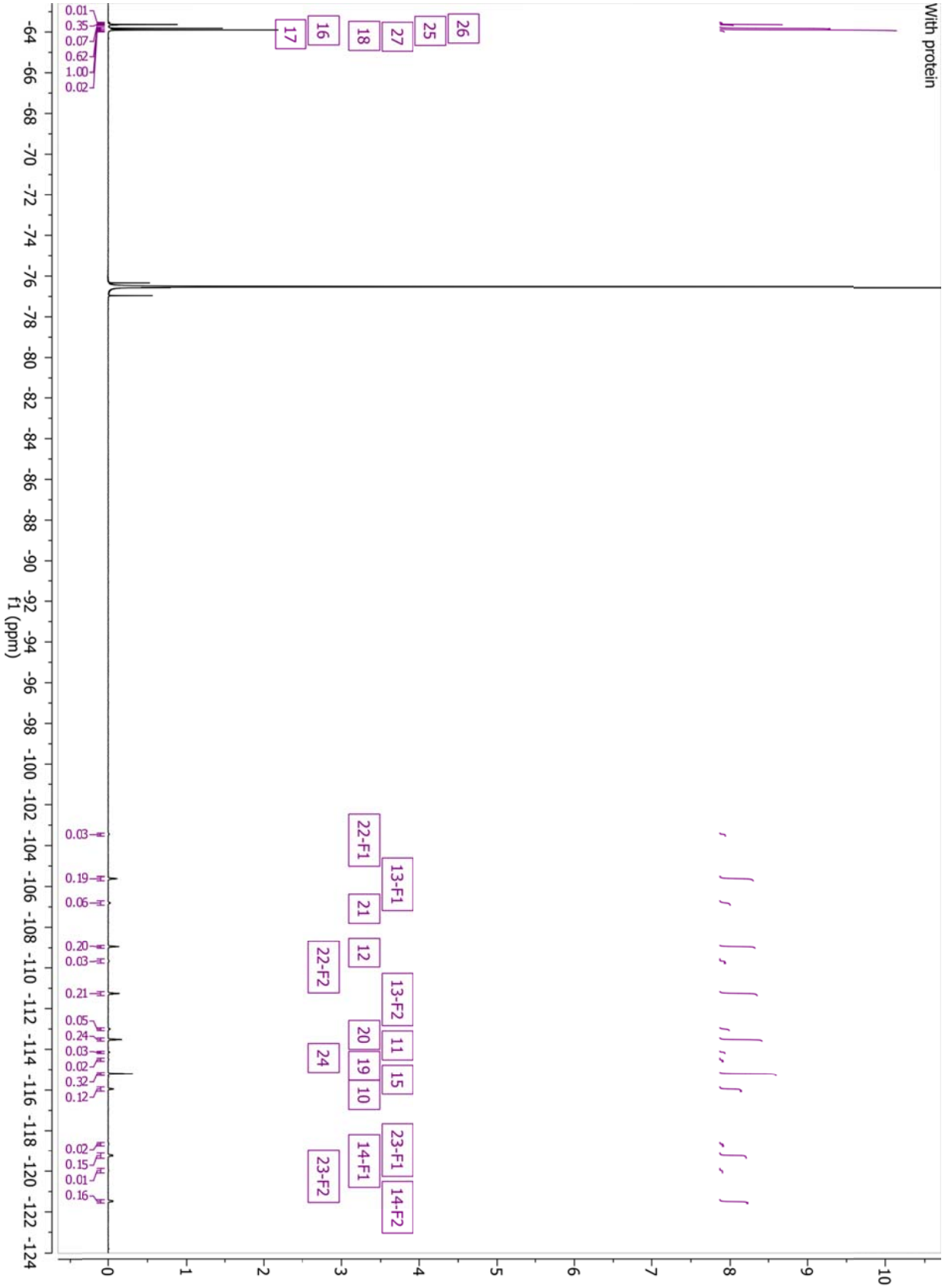
The ^{1}H-STD-NMR spectrum was recorded on a 600 MHz Varian Inova NMR spectrometer equipped with a HCN triple resonance probe; suppressing solvent peaks from water and DMSO using a WET pulse sequence.

Quantitative ^{19}F-NMR results

	Hydrazide	Acylhydrazone
10-minus	4.02	0.7
10-plus	3.37	0.35
11-minus	6.16	1.34
11-plus	6.92	1.55
12-minus	5.85	1.77
12-plus	5.88	1.65
13-F1-minus	5.75	1.02
13-F1-plus	5.54	0.89
13-F2-minus	6.55	0.98
13-F2-plus	6.24	0.72
14-F1-minus	5.75	0.62
14-F1-plus	4.37	0.38
14-F2-minus	6.04	0.68
14-F2-plus	4.58	0.42
15-minus	10.35	0.78
15-plus	9.28	0.75
16-minus	21.21	3.62
16-plus	10.3	2.38
17-minus	15.56	0.39
17-plus	17.83	0.31
18-minus	26.32	2.84
18-plus	29.08	0.39

Table S2: Peak areas of all library members compared to trifluoroacetic acid at 1000. X-minus depict compound X in the sample without protein, X-plus depicts compound X in the sample with protein.


Without protein
f1 (ppm)
26
16
27
18
25
17
22-F1
13-F1
21
22-F2
12
13-F2
20
11
24
19
15
10
23-F1
14-F1
23-F2
14-F2


With protein
f1 (ppm)
0.01
0.35
0.07
0.62
1.00
0.02
0.03
0.19
0.06
0.20
0.03
0.21
0.05
0.24
0.03
0.02
0.32
0.12
0.02
0.15
0.01
0.16
16
17
18
25
26
27
22-F1
13-F1
21
12
22-F2
13-F2
11
20
19
15
10
24
23-F1
14-F1
14-F2
23-F2

^{1}H-STD-NMR spectrum