Development of FBLD techniques for Intrinsically Disordered Proteins

Darius Vagrys

Doctor of Philosophy

University of York

Chemistry

February 2020

Abstract

Structural biology methods have determined 3D structures of many proteins describing their mechanism of action and providing insights into their biological function. In recent years, the link between structure and function has changed with the identification of intrinsically disordered proteins (IDPs) and regions (IDRs). These proteins do not have a defined structure but can specifically interact with their partners. Such interactions are being identified as important regulators in mammalian biology and there is great interest in their druggability. However, the identification of ligands that bind to IDPs or IDRs has been challenging.

Fragment-based lead discovery (FBLD) offers numerous advantages over conventional compound screening techniques. As well as sampling a potentially larger chemical space, fragments are more likely to bind to a protein, though with lower affinity. This requires biophysical methods to detect and characterise binding and this is the focus of this thesis.

Two IDPs were studied with their literature-reported compounds:

- Aggregation-prone tau K18 was screened against literature ligands by NMR spectroscopy. One ligand was identified as binding and its analogues were further examined using NMR, MST and SPR. Electrophilic fragment screening by MS and NMR yielded several covalent hits, whereas an SPR-based screen yielded no hits.
- One literature ligand was confirmed to bind to commercial cMyc peptide by NMR. Oncogenic cMyc peptides could not be produced for further studies due to degradation and purity problems.

Several fragment screens were performed against IDP and IDR:

- Hydrophilic acylated surface protein (HASPA), thought to be involved in leishmaniasis, was expressed and screened against a fragment library using NMR. The screening yielded no confirmed hits.
- Protein Tyrosine Phosphatase 1B (PTP1B)(1-301) and (1-393), containing an IDR, was produced and investigated using NMR and SPR. ^{19}F NMR fragment screening did not yield any hits binding to the IDR, whereas SPR identified several potentially IDR-directed fragments.

This study highlights the importance of orthogonal assays and direct observation of IDPs during the screening process in order to avoid potential artefacts.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my supervisors Ben Davis and Roderick Hubbard for providing me with this opportunity to work and learn more about intrinsically disordered proteins, fragment-based ligand discovery and their current challenges. The guidance they provided helped me to become a better scientist and science communicator.

The whole Vernalis RnD LtD staff has provided me with valuable discussions and insights during my project including but not limited to Hayley, Richard H., Heather, Natalia, Douglas, Loic, Nick, Emilie, James D., Alan R., Neil, Julia, Peter, Karen, Terry, Sharon and Samantha.

My deepest thanks also go to all of the collaboration supervisors, including Michael Plevin, Helena Danielson, Anna Moberg, Wolfgang Jahnke and Gyorgy Keseru. In addition to this, I would like to acknowledge multiple people I worked with during the secondments and collaborations:

- Uppsala University: Eldar, Vladimir, Edward and Giulia.
- Novartis: Lena
- University of York: Alex, Mike, Bas, Eleni
- RCNS: Peter

Last but not least I would like to thank, Iwan de Esch, Jacqueline van Muijlwijk-Koezen and all EU Marie-Curie FragNet PhD fellows for organizing and being a part of this superb international journey.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University with the exception of the PhD in Chemistry thesis by Sophie McKenna that included HASPA expression and purification work. All sources are acknowledged in Materials and Methods, and References sections. The experimental work was carried out solely by me with the exceptions listed below:

- Intact protein MS experiments and data were obtained and analyzed by Loic Le Strat at Vernalis RnD Ltd
- SPR experiments with literature-reported compounds against biotinylated tau K18 were performed by Edward Fitzgerald at Uppsala University
- X-filtered NMR experiments for tau K18 were carried out by Alex Heyam at University of York

1   Introduction

The classical view of proteins has closely related function to a stable three-dimensional (3D) structure and the paradigm of 'S*equence -> Structure -> Function'* has dominated scientific thinking since early in the 20th century [1], [2]. In 1936, Mirsky and Pauling deduced from a number of studies that the characteristic behaviour of a folded protein was related to its unique structural configuration [3]. For example, the experiments that showed a loss of enzymatic activity in the presence of alkali, acid or urea confirmed the need for proteins, or more specifically enzymes, to be properly folded in order to retain their function. By the time the first experimentally derived 3D models of protein structure were available, this '*structure->function'* relationship was widely accepted   and underpins our understanding and appreciation of most protein structures [4], [5].

In 2003, the human genome sequencing project was completed, indicating that there are over 25 000 protein-coding genes [6]. The actual number of proteins expressed is much higher due to the various splicing events that produce many variants and current estimates are that there are more than 200 000 transcripts [7]. To date, crystal structures of more than 16 000 different human proteins (> 90% homology) are deposited in the Protein Data Bank with many structures of the same protein in complex with different partners and ligands, providing a rich description of many of the processes that underpin human biology [8], [9].

It was in 1975 that the first ideas of a protein (in this case, glucagon, a 29 amino acid peptide) having a biological function without having a defined structure, emerged [10]. Over the last decade, it has also become apparent that over 30% of the eukaryotic proteome contains genes that encode for significant regions of intrinsically disordered proteins (IDPs) [11]. In addition to this, over 27% of the human genome is predicted to contain helix-forming molecular recognition features (MoRFs). These elements are thought to undergo disorder-to-order transitions upon interacting with their binding partner, providing a further example of IDPs and intrinsically disordered regions (IDRs) [12], [13]. The presence of potential IDRs have also been previously noted in a number of crystal structures where missing electron density was observed, including the flexible loops of tobacco mosaic virus protein and Bcl-X$_L$ [14], [15]. In order to understand such a potential prevalence of IDPs, it is important to understand their inherent functions in organisms.

1.1 Function of disordered proteins

Multiple IDPs have been identified to be involved in a wide range of protein-protein interactions (PPIs) that include cell signalling, transcription, apoptosis and autoregulatory functions [16]–[20]. It is important to mention that most of the IDPs and their functions so far have been noted by using mutants and knock-out systems rather than measurement of detectable activity which can used for tracking enzyme behaviour. The propensity of IDRs and IDPs to be appended with post-translational modifications (PTMs) has been noted to add an increased complexity to signalling pathways, including threshold responses, phosphorylation switches and molecular clocks [21]–[24]. Moreover, the IDRs have also been observed to link separate domains of proteins that may increase their affinity and specificity towards specific binding partners by providing additional conformational flexibility inside the cell [25], [26]. The ability of IDPs to interact with multiple targets due to internal dynamics and act as signalling hubs suggests that allosteric regulation may be important (**Figure 1**) [27]. This regulation allows the fine-tuning of the interactions and introduces additional complexity to the signalling cascades. In general, although IDPs and IDRs are prone to proteolytic degradation and aggregation, they presumably provide an important and beneficial mechanism of regulation [28], [29].
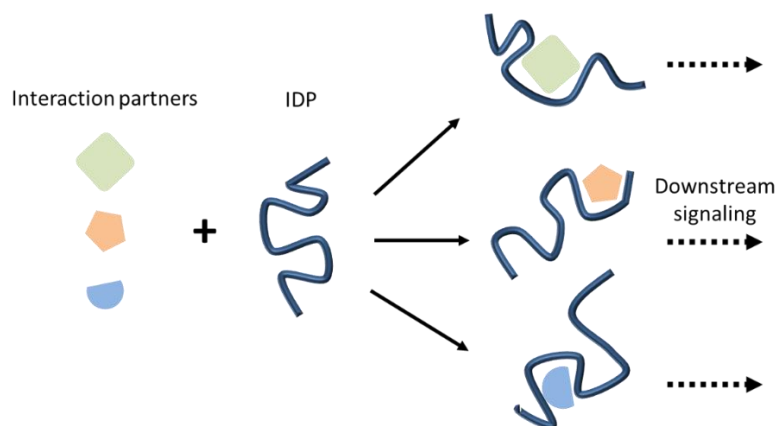
Figure 1. A representation of an IDP acting as a signalling hub.

In order to better understand IDPs and their involvement in PPIs, it is important to understand the sequence and structural differences that differentiate IDPs from their folded counterparts.

1.2 Structure of disordered proteins

The IDPs are a class of proteins that do not have a stable secondary or tertiary structure entirely or have an IDR as a part of an otherwise folded domain. The IDPs have also been shown to exhibit either collapsed (molten globule) or extended (coil or premolten globule) conformations [30], [31]. Interestingly, upon interactions with their binding partners, the structure of IDPs and IDRs can change in several ways, including disorder-to-order [32] or order-to-disorder [33] transitions.

A comparison of primary amino acid sequences between ordered and disordered proteins suggests that Pro, Glu, Ser, Gln and Lys amino acids are disorder-promoting, whereas Cys, Trp, Ile, Tyr, Phe, Leu, Met, His, Val and Asn are order-promoting amino acids [34]. The remaining amino acid residues of Ala, Gly, Thr, Arg and Asp were identified to be indifferent to protein structure. Using this data, the phenomenon of disorder can be predicted with some accuracy using primary sequence-based bioinformatics approaches [35]. A special note should be taken of Pro which is found in IDPs in significantly increased numbers in comparison to their ordered counterparts [30], [36]. Proline is known to not only affect the formation of secondary structures, but also participate in signalling events by assuming *cis* or *trans* conformations [37]. It is established that non-folded or misfolded proteins may aggregate due to hydrophobic interactions which decrease the solubility of the proteins. Such proteins may be degraded by internal proteasome systems, such as ubiquitin-proteasome [38]. It is evident, however, that the IDPs have evolved to avoid aggregation by having low hydrophobicity, large net charge and increased number of Pro residues that significantly inhibit the formation of β-sheets typically found in aggregates [39], [40].

As mentioned previously, PTMs of IDPs can provide additional complexity to the signalling pathways (**§1.1**). The apparent structure of IDPs can also be influenced by the level and type of PTMs, such as phosphorylation, acetylation and glycosylation. The PTMs can also provide additional interaction possibilities between parts of the IDP or with its target molecules.

It has also been shown that the surfaces of folded proteins, found at PPI interfaces, contain compact and specific regions of residues, or "hot-spots", that are essential for high affinity interactions [41]. Such regions have also been noted to structurally adapt and interact with multiple partners, as shown for immunoglobulin G [42] . In addition to previously mentioned

MoRFs in IDPs that undergo 'disorder-to-order' transition, supplementary interacting spots have also been identified for IDPs, including short linear motifs (SLiMs). SLiMs are short (3-10 amino acids) protein segments found in IDRs that are involved in proteolytic cleavage, PTMs and ligand binding [43], [44]. In such way, the motifs add yet another level of functionality to the proteins.

In order to better understand the IDPs and their apparent structure, it is also important to understand the folding phenomenon itself.

1.3   Folding phenomenon

Protein folding is a complex process from a mechanistic point of view. It is considered that the folding events depend primarily on the primary sequence of amino acids. It has been proposed that folding of proteins can be described through a schematic energy landscape that represents possible folding pathways of a protein (**Figure 2**) [45], [46].
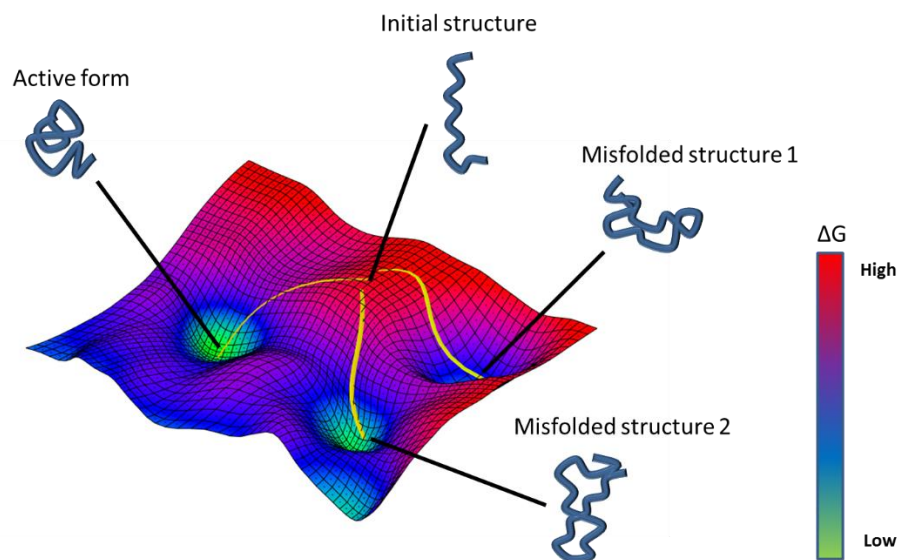
Figure 2. A schematic representation of energy landscape for protein folding. It also indicates that proteins may fold incorrectly without any internal or external guidance.

The landscape represents multiple conformations at different energy states that over time are funnelled towards protein's unique native structure that has the lowest or stable-enough energy state [47], [48]. It has been estimated that the number of possible conformational structures at any given moment during the folding process can be very large due to all possible bond angles (up to 10^8 for 150 amino acid protein) and it may be difficult to achieve the correct active conformation without any 'internal guidance' [49]. In some cases, this can even cause misfolding of the proteins. In contrast to folded proteins, IDPs may exist as a

conformational ensemble out of which several conformations can be in the lowest energy state (**Figure 3**). The IDP can assume different active conformations without much energy loss which allows it to participate in multiple PPIs, as mentioned previously (**§1.1, Figure 1, Figure 2**).
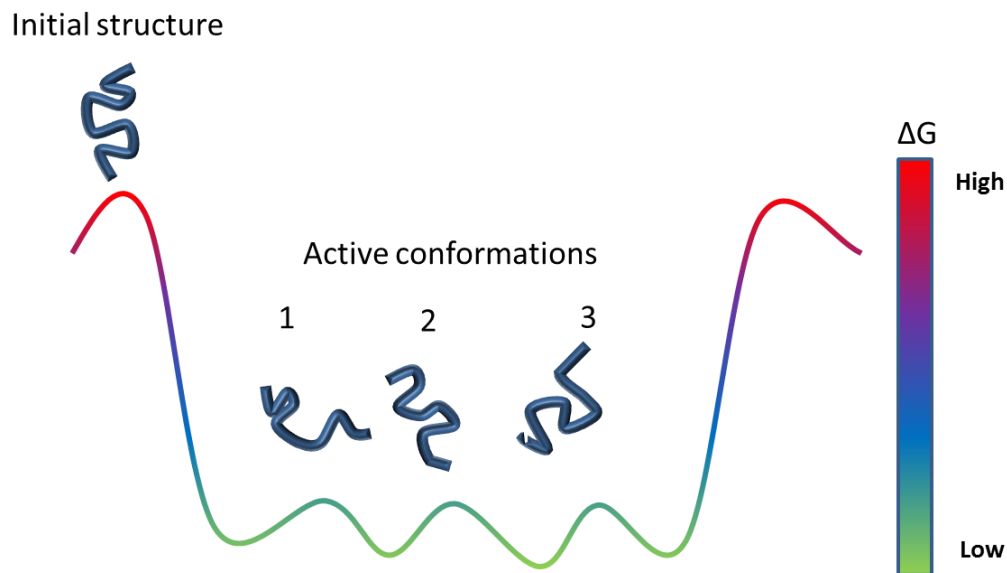
Figure 3. *A schematic representation of energy landscape for IDP folding which indicates that IDPs can exist as an ensemble of functionally active conformations.*

It has been postulated that the protein folding process may start with the formation of local elementary structures (LESs) [50]. LESs are pseudo secondary structures that form on their own in early stages of folding [51]. These structures are stabilized by hydrophobic residues, such as Trp, Phe Tyr, that act as local nucleation sites and guide the folding process. However, it is also important to mention that some proteins are unable to fold on their own and are assisted by molecular chaperones, such as Hsp70 [52], [53]. Chaperones stabilize specific, energetically unfavourable conformers of the target protein during the folding process.

The functional importance of IDPs and IDRs in therapeutic areas and their structural peculiarities led to multiple attempts to modulate the interactions of interest in a desirable manner with low molecular weight (LMW) ligands.

1.4   Changing the paradigm

As discussed previously, IDPs and IDRs have been shown to be involved in a wide range of PPIs affecting multiple therapeutic fields, including cardiovascular, neurodegenerative, oncogenic, transcriptional, and autoregulatory (**§1.1**). These examples show that proteins do not have to

have a pre-defined structure to exhibit a function, which is a change of the previously established paradigm of '*structure -> function*'.

The interfaces of PPIs are known to be highly polar which resemble the solvent-exposed surfaces of the proteins. The PPIs usually cover a larger surface area (1000-3000 $Å^2$) than the typical protein-ligand interactions (300-1000 $Å^2$) [54]–[56]. Due to the nature of PPIs, the binding surface is usually flat-like and is devoid of specific groves or fingers, while the affinities can range from mM to pM [16], [57]. The specificity of PPIs that involve IDPs is thought to be achieved by partial desolvation of the interface, followed by ion pairing, hydrogen bonding which allows rapid binding kinetics [58], [59]. Even so, the PPIs that include IDPs have been measured to form ultra-high affinity complexes with pM affinities [60].

Interestingly, despite the flexibility of IDPs, several studies have shown that the PPIs between IDPs and their binding partners can be modulated by LMW ligands.

Multiple IDR-containing protein p53 has been identified to be involved in control of apoptosis, cell-cycle arrest and DNA repair [61]. Due to specific mutations in p53 gene, p53 may become constitutively deactivated which causes tumorigenesis. In addition to this, specific changes in typical p53-HDM2 interactions also inactivate wild-type p53. The disordered N-terminal transactivation region was identified to be important in p53-HDM2 interactions. The IDR was found to undergo a "disorder-to-order" transition and form a rigid structure in complex with HMD2. The determined structure was then used to develop multiple LMW ligands, such as Nutlin-3a and MI-219, that disrupted the complex formation [62], [63].

The disruption of PPIs by LMW ligand has also been shown for EWS-FLI1:RHA complex. The EWS-FLI1 is an intrinsically disordered fusion protein that was shown to be essential for tumor maintenance in Ewing's sarcoma family of tumors. EWS-FLI1 has many interactions partners but RNA helicase A (RHA) was determined to be its most common target using phage display [64]. A single compound NSC635437 was identified from 3000 compound library using SPR assay to bind to the disordered EWS-FLI1 fusion protein. The ligand was further developed to yield YK-4-279 with K_D of 9.48 µM to the EWS-FLI1 complex [65]. YK-4-279 was shown to displace the RHA binding site-representing peptide E9R from the EWS-FLI1 complex. Due to disordered nature, however, the exact mechanism of action has not been established.

The interactions between oncoprotein cMyc and its partner Max have also been shown to be inhibited by multiple LMW ligands, including 10074-G5 and 10058-F4 [66]–[68]. The structural data of 10074-G5 and 10058-F4 interactions with cMyc have been confirmed by NMR spectroscopy [69].

These examples show that effective targeting and modulation of IDPs with LMW ligands is possible, opening new opportunities in drug discovery. However, sensitive, robust and reasonable approaches are required to identify and characterize the IDP:ligand interactions. Fragment-based ligand discovery (FBLD) may potentially be used for this purpose.

1.5   Fragment-based ligand discovery

The typical approach in order to identify a ligand towards a biologically relevant folded protein would be to screen a library of compounds using protein-dependant biochemical and biophysical techniques to detect binding of a ligand. The resulting hits would then be validated using additional orthogonal assays to confirm the mechanism of action and binding mode. Typical high throughput screening (HTS) employs the screening of tens of thousands of synthetic and natural compounds against the target to identify a hit compound. This is then optimised to improve binding and to incorporate suitable physicochemical and biological properties (**Figure 4**) [70]–[73]. HTS often fails to identify suitable hit compounds for new classes of protein target, as most libraries of compounds contain molecules which have been synthesised for existing types of proteins. Although large pharmaceutical libraries may contain millions of compounds, they represent only a small fraction of the estimated 10^{60} small molecules that can be synthesized to cover the chemical space of drug-like molecules [74].

Figure 4. A representation of a typical HTS approach.

An alternate approach to identifying hit compounds is fragment-based ligand discovery (FBLD). The initial concepts and examples of FBLD emerged more than 20 years ago and its use has been increasing [75]–[78]. The general idea of FBLD is to identify fragment-like molecules from a fragment library that bind weakly but efficiently to the target protein.

1.5.1 Design of fragment library

The fragment screening library is usually designed in such a way that there is a higher chance of fragment binding to the target due to fewer non-favorable interactions and increased sampling of chemical space. It is also important to mention that a fragment has also to have sufficient complexity to make target-specific interactions [79]. For example, for fragment-based NMR screening library, the fragment structures can be selected using a multi-step approach. First, the fragments are accumulated *in silico* according to desired properties, including fragment-like MW (100 < MW < 250 Da), functionality and virtual solubility [80]. Following this, the library is further filtered by novelty, virtual screening or pharmacophore. The 2D 3-point pharmacophore fingerprint is then used to further narrow down the fragments. Finally, the availability, tractability, experimental stability and solubility results in a curated library that can be used in multiple fragment screening campaigns.

There has also been suggestions to use a "rule of 3" (Ro3) approach for designing fragment libraries in order to limit the chemical complexity and maintain favourable physicochemical "fragment-like" properties [81], [82]. However, in some cases fragment hits have been shown to disregard this rule and were still valid for hit-to-lead development. Such observations suggest that Ro3 should not be followed blindly but rather used as a reasonable suggestion in the library design.

1.5.2 Detection of fragment hits to proteins using biophysical techniques

In order to enable FBLD, sensitive and state-of-the-art biophysical methods have to be used. The typical biophysical techniques used to identify fragment hits against a folded protein are X-ray crystallography, NMR, SPR and MST [83]. Out of these techniques, only X-ray and NMR can provide atomic scale structural information about protein:ligand interactions.

1.5.2.1 X-ray crystallography

By crystallizing the protein together with fragment cocktail (co-crystallizing) or by soaking an already preformed crystal in fragment cocktail, it is possible to produce a protein:ligand complex. The formed complex can then be evaluated in atomic scale by studying the data of X-ray crystal diffraction pattern [78]. The resolved structure can provide a potential mechanism of action and further guide "hit-to-lead" studies. In addition to this, very weak binders can also be detected due to crystallization conditions allowing high concentrations of fragment solutions. However, some potential drawbacks also may be present, including

crystal stability, packing patterns, multiple attempts of crystallization and synchrotron time. It has been suggested that crystallization approach may be an expensive primary screening tool to get a straightforward "yes or no" answer [84].

1.5.2.2 *Nuclear Magnetic Resonance (NMR)*

NMR-based methodologies have increased throughout the years and today offer a variety of ligand- and protein-based techniques to be used for FBLD projects. In general, due to relative insensitivity of the technology when compared to other orthogonal assays, high protein concentrations are required for these type of experiments: around 5-10 µM of protein for ligand-observed and 50-1000 µM of protein for protein-observed NMR approaches (~1 mg of 20 kDa protein at 100 µM) in low salt buffer solution.

Ligand-observed NMR

In early FBLD stages, ligand-based NMR can help to identify any potential binders to the target protein by using such techniques as saturation transfer difference (STD), water-ligand observed via gradient spectroscopy (wLOGSY), Carr–Purcell–Meiboom–Gill (CPMG)-filtered and ^{19}F CPMG-filtered.

STD technique is based on selectively saturating protein resonance which would cause the magnetization to spread through the protein due to spin diffusion and progressively saturate any bound ligands [85]. The dissociation constants between 10^{-8} to 10^{-3} M can be detected by this method.

wLOGSY technique works by transferring the magnetization from water molecules to the ligand [86]. If the ligand is interacting with water via protein-ligand or protein-ligand-water complex, the observed NOEs are of an opposite sign than the NOEs of non-bound ligand in solution that only interacts with bulk water.

CPMG-filtered method measures the changes in transverse relaxation rate for the ligand [87]. The T_2 relaxation rate increases for the bound ligand due to decreased rotational diffusion rate and increased rotational correlation time for the ligand. This is observed as line broadening for ligand-specific resonances in the spectra after CPMG filter. Similarly, ^{19}F containing compounds can be detected using the same working principle. Due to ^{19}F sensitivity to its chemical environment and broad spectral width, cocktails with more than 20 compounds can be effectively screened in one experiment [88].

Protein-observed NMR

Once hit fragments have been identified, protein-observed NMR can be used as an orthogonal approach to validate the hits by performing 2D experiments in order to identify a potential binding pocket in addition to measuring the affinity. However, these approaches have several limitations. Most importantly, ^{15}N and/or ^{13}C isotopic labelling of the protein is required which may become expensive. Also, the target protein has to be soluble and not precipitate or aggregate at required concentrations. The protein-size is also limited to ~35 kDa due to increased relaxation rates and lack of spectral resolution. This can be overcome with more robust labelling approaches, such as ^{2}H labelling, side-chain labelling and isotopic unlabelling [89], [90]. This can be especially useful for IDP systems where spectral overlap can be a problem even for small proteins (~10 kDa).

2D NMR spectra that are used for determining potential binding are usually obtained with heteronuclear single quantum coherence (HSQC) and heteronuclear multiple quantum coherence (HMQC) experiments. The resulting spectra represent backbone amide cross peaks via ^{1}H-^{15}N correlation or carbon cross peaks via ^{1}H-^{13}C correlation. Additional experiments, such as HNCO, HNCACB and HN(CO)CACB, can be performed in order to assign the HSQC cross peaks to specific amino acids in the sequence. If needed, NMR-guided modelling (NGM) may be used if X-ray crystallography approach is not available [91]. NGM experiments identify which parts of a ligand interacts with which part of the protein. This data can then be used as conformational constraints for computational modelling systems, such as CYANA, to find the optimal conformation of the ligand in contact with the protein which may be a relevant approach for IDP:fragment interactions [92].

1.5.2.3   Surface Plasmon Resonance (SPR)

SPR-based technology has become an important part of a variety of biosensors in the field of drug discovery, biology and medicine due to its non-invasive and real-time ability to observe ligand:analyte interactions. Since the first SPR-based instrument was introduced in 1990, over 20000 papers that include commercial SPR biosensors were published. However, before the first biosensors were released, it was essential to understand the physical basis of SPR in order to use it effectively and beneficially.

When a polarized light is shone through a prism onto the metal-coated surface, the light is reflected to a detector and its intensity is measured. At a certain angle, there is a decrease in

intensity of the reflected light which is caused by interactions between the polarized light and the free electrons of the metal layer of the surface. The angle which this decrease is observed at is called resonance angle. Most importantly, such resonance angle also depends on the optical characteristics of the surface which can be affected by the change of accumulated mass in the immediate vicinity of the surface. Therefore, the dynamic change of the resonance angle over time can directly provide information about local binding events happening at the surface in real time.

In SPR-based biosensor instruments, the surface of a sensor chip is usually made out of glass which is coated with a thin layer of gold followed by a hydrophilic dextran matrix. By using a well-defined chemistry, a variety of ligands, or proteins, can be immobilized on the matrix, including LMW compounds, enzymes, antibodies, other proteins and DNA/RNA. After immobilization of the ligand, the microfluidic system is used to inject the analyte onto the modified surface and any binding interactions can be observed. The occurring changes in resonance angle are represented in response units (RU) which allows the determination of interaction parameters via established mathematical models. The parameters of interest include kinetic parameters (k_a, k_d), affinity (K_D), stoichiometry and thermodynamic parameters (ΔG, ΔS, ΔH). SPR-based assays can also be used as a main or orthogonal approach for drug discovery projects, immunogenicity, binding site mapping, or proteomics in order to better understand binding kinetics and ligand:analyte interactions. Due to sensitivity of the technology, the binding of compounds from pM to low mM affinity can be detected.

A typical SPR sensorgram consists of a baseline which is established by injection of running buffer (**Figure 5**). This is followed by the sample injection during which an increase of signal is observed due to association of the analyte at specific rate constant (k_a) with the ligand until a point of saturation. After this, running buffer is injected again which causes the analyte to dissociate from the ligand at a specific rate constant (k_d). A regeneration solution can be used, if necessary, to remove any remaining bound sample from the surface.
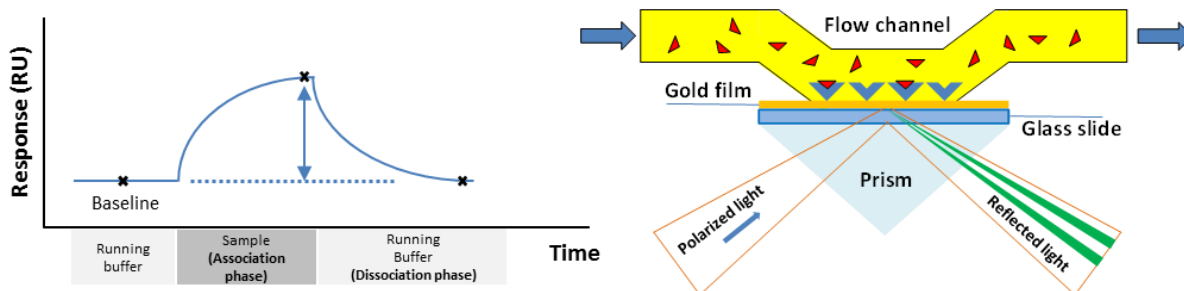
Figure 5. *A representation of a typical SPR sensorgram. (**Left**) A graphical representation of the SPR biosensor working principle and components (**Right**).*

However, the main disadvantage of the SPR assay is that the protein has to be immobilized on the surface which may conformationally restrict the ligand and prevent ligand:analyte interactions. This potential problem may be overcome by employing different immobilization strategies, including thiol coupling, amide coupling, Ni-NTA coupling via His tag and streptavidin:biotin conjugates [93]. Therefore, a good understanding of the protein system is required in order to accurately set up the assay and interpret the results.

SPR technology is used in FBLD to quickly identify potent fragments and evolve them into more potent ligands using medicinal chemistry and orthogonal biophysical assays [94]. In addition to this, off-rate screening approach allows to quickly evaluate Structure-Activity Relationship (SAR) of the synthesized compounds by evaluating any changes in ligand dissociation rates from the receptor [95], [96].

1.5.2.4 Microscale Thermophoresis (MST)

Microscale thermophoresis (MST) emerged as a new technology to investigate and characterize protein-ligand, protein-protein and protein-DNA interactions [97]. MST measures the directed movement of fluorescent molecules in microscopic temperature gradients. Due to low intrinsic fluorescence of proteins, various dyes are used to label the protein of interest. Each labelled protein has distinct thermophoretic properties that are affected by protein's size, charge and hydration shell. The binding of a ligand may change one of these parameters which directly affects the observed thermophoresis data. This change can be used to derive K_D values (**Figure 6**).
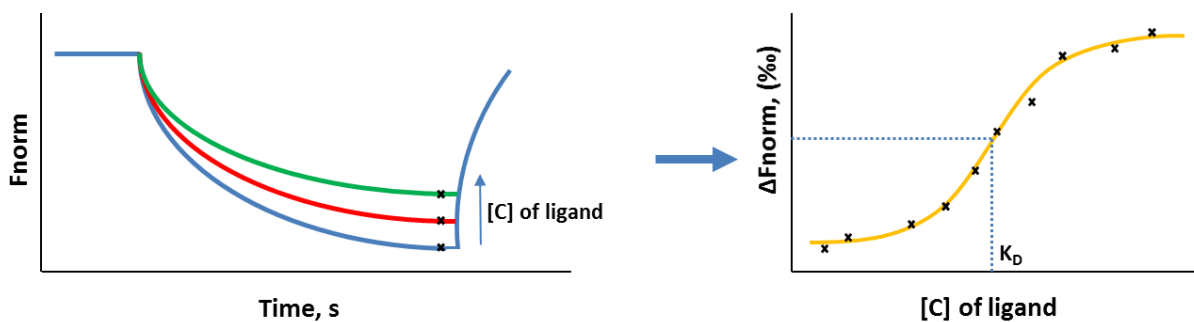
Figure 6. A representation of typical MST binding trace pattern and data analysis used to obtain K_D value.

MST has also been shown to be applicable in FBLD campaigns to identify potent binders [98]. In addition to this, MST can also identify potential artefacts from MST traces, including protein aggregation, denaturation or unfolding due to fragments which offers important early QC information.

However, it is important to be aware of potential problems when using labelling approach for MST. This includes the possibility of the dye blocking the binding site or the dye interacting with the fragment rather than the protein. Therefore, orthogonal assays are important in order to avoid false-positive hits.

1.5.3 Fragment evolution

After identifying and validating the fragment hits, the next step in FBLD involves the development and optimization of the identified compounds by Structure-Activity Relationship (SAR) studies. The ligand development can be achieved by either linking or 'growing' the fragments (**Figure 7, Figure 8**). If two or more fragments are determined to bind close in space or to the same binding site, it is possible to covalently link them via chemical linker followed by optimization using medicinal chemistry. Alternatively, it is possible to chemically expand the identified fragment further by adding additional chemical groups to it, which would potentially increase the affinity and selectivity towards the target. In both cases, this would allow obtaining a novel molecule exhibiting both higher affinity and selectivity towards the target protein.
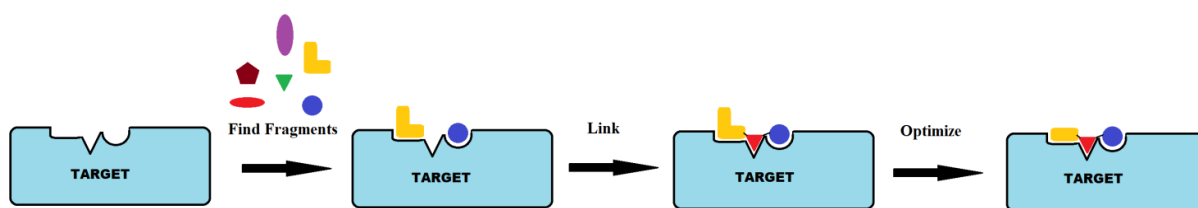
Figure 7. A representation of a typical FBLD approach by linking the fragments.
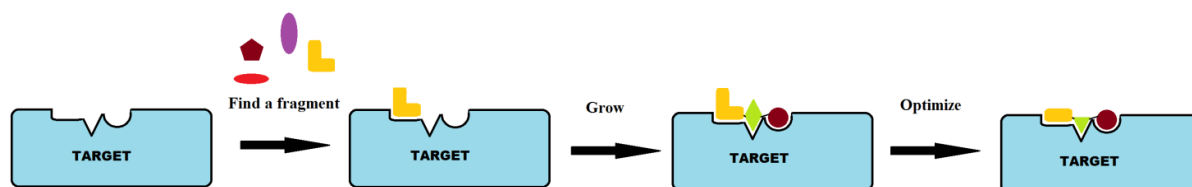
Figure 8. A representation of a typical FBLD by 'growing' the fragments.

One of a few values used to follow the effectiveness of the evolved fragment is ligand efficiency (LE) which measures how much of energy is gained per non-hydrogen atom in the structure [99]. It is common that as the ligands are optimized, the LE value decreases due to possibly not the most efficient but potentially necessary additions. While enzyme inhibitors exhibit LE usually above 0.4, the developed ligands for inhibiting PPIs have LE values of 0.1-0.35 due to larger interaction interfaces [16], [100]. These examples show that it can become a progressively difficult task to evolve the molecule into a higher-affinity one while still maintaining drug-like properties, including oral availability and permeability, as defined by Lipinski's "rule-of-5" (Ro5) [101]. Nevertheless, some marketed drugs do not follow these rules, such as macrocyclic peptides [102], [103]. This may be relevant for IDPs since the initial hits may have low LE values from the very beginning due to lack of protein structure and defined structural elements, including pockets, grooves or fingers, in large and diffuse interfaces. Moreover, a potential problem with establishing a reliable screening assay for IDPs or PPIs in general is that the proteins may behave differently *in vitro* than they do *in vivo* due to the presence of certain complexes, cofactors and specific conformations.

To date, there are three approved fragment-based drugs (erdafitinib, venetoclax and vemurafenib) in the market with over 30 more which are currently in various stages of preclinical or clinical trials (**Figure 9**).
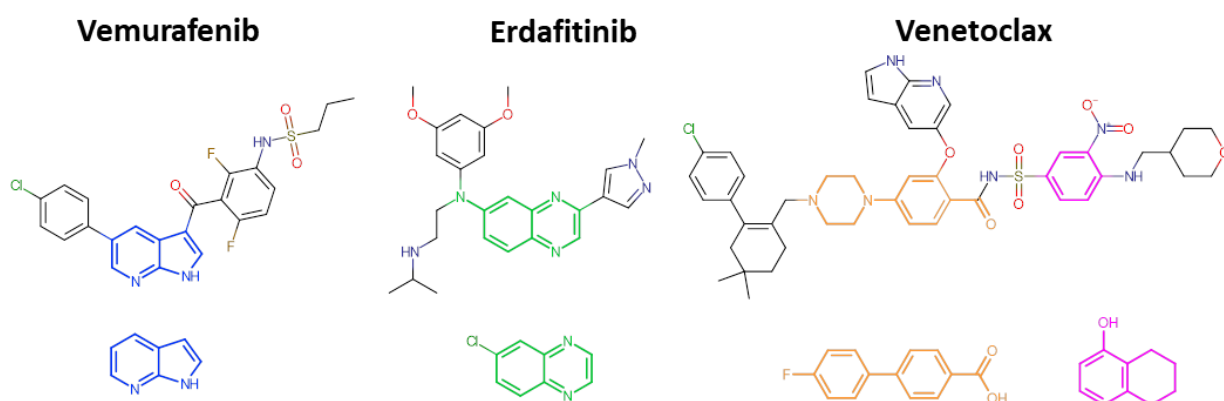
Figure 9. Currently approved drugs in the drug market that initially started as fragments.

1.5.4 Identification of ligands and characterization of IDP:ligand interactions

The same set biophysical techniques may be applicable to IDPs in FBLD campaigns, despite them not having a rigid 3D structure, except for X-ray crystallography (**§1.5.2**). In addition to this, several other biophysical techniques may be used to better characterize IDP:ligand interactions. These techniques include circular dichroism (CD) [104], fluorescence resonance energy transfer (FRET) [105], Fourier transform infrared (FTIR) spectroscopy [106], dynamic light scattering (DLS) [105] and small-angle X-ray scattering (SAXS) [107] which can help to determine frequency of secondary structures, end-to-end distances, protein folding, hydrodynamic radius R_H and radius of gyration R_g, respectively (**Figure 10**).
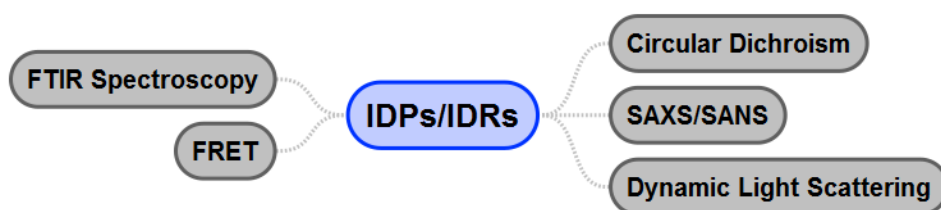
Figure 10. Additional techniques used for obtaining structural data on IDPs and/or IDRs.

1.6 Aims of the project

As has been discussed, there is an increasing interest to explore therapeutically relevant IDP systems in the drug discovery field. It would be of great interest to evaluate current biophysical methods using IDPs and their literature-reported ligands. The results would help

identifying any potential limitations or artefacts of the techniques which could then be extended out to FBLD approaches against IDPs.

In order to expand the current knowledge about IDPs and their applicability in FBLD, the following tasks had to be accomplished:

- Identify suitable IDP systems with tractable expression, purification, stability and behaviour in aqueous solution
- Evaluate literature-reported studies of ligands binding to IDPs using various biophysical techniques
- Evaluate, develop and perform fragment screening to identify fragments which bind to the IDPs
- Validate and evolve initial fragments to show enhanced potency and characterize the response of IDPs to these ligands

2 Hydrophilic acylated surface protein A (HASPA)

2.1 Introduction

Despite the advancing medical understanding of diseases and the evolution of drug discovery, there are still numerous groups of illnesses that are currently neglected but still affect millions of people around the world. One of such disease groups is called neglected tropical diseases (NTDs). These pathologies occur mainly in low-income tropical areas such as African, Asian and South American countries and affect more than a billion people in over 149 countries every year which costs billions of dollars, as reported by World Health Organization (WHO) [108]. One of the reasons why NTDs are overlooked is that they are not as fatal as other commonly known illnesses, such as tuberculosis. Moreover, the NTDs do not have easily distinguishable phenotypic symptoms. These effects make the current research of novel treatments for NTDs underfunded [109]. As of today, seventeen diseases are recognized to be in the group of NTDs, including rabies, sleeping sickness, trachoma and leishmaniasis.

Leishmaniasis is caused by intracellular protozoan parasites from genus *Leishmania* that are usually found in sandflies of genera *Phlebotomus* in the Old World and *Lutzomyia* in the New World [110], [111]. The sandfly acts as a vector for the parasite that consequently infects other organisms when the sandfly is feeding on blood. The different species of genus *Leishmania* can cause different forms of the disease in humans, including visceral, cutaneous and mucocutaneous leishmaniasis [110]. Currently, there are numerous drugs available for treatment purposes, such as pentavalent antimonials, amphotericin B, pentamidine, miltefosine, paromomycin and sitamaquine (**Figure 11**). Unfortunately, the efficacy and access to these drugs varies depending on the sub-type of the parasite and type of disease. For example, pentavalent antimonies are effective drugs against only certain subtypes of *Leishmania* parasites, whereas the exhibited toxicity and side effects of Amphotericin B prevents its wide usage outside hospitals despite its proven effectiveness [112]. However, the recent advances in treating leishmaniasis have shown that synthetic delivery systems can make the compounds less toxic or expensive. This comes as a contrast to the current formulations, which can exhibit serious side effects [113].

It has been known for some time that during its life cycle in the vector, the parasite can exist as either intracellular amastigotes or extracellular promastigotes [114]. The promastigotes

undergo certain biochemical and molecular changes inside their vector organism which is thought to increase the virulence of the parasite. Some drugs like pentavalent antimonials were determined to be effective only against amastigotes while showing a decreased effectiveness against promastigotes. This is thought to depend on internal differences in the structure and gene expression of the parasite.

In the past decade, the potential virulence factors have been associated with the increased presence of *Leishmania* species-specific molecules, including hydrophilic acylated surface proteins (HASPs) and small hydrophilic endoplasmic reticulum-associated protein (SHERP), located in cDNA16 locus on chromosome 23 [115]–[117]. HASPs have been shown to exhibit intrinsic disorder and undergo amino acid sequence–specific S-palmitoylation and N-myristoylation which allows the proteins to be localized to the cellular plasma membrane and increase the virulence of the parasite [118], [119]. It has been also shown that deletion of one of the HASP-coding genes from the gene locus, namely *HASPA*, reduces the virulence of the *Leishmania* parasite [120].
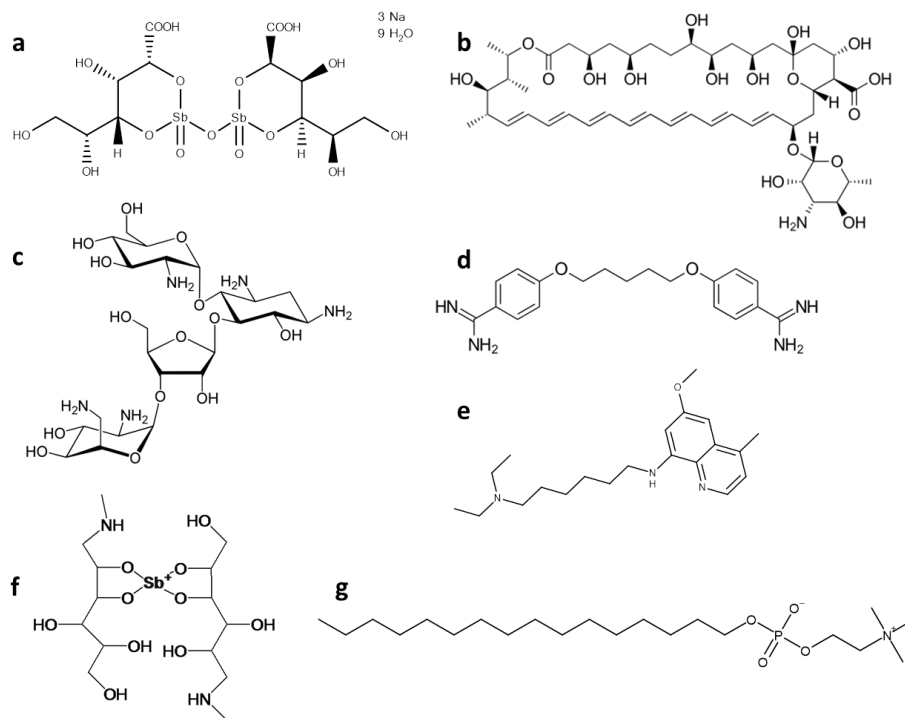
Figure 11. *The chemical structures of current drugs on market against leishmaniasis.* ***A.*** *Sodium stibogluconate.* ***B.*** *Amphotericin.* ***C.*** *Paromomycin.* ***D.*** *Pentamidine.* ***E.*** *Sitamaquine.* ***F.*** *Meglumine antimoniate.* ***G.*** *Miltefosine*

HASPA is an intrinsically disordered protein (IDP) with a structural rather than an enzymatic function, which makes the identification approaches for LMW modulators challenging. To date, no successful structural biology studies of HASPA have been reported.

The identification of LMW ligands through biophysical fragment screening may identify initial molecules that could be developed further to increase their affinity towards HASPA and block therapeutically relevant PPIs. It has been reported that HASPB is released from metacyclic parasite's surface and binds to macrophages in host organism [121]. A similar mechanism is also considered to involve HASPA. By successfully targeting HASPA with LMW ligands, it would be of interest to see if the modulation of HASPA affects the parasite's virulence. Moreover, HASPA would serve as a model system for exploring other IDP:ligand interactions. There are numerous significant advantages that nuclear magnetic resonance (NMR) technique provides during fragment screening campaigns, including the ability to identify low affinity binders, directly observe protein/ligand behaviour in solution and assess pan-assay interference (PAINS) compounds. While protein-observed NMR can provide information about the binding site of the putative hit, ligand-observed CPMG-filtered experiments could potentially provide additional information about behaviour of the ligand in solution, such as increased relaxation rates due to binding events, serving as an additional control.

In this study, isotopically labelled HASPA was produced which was followed by protein- and ligand-observed NMR fragment screening campaign. After initial hits were identified, additional NMR techniques were used for confirmation purposes. The obtained results yielded further insights about potential artefacts during screening campaigns against IDPs and provided a better understanding of the methodology.

2.2   Protein production

The isotopically labelled ^{15}N-HASPA protein was required to be produced in order to enable 1D and 2D NMR fragment screening campaign. ^{15}N-labelled HASPA with C-terminal 6xHis tag was expressed and purified, as determined via SDS-PAGE and 2D NMR (**Figure 12, Figure 13**) (**§7.3.2, §7.3.3.1**). HASPA was noted to run as a higher MW species which was also observed previously [122] (**expected:** 9.7 kDa, **observed**: ~18 kDa). The protein was previously confirmed to exist as monomeric species by SEC-MALS [119], [123]. Therefore, the higher MW observation could not be attributed to any formation of dimers. It has been determined that HASPA contains a significant number of negatively charged amino acids (Glu, Asp: 23 % of total sequence) in comparison to positively charged amino acids (Lys, Arg: 7 % of total sequence) (**Figure 14**). The larger number of acidic than basic amino acids makes HASPA protein significantly acidic (theoretical pI = 4.91). The increased negative charge of the protein

may prevent efficient association of HASPA with sodium dodecyl sulphate (SDS) and have an effect on protein's migration pattern on the gel. A similar behaviour for the highly negatively charged protein hSOD1 has also been observed [124].
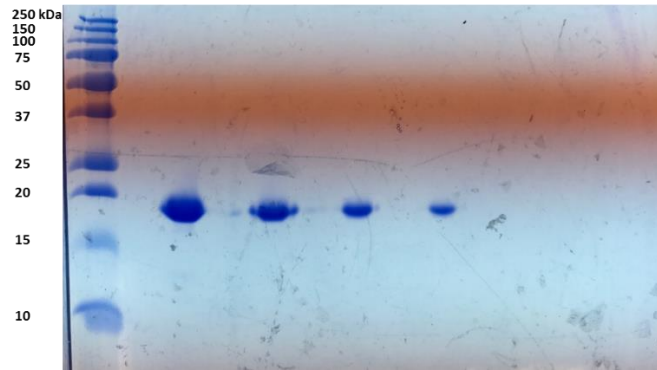
Figure 12. The QC of ^{15}N-HASPA with SDS-PAGE on 15% polyacrylamide gel.
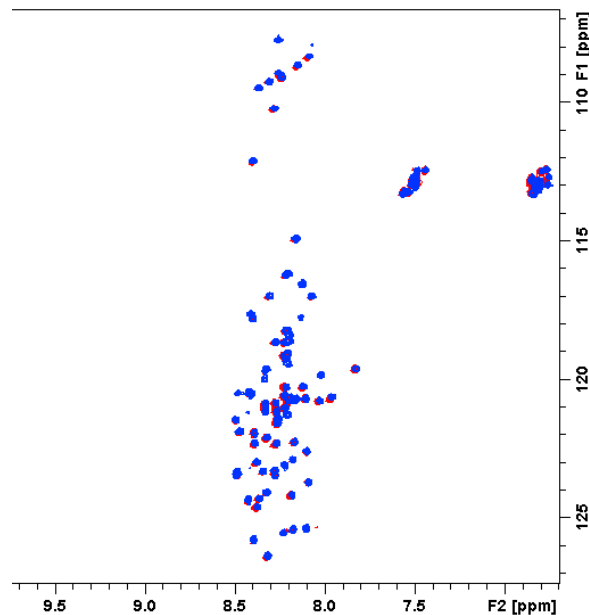
*Figure 13. A comparison of 2D SOFAST-HMQC (SF-HMQC) spectra of ^{15}N-HASPA (**Blue**) with previously purified ^{15}N-HASPA (**Red**).*

*Figure 14. The amino acid sequence of HASPA, indicating His tag (**Green**), Lys, Arg (**Cyan**) and Glu, Asp (**Red**).*

One of the approaches for determining protein concentration is to use A_{280} and the extinction coefficient of the protein. However, the value of extinction coefficient is influenced by the presence of aromatic amino acids in the sequence, namely Phe, Tyr and Trp. In case of HASPA, only two aromatic amino acids are present in the sequence (44Phe and 3Tyr) which resulted

38

in low extinction coefficient (ε = 1490 1/M*cm) (**Figure 14**). A low extinction coefficient can cause larger than 10% variability in A_{280} data, as reported previously [125]. It was suggested to use *ortho*-phthalaldehyde (OPA) assay for a more accurate protein concentration determination [119], [123] (**§7.3.8**). Under basic conditions OPA reacts with primary amines of the peptide/protein in the presence of 2-mercaptoethanol and forms fluorescent species (λ_{ex} = 340 nm, λ_{em} = 455 nm). Bovine serum albumin (BSA) was used as a standard which has been shown to exhibit linear concentration dependence from 31.25 to 500 µg/mL [126]. However, fewer Lys in HASPA sequence (~7% of total sequence) in comparison to BSA (~10% of total sequence) could have influenced the final readout values. It is suggested that a similar protein to the target (size, composition) should be used as a standard for calibration curves for more accurate results. Therefore, the actual concentration of HASPA may have had some variation. Alternatively, bicinchoninic acid (BCA) assay could have also been used because its colour development while reacting with the protein also depends on the backbone of the protein rather than just specific amino acids in the sequence. This working principle minimizes the variability between samples. In contrast, Bradford assay has been shown to produce larger protein-to-protein variability due to the method based on Coomassie dye which does not interact with the backbone of proteins.

The final concentration of the purified ^{15}N-HASPA sample was determined to be 6.85 mg/mL by OPA approach (**Figure 15**).
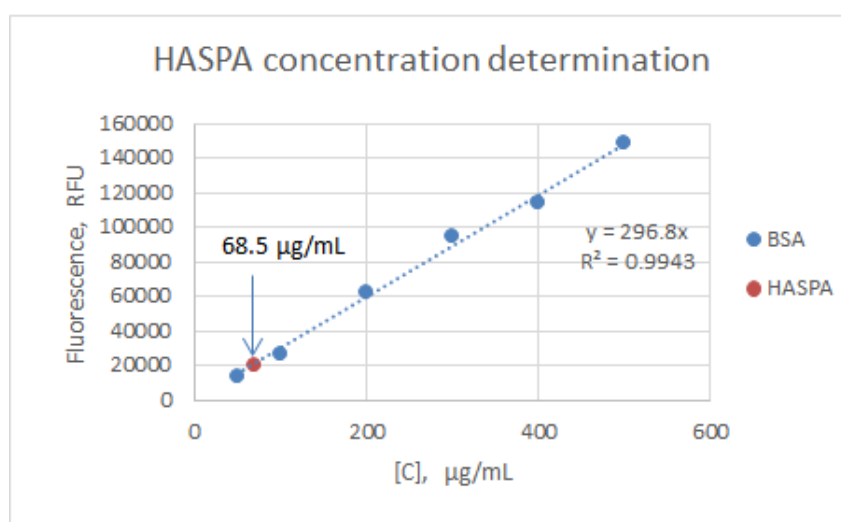
Figure 15. *Concentration determination of 1:100 diluted ^{15}N-HASPA sample using OPA assay with BSA as control sample.*

After the production of ^{15}N-labelled HASPA, an NMR fragment screening campaign was initiated.

2.3 NMR fragment screening

2.3.1 Protein control experiments

Before the actual fragment screening, d_6-DMSO, acetic acid and ethanolamine were titrated on pure ^{15}N-labelled HASPA to observe any changes that may be caused by solvation effects, or pH changes, associated with the potential presence of high concentrations of charged fragments in the screening samples (**Table 1**). The obtained NMR spectra would also serve as controls for evaluating the stability of the protein over time (**§7.4.2, §7.4.2.1**).

Table 1. Used concentrations of chemical compounds for ^{15}N-labelled HASPA in pH and DMSO control experiments.

Sample 1	Sample 2	Sample 3
d_6-DMSO, %	Ethanolamine, mM	Acetic acid, mM
1	0.25	0.25
2	0.5	0.5
3	1	1
4	2	2
5	4	4

The solvent-exposed residues are expected to be affected by the changes in the buffer solution (pH, salt concentration, organic solvents) more than their folded and buried counterparts. However, all of the residues may be affected if there are any significant conformational changes in the system. In case of IDRs and IDPs, the majority of amino acids are exposed to the solvent due to lack of compact 3D structure. Therefore, the majority of NMR resonance peaks were expected to be affected by the changed chemical environment. This is because the changes affect the unique local magnetic field of the nuclei which translates to chemical shift perturbations (CSPs) [127] .

The acid and base titration experiments concluded that buffering strength of 50 mM was enough to prevent pH changes of more than 0.2 pH units (**Figure 16**). The ΔpH limit of 0.2 was chosen to account for possible variations during pH measuring with a pH meter. In addition to this, it was important to notice any small pH changes which may influence the observed experimental results during fragment screening. Ethanolamine and acetic acid represented simple mono-charged species, as some fragments can be. Most importantly, the titration experiments have shown how a known concentration of acid and/or base affected HASPA and which CSPs were to be expected from charged fragments. The fragment mixtures used in the fragment screening were designed to contain a maximum of 4 equivalents of acid and/or base

moieties, which would translate to the top final concentration of 2 mM of ethanolamine or acetic acid in the screening solution due to 500 µM of each fragment in the mix.
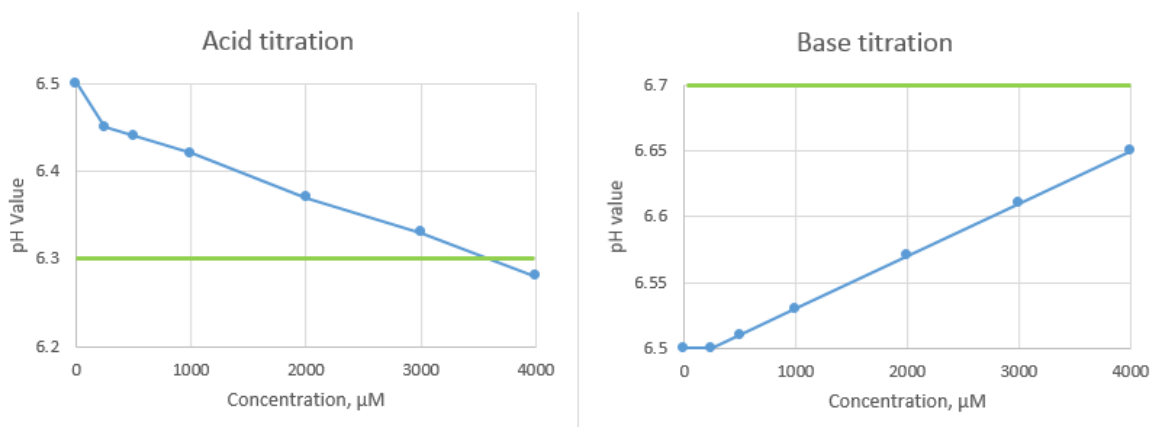
Figure 16. *The observed pH changes in 50 mM NaP$_i$ buffer solution at different concentrations of acetic acid and ethanolamine. Green line marks the pH change limit value of 0.2 units. The pH values were measured using bench-top pH meter (VWR) calibrated with standard pH solutions.*

Interestingly, several unique CSPs were observed for HASPA residues (H32, V33, Q34, H55 and N56) in 2D NMR spectra during acid and base titration (**Figure 17, Figure 18**). The CSPs, including the ones for unassigned X1-X4 cross peaks, were of similar level but in opposite vector which was dependent on the base/acid presence. The unassigned cross peaks are likely to be His residues from the C-terminal 6xHis tag protein which are expected to be sensitive to small pH changes in the observed range. The theoretical pKa of His side chain (pKa = 6.6) is close to the pH of the solution (pH = 6.5). Due to relatively large buffering capacity of buffer solution (50 mM), pH change was not expected to change by a significant amount, as confirmed experimentally (**Figure 16**). However, NH of imidazole in His is on the steepest part of the pKa titration curve. This means that even small pH changes can translate to observable CSPs due to population change (**Figure 19**).

The increasing concentration of d$_6$-DMSO has caused general CSPs in 2D ^{1}H-^{15}N SF-HMQC spectra where the majority of the cross peaks have shifted upfield in both dimensions (**Figure 17**). In addition to this, the CSPs were not of the same value for the backbone amides suggesting that some parts of primary amino acid sequence may have an inherent propensity to form secondary structures which can be affected by different additives in the solvent, such as methanol, DMSO, trifluoroethanol or SDS [128]–[130] (**Figure 18**). Moreover, the DMSO and water resonance peaks in ^{1}H spectra were not perturbed during titration indicating that the referencing was the same in all experiments (**Figure 20**).

^{1}H-^{15}N SF-HMQC
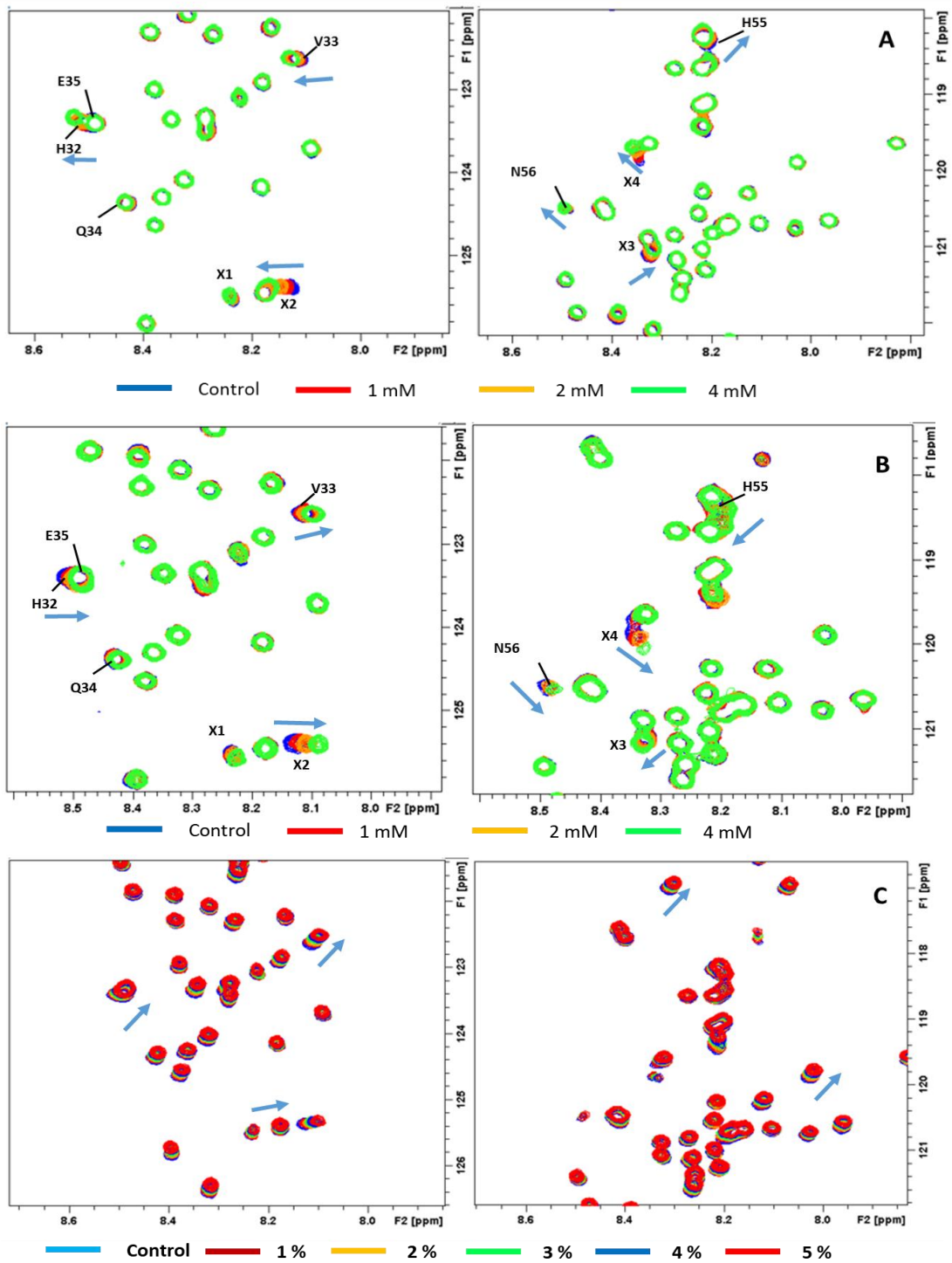
Figure 17. *The observed CSPs for 25 µM ^{15}N-HASPA in the presence of different concentrations of acetic acid (**A**), ethanolamine (**B**) and DMSO (**C**). The vectors of the observed CSPs are indicated with blue arrows.*
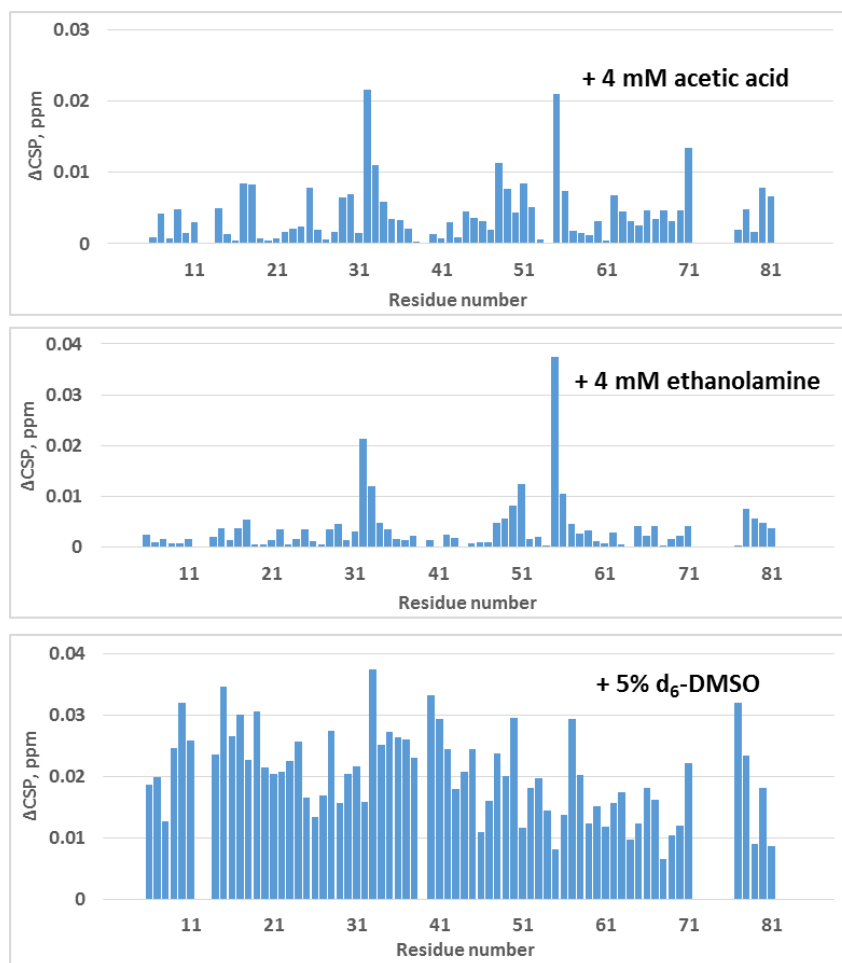
Figure 18. *A comparison of ΔCSPs for 25 μM ¹⁵N-HASPA amino acid residues in the presence of 4 mM acetic acid, ethanolamine and DMSO.*
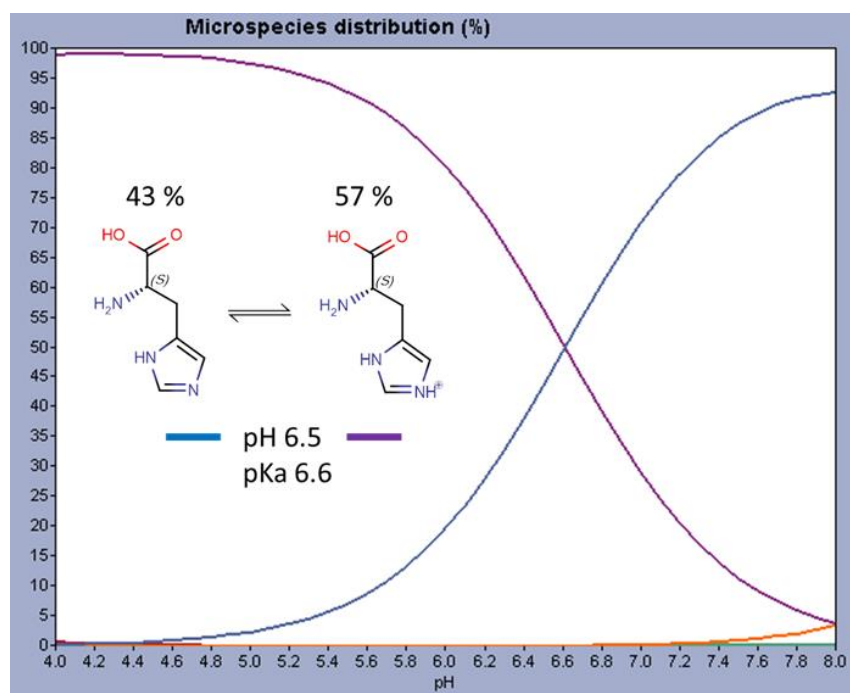
Figure 19. *The microspecies distribution of His at different pH levels (MarvinSketch).*
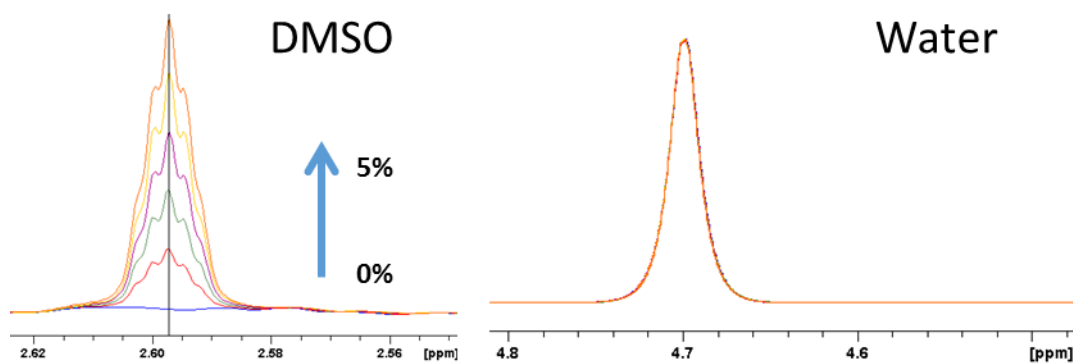
Figure 20. The resonance peaks of DMSO (Left) and water (Right) at different concentrations of d_6-DMSO (0-5%) in the presence of 25 μM ^{15}N-HASPA.

The protein control experiments, were followed by the acquisition of the reference ^1H and ^1H CPMG-filtered spectra for fragment mixtures.

2.3.2   Fragment reference experiments

The reference spectra without protein of 609 fragments in 93 mixtures were obtained using ^1H and ^1H CPMG-filtered NMR techniques in order to assign resonance peaks to specific compounds (**Figure 21**) (**§7.4.2**). The acquired spectra also enabled following of peak shape and intensity changes at different relaxation delays in the follow-up experiments with HASPA present (**§2.3.3**).
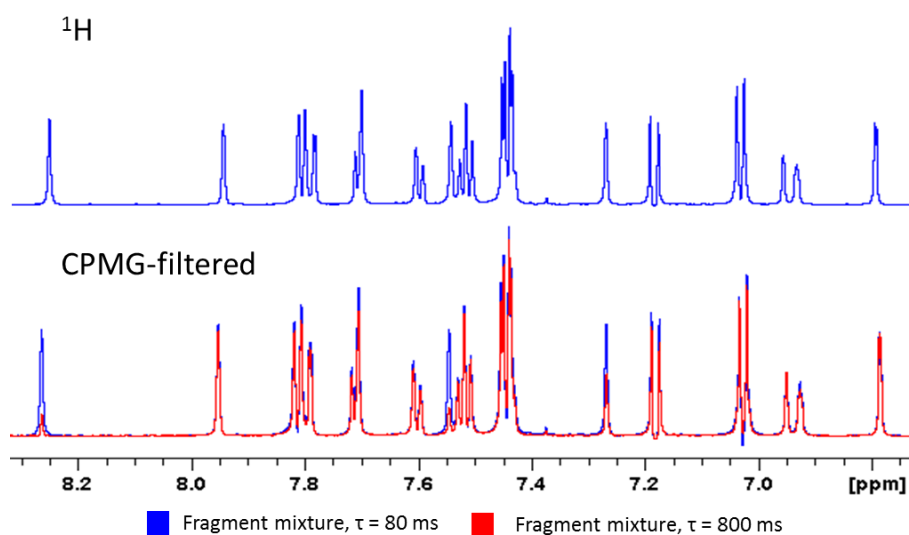
Figure 21. The reference ^1H and ^1H CPMG-filtered spectra for a fragment mixture.

It was noted that some fragment-specific peaks showed a partial decrease in intensity in ^1H CPMG-filtered data even without protein present (**Figure 22**). In general, LMW ligand resonance signals have slow T_2 relaxation rates which are directly affected by rotational correlation time [87]. Small organic molecules tumble relatively fast in solution which makes

the rotational correlation time low. In contrast, proteins have faster T_2 relaxation as they are rotating more slowly in the solution. The correlation time for ligand increases upon binding to the target protein. Therefore, T_2 value for ligand is increased which is observed as resonance peak broadening in ^{1}H CPMG-filtered spectra.
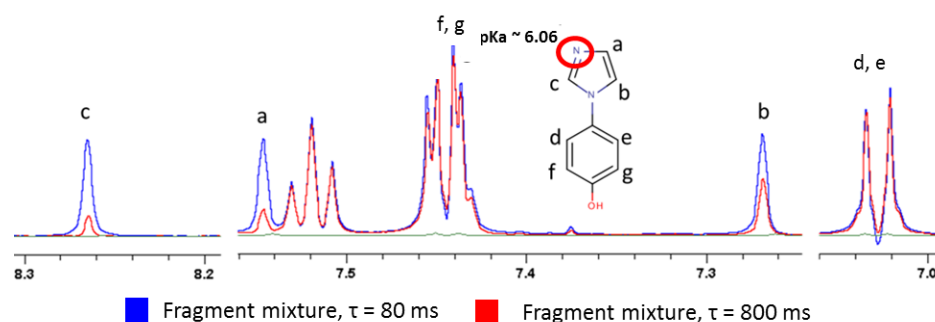
Figure 22. A comparison of ^{1}H CPMG-filtered data for a fragment at different delays showing increased proton relaxation in imidazole ring, but not in benzene ring.

The fact that the relaxation was observed without any protein present suggested the presence of internal events. The partial signal decrease can potentially be associated with chemical exchange due to the theoretical pKa value of the nitrogen in imidazole ring (pKa = 6.06) being relatively similar to the pH of the buffering solution (pH = 6.5). The protons, neighbouring a proton-exchangeable group, can have their T_2 relaxation rates affected as their chemical environment is perturbed, especially if pH is close to the pKa of the labile proton [127], [131]. The theoretical microspecies distribution chart suggests that the fragment of interest exists in two states with a ratio of 76:24 at pH 6.5 (**Figure 23**).
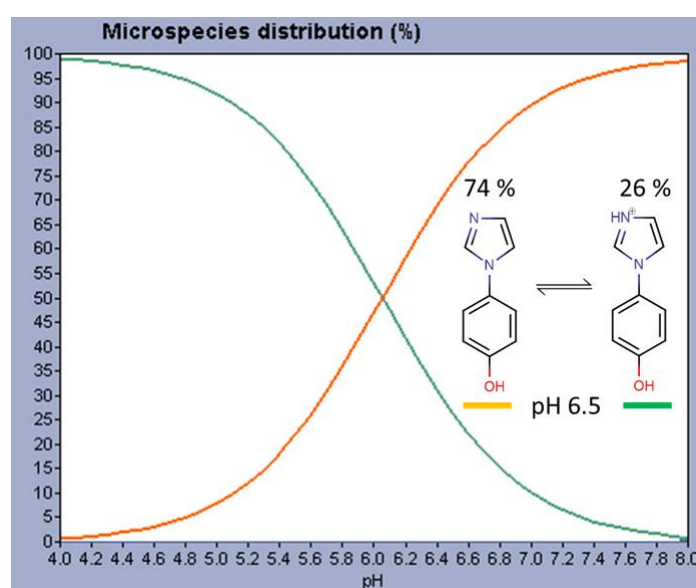
Figure 23. The distribution of microspecies of a compound with proton-exchangeable group at different pH values. (MarvinSketch)

The chart also suggests that any small pH changes may change the equilibrium towards a more protonated or deprotonated form due to the pH value being on the steep part of the curve.

Additional broadening of fragment signals without protein present may also be observed if the chemical exchange time constants are on the same timescale as the delays between pulses. This would cause significant number of spins to change between bound and non-bound states, and not get refocused completely. This also means that if a labile proton is exchanging between two states, the neighbouring protons will also have their T_2 relaxation affected due to the changed chemical environment which is what was observed.

With the reference spectra acquired, the fragment screening campaign was continued with ^{15}N-labelled HASPA present in the mixtures.

2.3.3 Fragment screening

Following the fragment mixture reference experiments, the fragment library was screened and evaluated against ^{15}N-labelled HASPA protein using 2D ^{1}H-^{15}N SF-HMQC and ^{1}H CPMG-filtered NMR techniques at fragment concentrations of 500 µM in the mixture.

It was noted that the observed CSPs in ^{15}N-labelled HASPA 2D SF-HMQC spectra were comparable to the ones observed during acetic acid and ethanolamine titration studies. Such results indicated possible pH- or charge-related artefacts (**Figure 17, Figure 18, Figure 24**).

^{1}H-^{15}N SF-HMQC
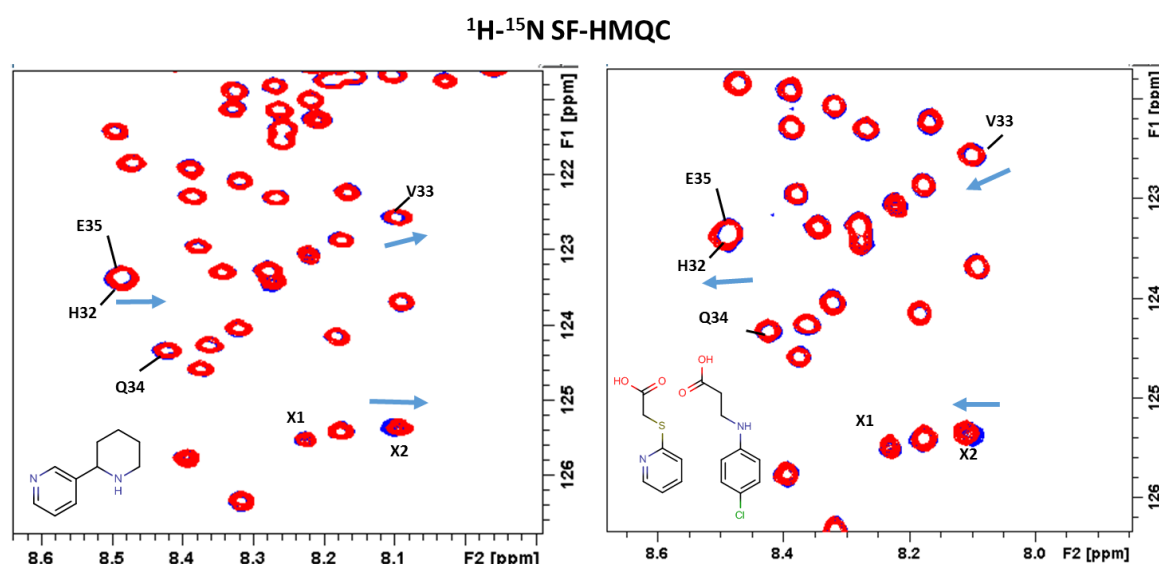
Figure 24. *A comparison of ^{15}N HASPA CSPs potentially caused by basic (**Left**) and acidic (**Right**) compounds in the fragment mixtures. Arrows mark the direction of cross peak movement.*

The significant enough pH changes can affect many protein structural characteristics, including hydrogen bonding, protein-solvent interactions and protonation state of amino acid side-chains [127]. In addition to this, the charged ligands can interact non-specifically with parts of protein of opposite charge which also can cause observable CSPs in NMR spectra. As mentioned previously, the fragment mixtures contained multiple basic and/or acidic compounds, suggesting that a slight perturbation of sample pH or charge-charge interactions could have caused some artefactual CSPs (**Figure 24**). As no compound-specific CSPs were observed, the potential hits were further selected using ^{1}H CPMG-filtered data.

The initial selection of fragment hits was based solely on the level of signal broadening after ^{1}H CPMG-filter (>30% intensity decrease). Due to a large number of potential hits (N = 52), the fragments had to be prioritized for follow-up 2D SF-HMQC and wLOGSY titration experiments, using the following prioritization steps:

1. A hit had to have a significant signal broadening when compared with reference ^{1}H CPMG-filtered spectra (>50% intensity decrease) (**Figure 25**)

2. If number of hits was still too large, a hit was expected not to have CSPs of more than half of a peak width in ^{1}H spectra when mixed with protein, as it would be indicative of a potential pH shift or aggregate formation (**Figure 26**) [132]

3. If number of hits was still too large, the theoretical pKa of a hit had to differ by more than 1.5 – 2 units from the pH of the solution avoiding mixed protonation states which would lead to potential false positives due to chemical exchange (**Figure 22**)
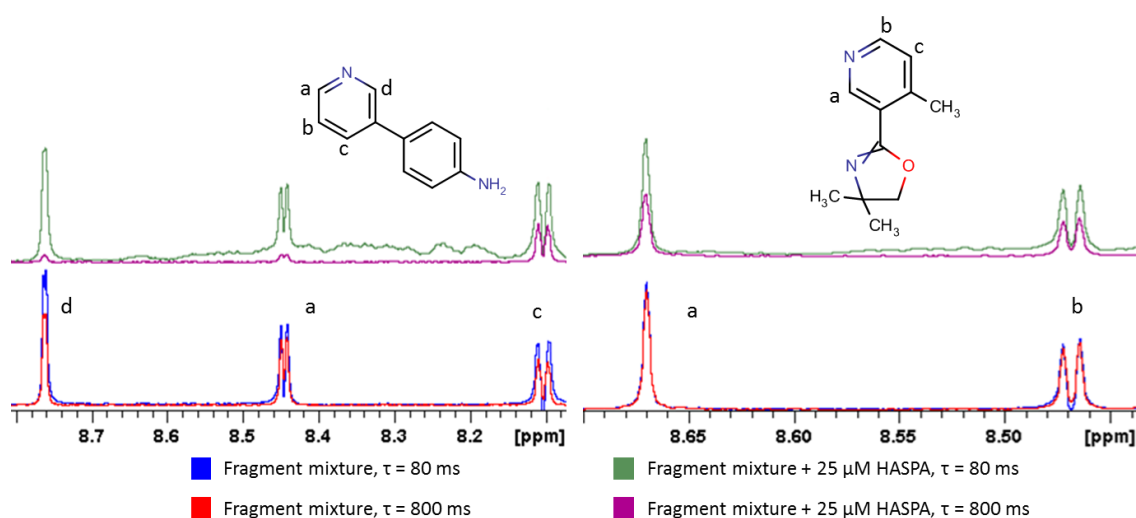
Figure 25. *A comparison of ^{1}H CPMG-filtered spectra of two fragments, where fragment A (**Left**) exhibits a larger signal broadening than fragment B (**Right**).*
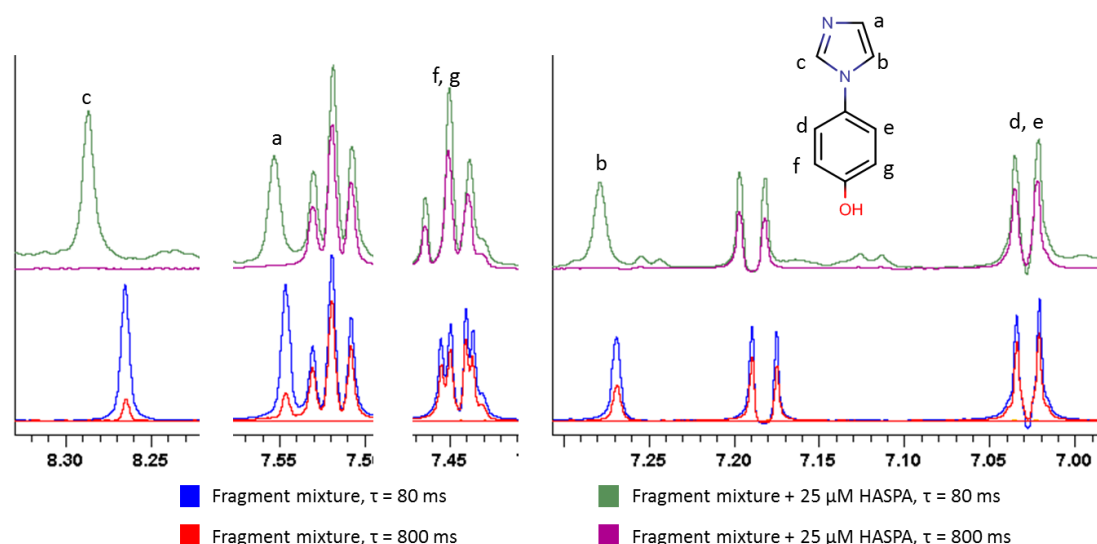
Figure 26. A comparison of ^1H CPMG-filtered spectra of a fragment in the presence and absence of ^{15}N-labelled HASPA.

It is important to mention that even though the observed signal broadening of ligand-specific resonances in the presence of ^{15}N-labelled HASPA was larger than in reference spectrum, the other ligand-specific resonances did not decrease for any of the prioritized hits (**Figure 25, Right, Figure 26**). The relaxation of all protons in the molecule was expected to be perturbed to some degree due to the general tumbling rate change upon binding. The observed local change could have been an indication that the chemical exchange rate of the ligand has changed due to pH shift or aggregation, as discussed previously (**§2.3.2**). It is also possible that HASPA had some influence on the pH due to its negatively charged nature (-10 net charge at pH 6.5) even at 50 mM buffering strength. The small pH changes associated with the presence of 25 μM HASPA could have shifted the ratio of protonated/deprotonated ligand which in turn influenced chemical exchange and T_2 relaxation rates of protons close to the exchangeable group. However, the hits with partial relaxation behaviour were still approved for titration experiments due to not knowing exactly if that was an actual artefact.

The prioritization of hits resulted in 20 fragments that were titrated as singletons up to 4 mM in the presence of ^{15}N-labelled HASPA. The selection included 9 fragments from Maybridge library, 4 fragments from ESR2 library (Hanna Klein, personal communication) and 7 outliers for control purposes (**Figure 27, Figure 28, Figure 29**). It is important to mention, however, that the majority of the prioritized hits that contained pyridine or imidazole moieties did not show a decrease in resonance intensity for all of the protons after the ^1H CPMG filter, as discussed previously (**Figure 22, Figure 26**).
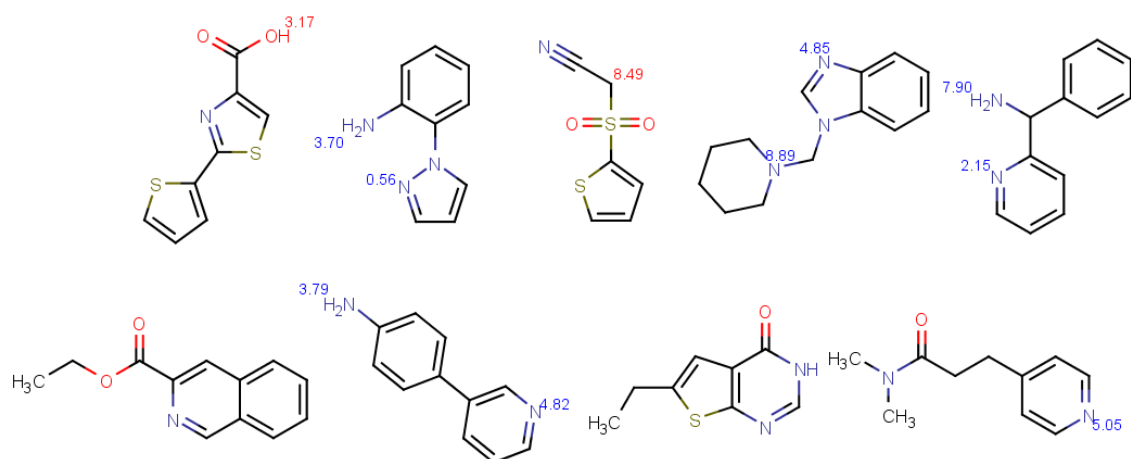
Figure 27. The chemical structures of selected compounds from Maybridge library. The theoretical pKa values were calculated using Marvin Sketch software.
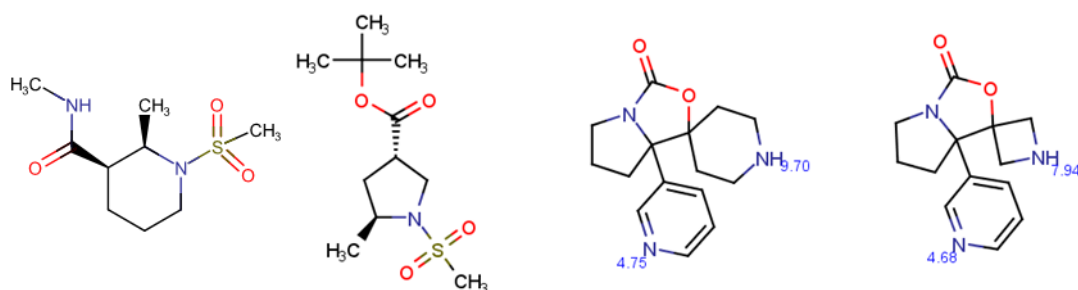
Figure 28. The chemical structures of selected compounds from ESR2 (Hanna Klein) library. The theoretical pKa values were calculated using Marvin Sketch software.
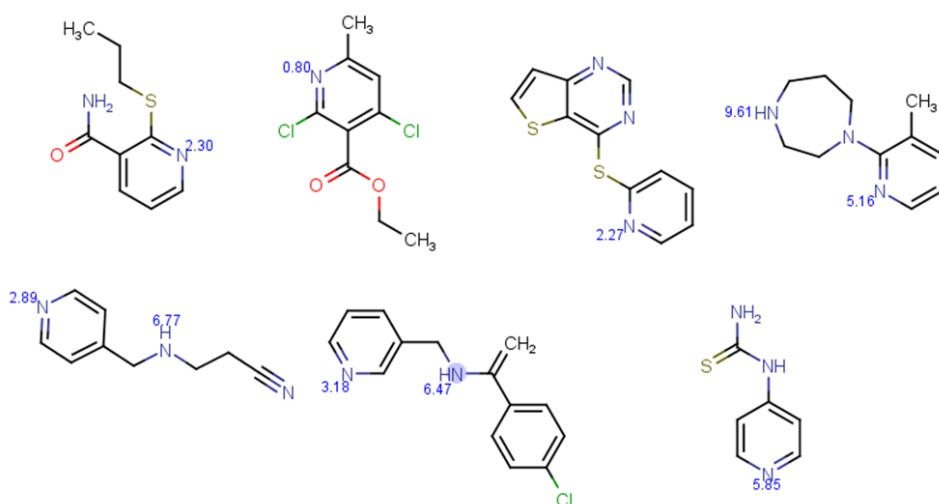
Figure 29. The chemical structures of outlier fragments to be used as controls. The theoretical pKa values were calculated using Marvin Sketch software.

No fragment-specific CSPs were observed during the titration experiments (**Figure 30**). Protein cross peaks in 2D SF-HMQC spectra were unperturbed at different fragment concentrations which was indicative of a non-binding ligand. The fragment-specific resonances increased in a concentration-dependent manner while having a negative phase in wLOGSY spectra which

was also consistent with no binding events to ^{15}N-labelled HASPA. The same level of partial signal broadening was also observed in ^{1}H CPMG-filtered spectra at different concentrations, as during the initial screen without any protein (**Figure 21, Figure 22**). This observation was again consistent with a non-binding ligand.
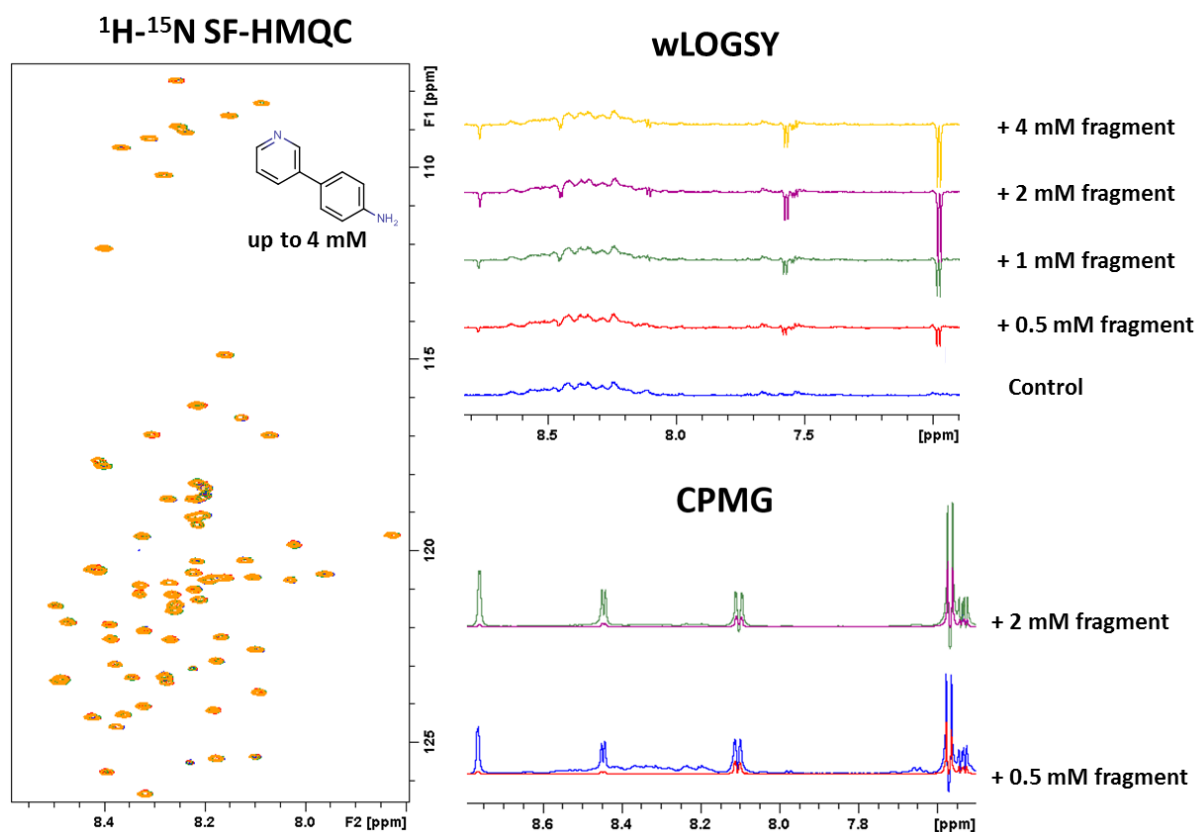
Figure 30. A representation of 2D SF-HMQC, wLOGSY, CPMG-filtered spectra in the presence of 25 μM ^{15}N-HASPA and different concentrations of a single fragment.

Additional data analysis revealed that there were an additional 23 pyridine and 13 imidazole ring-containing compounds in the initial fragment library that did not show any indicative binding signs in initial ^{1}H CPMG-filtered or 2D SF-HMQC spectra. Further SAR analysis revealed that the proton-exchangeable groups had their pKa values differing by more than 1.5 units from the pH of solution (pH = 6.5) which may have prevented the previously observed chemical exchange events and appearance as a hit in ^{1}H CPMG-filtered experiments (**Figure 31**). The non-hit pyridine-containing compounds had an *ortho*-substitution which included -S-, -NH$_2$, -NH-, -N=, -O-, -OH, -CH$_3$ and -Cl groups, modulating the theoretical pKa value of nitrogen significantly.
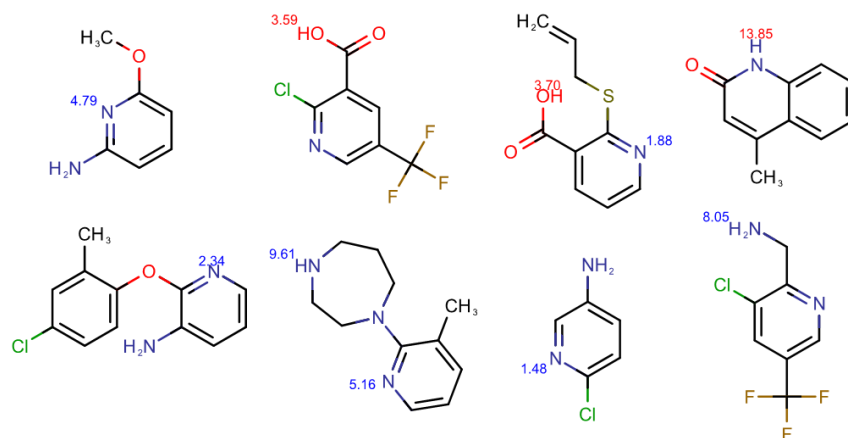
Figure 31. *The chemical structures of ortho-substituted pyridine-containing fragments (pKa values of pyridine nitrogen) that were not identified as hits in 2D SF-HMQC and ^{1}H CPMG-filtered experiments. The pKa values were calculated using Marvin Sketch software.*

In case of imidazoles, the non-hit fragments contained neighbouring substitutions as well (-S-, -CH₃, -Cl). However, even though the theoretical pKa of the imidazole nitrogen was closer to the pH of buffer solution (pH = 6.5), they were still identified as non-hits via ^{1}H CPMG-filtered experiments (**Figure 32**).
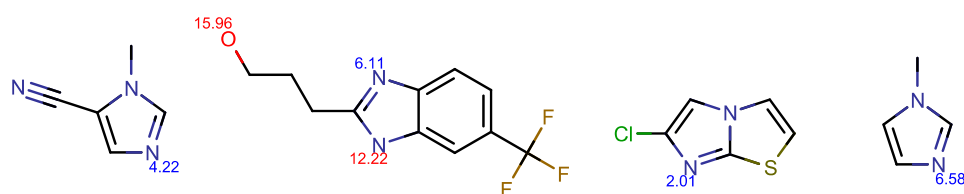
Figure 32. *Chemical structures of o-substituted imidazole-containing fragments (pKa values of imidazole nitrogen) that were not identified as hits in 2D SF-HMQC and ^{1}H CPMG-filtered experiments. The pKa values were calculated using Marvin Sketch software.*

The examples of pyridine and imidazole ring-containing fragments show that the awareness of potential similarity between theoretical pKa of labile protons and pH values is an important factor for objective ^{1}H CPMG-filtered result evaluation due to chemical exchange.

The fragment screening results concluded that the fragment library did not contain any ligands which interacted with ^{15}N-labelled HASPA in a specific manner; certainly, no compounds were identified from the screening library which showed data to be consistent with binding to HASPA, but rather all potential hits appeared to be false positives resulting from changes in chemical exchange associated with changes in the pH of the buffer solution. It is also possible that the affinities of fragments were too low to be detected. A recent example has shown that it is possible detect fragments with very low affinities (10^{-2} - 10^{-1} M)

by using reverse micelle encapsulation method by NMR which locally concentrates protein and fragments and increases the possibility of interactions [133].

2.4 Conclusions

^{15}N-labelled HASPA was produced and screened against a fragment library using 1D ligand-observed and 2D protein-observed NMR approaches. The importance of reference experiments was shown by noting that specific ^{15}N-labelled HASPA CSPs were caused by basic or acidic fragments due to a slight pH change. After data analysis and prioritization steps, no verified hits were identified using both 1D and 2D NMR approaches.

One of main observations was that if the pKa value of the proton-exchangeable group in the molecule is close to the pH of the solution, an increased relaxation rate should be expected for neighbouring protons in ^{1}H CPMG-filtered experiments. This is explained by change of chemical exchange which in turn affects T_2 relaxation of the molecule. Therefore, these parameters should be taken into consideration when designing and evaluating the ^{1}H CPMG-filtered experiments and results to avoid false positives. In order to achieve that, the in-depth knowledge of the NMR techniques and physical processes is required.

It can be suggested that a higher MW fragment library may provide higher chances of finding a specific hit due to a higher number of potential interactions possible thanks to the larger structure. In addition to this, it may just be that some of the tested fragments are actually binding to ^{15}N-labelled HASPA, just with affinities too low to be detected by NMR.

3 cMyc

3.1 Introduction

According to the World Health Organization (WHO), cancer is the second leading cause of death in the world with almost 10 million deaths from it in 2018 [134]. The risk of cancer is increased with tobacco, alcohol use, unhealthy diet and physical inactivity. While there are multiple treatments already available for various carcinomas, the research on better understanding of oncogenic malignancies and treatments is still ongoing. While some cases of oncogenic diseases are related to specific mutations in proteins that cause the over-activation of cell signalling pathways, the processes of gene amplification, chromosome rearrangement and protein overexpression are also involved [135].

cMyc is an oncogenic intrinsically disordered transcription factor which dimerizes with its intrinsically disordered partner protein Max via basic-helix-loop-helix-leucine-zipper (bHLHZip). The formed dimer interacts with E-box sequence (5'- CACGTG-3') in cells and promotes gene expression (**Figure 33**) [136], [137]. It has been determined that the majority of cancerous tumours are dependent on cMyc expression to variable degree and the suppression of elevated cMyc downstream interactions may inhibit tumour growth [138].
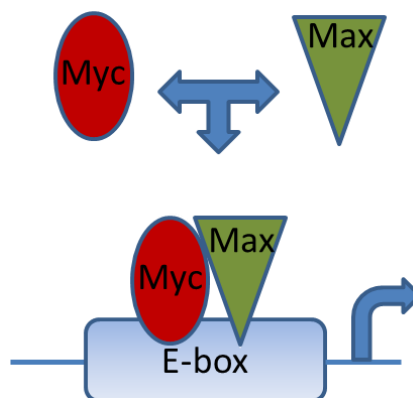
Figure 33. A graphical representation of a cMyc-Max dimer formation and the promotion of gene expression.

The inhibition of downstream gene expression may be achieved in several ways. cMyc or Max proteins can be targeted separately with a LMW ligand in order to disrupt the occurring PPIs (**Figure 34, Left**). It has been noted, however, that there may be several potential problems with specifically targeting cMyc protein. The general lack of cMyc mutations in oncogenic cells may cause undesired toxicities because cMyc is expressed in all proliferating cells. Nevertheless, several exceptions have been identified for Burkitt's lymphoma with specific

mutations to Thr58 and Ser62 of cMyc [139]. Moreover, the concerns about toxicity may not be profound, because cMyc is overexpressed only in oncogenic cells, and non-mutated cells would possibly not be affected by cMyc-targeting drugs due to generally low cMyc levels.
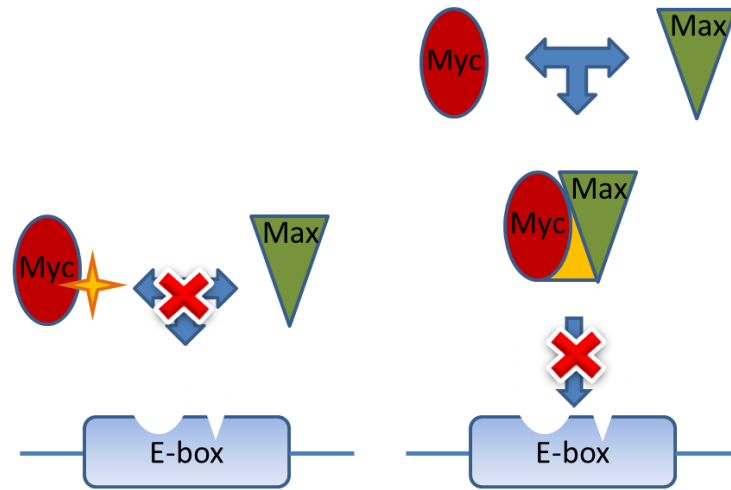
*Figure 34. Potential inhibitory pathways of gene expression: an inhibition of dimerization (**Left**) and inhibition of dimer interactions with E-box (**Right**).*

Alternatively, the formed cMyc-Max dimer itself could be targeted in order to prevent it from interacting with its target DNA sequence (**Figure 34, Right**). It has been shown previously that single amino acid mutations can disrupt cMyc-Max dimer formation. This suggests that LMW ligands could potentially also have a similar effect if it targets 'hotspots' that are essential for high affinity interactions, as shown for Bcl-2-Bak system [140], [141].

Since the first cMyc and Max dimerization inhibitors were discovered back in 2002, multiple investigations have been performed in search of additional LMW ligands that disrupt the dimerization process (**Figure 35**) [66]. 10058-F4 and 10074-G5 were identified to inhibit the formation of cMyc-Max dimers using yeast two-hybrid approach [142]. This was followed by a more in-depth structural study which identified 10074-G5 interacting with cMyc(363-381) and 10058-F4 with cMyc(402-412) with micromolar affinities using NMR and CD approaches [69]. Later, 10074-G5 and 10058-F4 were determined to interact with cMyc(353-437) by SPR with K_D of 31.7 and 39.7 μM, respectively [143]. A more soluble 10074-G5 derivative JY-3-094 was developed to interact with cMyc peptide via a pharmacophore identification approach [144]. In addition to this, another 10074-G5 derivative JKY-2-169 was designed to interact specifically with cMyc-Max dimer complex by mimicking an α-helix [145]. Multiple other molecules were also identified to perturb cMyc-Max dimer formation (**Figure 35**). E07 and

N12, derivatives of 10058-F4 and 10074-G5, were identified to form dimers and interact with cMyc bHLHZip domain with increased affinity, as determined by SPR [146].
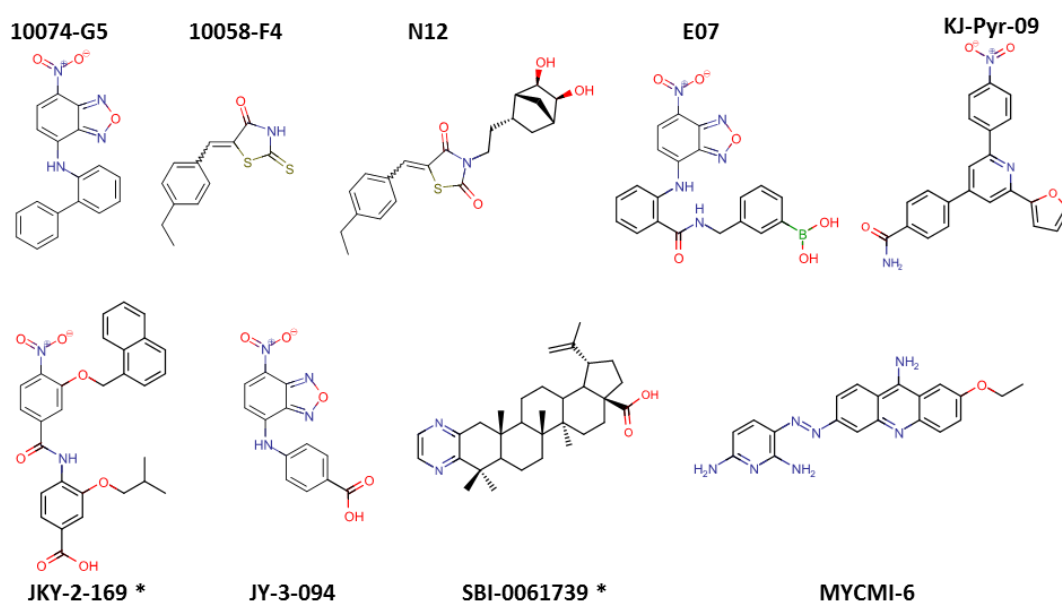
Figure 35. Chemical structures of compounds that were identified to directly interact with cMyc or cMyc-Max dimer (*).

KJ-Pyr-9 was noted to bind to both monomeric cMyc and cMyc-Max dimers with K_D of 6.5 nM, as determined by backscatter light interferometry, and prevent cMyc-Max dimer interactions with its target DNA sequence [147]. However, low solubility of the molecule prevented the use of other biophysical methods to better understand the binding mode. Interestingly, the results could not be reproduced with SPR technology, suggesting some assay-dependent differences [148]. Celastrol and its derivatives were also noted to disrupt the structure of preformed cMyc-Max dimers, and disrupt the dimer's DNA binding ability, as shown by 2D NMR and SPR [149]. However, more robust biophysical data are required to better understand the interactions occurring. MYCMI-6 was identified to interact with cMyc using cell-based protein interaction assay and NCI/DTP Open Chemical Repository library [148]. The specific interactions to cMyc were also observed using MST and SPR approach with K_D of 1.6 µM.

In this chapter, multiple cMyc peptide production approaches are described. The attempts of cMyc peptide production are followed by characterization of several cMyc binders using commercially available cMyc peptides by NMR in order to probe and learn more about IDP:ligand interactions and possible caveats.

3.2 Peptide production as soluble fusion protein

The production of cMyc(363-381) and (402-412) peptides was initiated to enable structural investigation of the interactions between the peptides and literature-reported ligands. Typically, short peptide synthesis can be achieved using solid-phase peptide synthesis via well-established coupling and deprotection chemistry [150]. However, if isotopically labelled peptides are required for structural NMR studies, the synthetic approach can be costly due to high prices of individual isotopically labelled amino acids. By fusing the peptides with a well-expressed fusion partner it is possible to produce isotopically labelled proteins in high yields. After releasing the peptides from its fusion partner, further purification can be performed via HPLC.

Initial soluble fusion protein constructs were designed to contain cMyc peptides at the C-terminus of well-expressed KRas protein with an enterokinase (EK) cleavage site (**DDDDK|**) (**Figure 36**). The selected fusion partner KRas contains an unstructured C-terminal region which was used to attach cMyc peptides to [151]. EK was chosen due to its ability to remove the peptide from fusion protein without leaving any residual amino acids at the N-terminus of target peptide.

cMyc(363-381): ERQRRNELKRSFFALRDQI

cMyc(402-412): YILSVQAEEQK
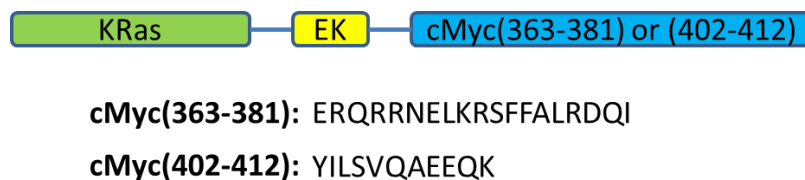
Figure 36. A representative scheme of a soluble fusion protein construct with C-terminal cMyc peptides.

The designed constructs were expressed and purified in acceptable yields of ~30 mg for cMyc(363-381) construct and ~20 mg for cMyc(402-412) from 10 L of LB media, as determined with A_{280} (**Figure 37**) (**§7.3.1**, **§7.3.3, §7.3.6**).
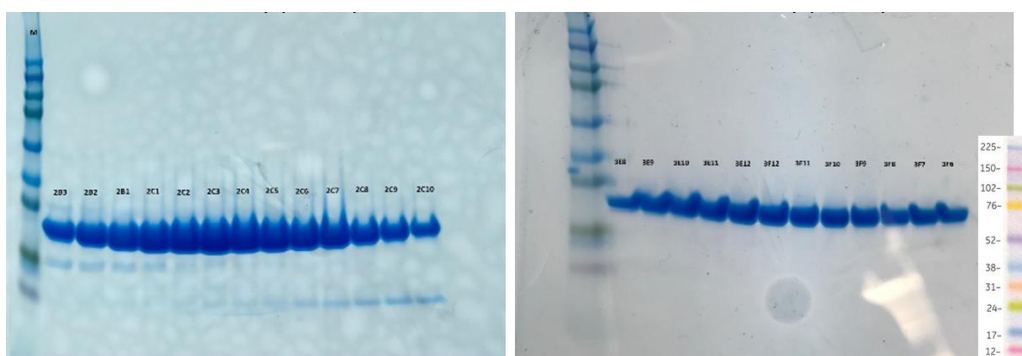
*Figure 37. The QC step of KRas-EK-cMyc(363-381) **(Left)** and KRas-EK-cMyc(402-412) **(Right)** using SDS-PAGE.*

Following this, the protein samples were incubated in EK cleavage buffer with recommended amount of EK (1200 units for cMyc(363-381) and 800 units for cMyc(402-412)) for 16 hours to release cMyc peptides from the fusion protein (**§7.3.3.2**). HPLC-MS detected only trace amounts of cMyc(363-381) and (402-412) peptides indicating low EK cleavage efficiency *in vitro* (**Figure 38**). It is possible that the cleavage site for EK was conformationally restricted which in turn prevented efficient cleavage.

In addition to this, while the theoretical MW of cMyc(402-412) matched the experimental MW, the experimental MW of cMyc(363-381) was determined to be lower than the theoretical MW (**Table 2**). It was calculated that cMyc(363-381) peptide was truncated by two amino acids at C-terminus, namely QI. It is possible that the protein or peptides started to degrade during cleavage process. EK has been shown previously to cleave non-specifically in some protein systems [152]. Moreover, low activity of EK also indicated that the cleavage site was potentially not as accessible as initially expected.
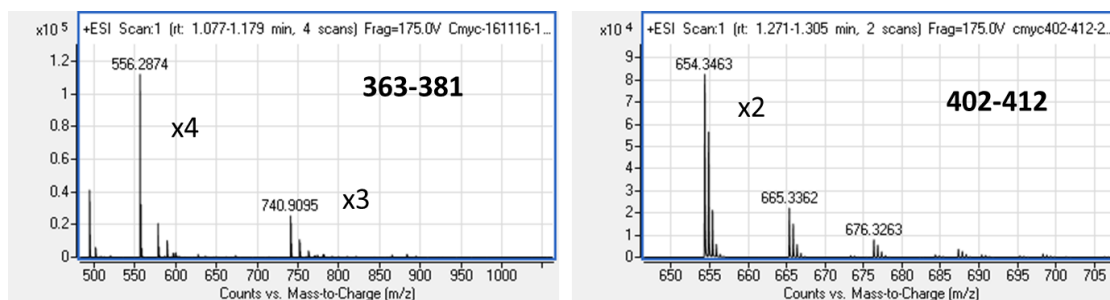
Figure 38. HPLC-MS spectra of digested fusion protein with traces of cMyc(363-381) and 402-412 peptides.

Table 2. The theoretical and experimentally determined MWs of cMyc peptides.

Peptide	Theoretical MW, Da	Experimental MW (as H^+ adduct), Da
cMyc(363-381)	2462.8	2222.7
cMyc(402-412)	1307.5	1308.7

As a result, the soluble fusion protein approach with enzymatic cleavage was not satisfactory. This was followed by an attempt to express cMyc peptides as an insoluble fusion protein using ketosteroid isomerase (KSI) [153], [154].

3.3 Peptide production via insoluble fusion protein

KSI readily forms inclusion bodies upon expression which can protect the attached peptides from proteolytic cleavage *in vivo*. C-terminal 6xHis tag is usually used to purify the KSI fusion protein under denaturing conditions before chemical cleavage with CNBr via methionine

residues. In this case, the introduction of C-terminal 6xHis tag was not an option due to the short length of cMyc peptides (**Figure 39**). It was predicted that the produced inclusion bodies would be pure enough for further downstream processing. The fusion constructs were designed to contain C-terminal cMyc peptides which would then be chemically cleaved from KSI using CNBr via two available methionine residues (M1 and M126).

A

```
        10         20         30         40         50         60
MHTPEHITAV VQRFVAALNA GDLDGIVALF ADDATVEDPV GSEPRSGTAA IREFYANSLK

        70         80         90        100        110        120
LPLAVELTQE VRAVANEAAF AFTVSFEYQG RKTVVAPIDH FRFNGAGKVV SIRALFGEKN

       130        140
IHACQMERQR RNELKRSFFA LRDQI
```

B

```
        10         20         30         40         50         60
MHTPEHITAV VQRFVAALNA GDLDGIVALF ADDATVEDPV GSEPRSGTAA IREFYANSLK

        70         80         90        100        110        120
LPLAVELTQE VRAVANEAAF AFTVSFEYQG RKTVVAPIDH FRFNGAGKVV SIRALFGEKN

       130
IHACQMYILS VQAEEQK
```

*Figure 39. The KSI constructs with C-terminal cMyc(363-381)(**A**) and cMyc(402-412)(**B**). cMyc peptides are marked in yellow and methionines in red.*

After obtaining cMyc constructs in pET-31b plasmid from commercial sources, the expression and purification of fusion proteins was performed, as indicated by SDS-PAGE and HPLC-MS data (**Figure 40**) (**§7.3.1, §7.3.3.2**).
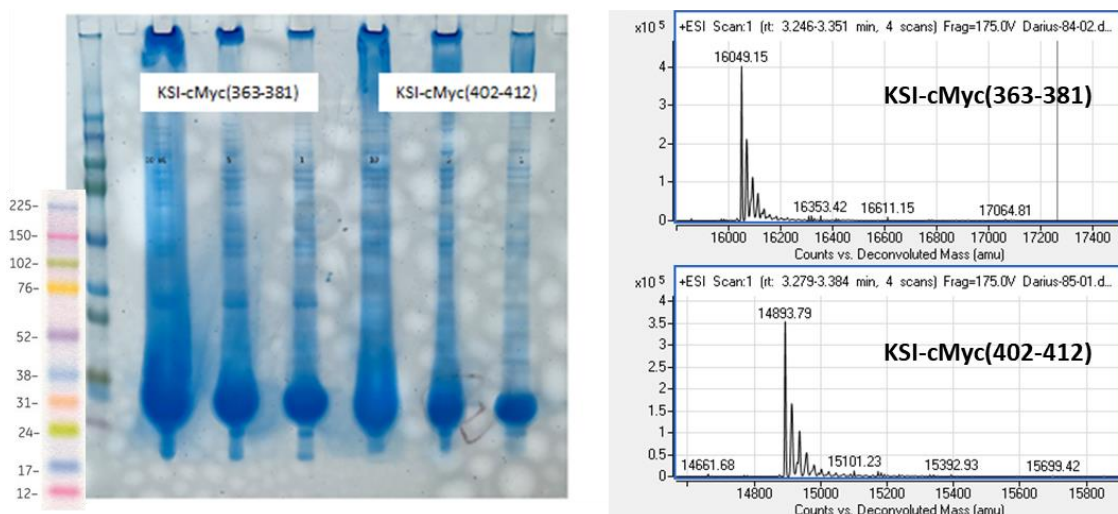
*Figure 40. SDS-PAGE gel images of washed pellets of KSI-cMyc(363-381), (402-412) (**Left**) and HPLC-MS QC data (**Right**).*

After pellet extraction and washing, the pellets were solubilized in ~5 mL of Buffer E (**§7.3.3.2**). The concentration was estimated by A_{280} to be ~12 mg/mL (740 µM) for cMyc(363-381) and 7 mg/mL (470 µM), for cMyc(402-412). The cleavage reaction was performed in 70% (v/v) formic acid in the presence of 100-fold molar excess of CNBr over methionine residues. Because the protein sequence contains two methionine residues, it was calculated that the chemical cleavage would require the final concentration of CNBr to be at 148 mM for cMyc(363-381) and 94 mM for cMyc(402-412) (740 µM x two methionine residues x 100 excess of CNBr = 148 mM of CNBr for KSI-cMyc(363-381). Same reasoning was applied to KSI-cMyc(402-412)).

After the overnight reaction only low amounts of intact MW cMyc peptides were detected via HPLC-MS (**Figure 41**). It was also noted that cMyc(402-412) construct had reacted more efficiently than 363-381 construct. However, HPLC-MS spectra for both constructs indicated a significant amount of noise peaks which would make the downstream purification difficult.
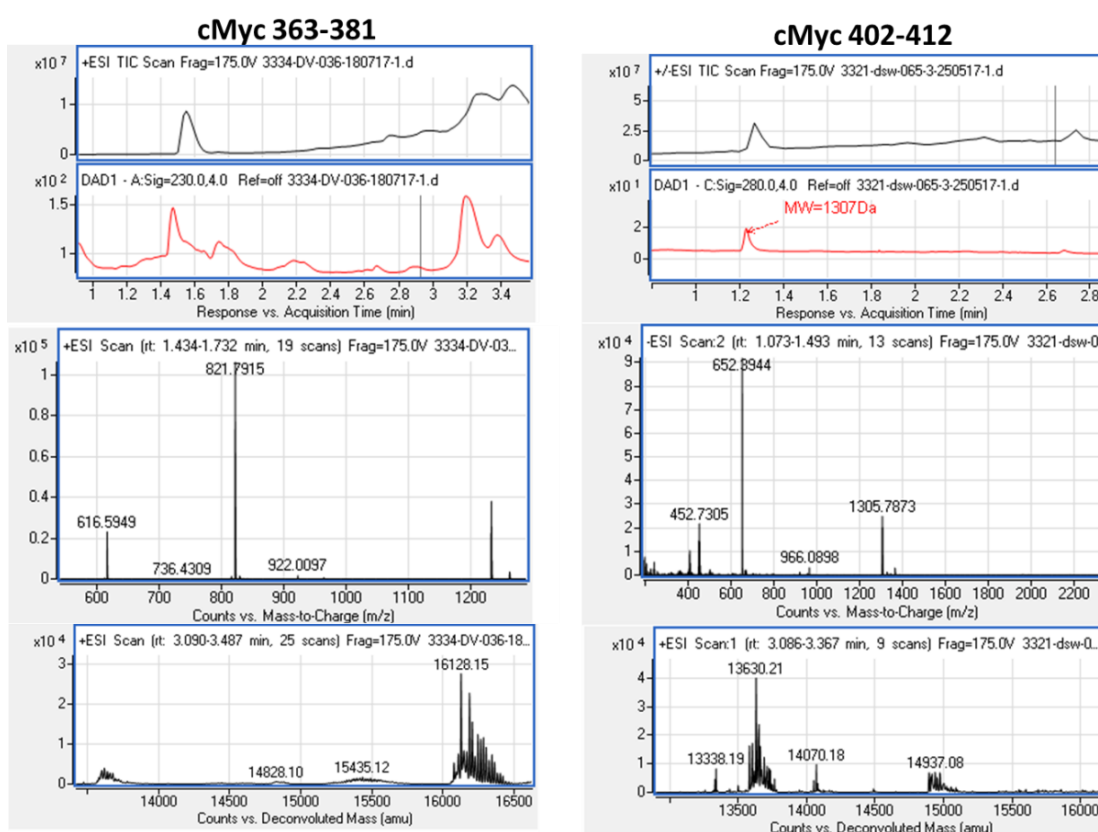
Figure 41. The HPLC-MS data of cMyc samples after cleavage with CNBr.

The inefficient cleavage may have occurred due to the sample being only partially pure, which resulted in the presence of too many methionine residues from contaminants, even with 100x

molar excess of CNBr (**Figure 40**). A possible solution to this would be to introduce N-terminal 6xHis tag to KSI and further purify the construct using IMAC under denaturing conditions in urea. This would produce a cleaner sample to be used for CNBr cleavage more efficiently. As an alternative, the expression of soluble longer constructs of cMyc protein would also allow the direct use of 6xHis tag for purification purposes.

It was concluded that the insoluble protein approach exhibited several problems, such as low yield cleavage, which was probably caused by not pure enough initial sample. Due to prioritization of other projects, it was decided to stop any further method development for cMyc peptide production.

3.4 Literature compound evaluation

At this stage, it was decided to test commercially available unlabelled cMyc peptides with the readily available literature-reported ligands using NMR. This was done in order to evaluate if further development of production of the peptides for isotopic labelling for NMR experiments is reasonable (**Figure 42**) (**§7.4.2**).
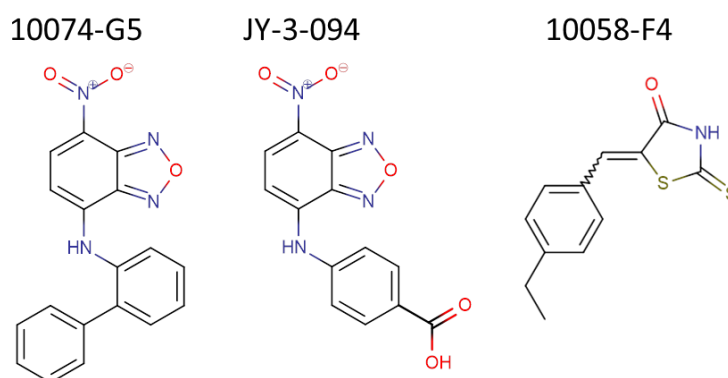
*Figure 42. Literature-reported ligands for cMyc(363-381) (**10074-G5, JY-3-094**) and cMyc(402-412) (**10058-F4**).*

10074-G5 and 10058-F4 were determined to have limited low aqueous solubility (80 and 60 µM, respectively) which agreed with previous literature observations [143]. Carboxylic derivative of 10074-G5, JY-3-094, was determined to be soluble up to 4 mM.

10058-F4 and 10074-G5 ligands have been previously shown to interact with cMyc peptides using protein-observed 2D NMR. However, there was lack of information about 1D ligand-observed NMR approaches [69]. Interestingly, the concentrations of the ligands used in literature-reported 2D protein-observed NMR experiments were around eight-fold higher than the determined aqueous solubility in this project [69]. This raised concern about the

quality of the data, experimental conditions and the chemical integrity of compounds. There are multiple publications that report lower solubility of 10058-F4 and 10074-G5 compounds than the original paper [143]. In addition to this, the compounds are PAINS which have been shown to give false positive in high throughput assays and could be interacting with cMyc in a non-specific way.

The 1D ligand- and protein-observed NMR data suggested no observable interactions between 10074-G5 for cMyc(363-381) and 10058-F4 for cMyc(402-412). ^{1}H STD experiments are usually performed at higher ligand:protein ratios, such as 25:1 or 50:1, but due to limited aqueous solubility, the ligand:protein ratio of 1:1 was used (**Figure 43, Figure 44, Figure 45**).

■ 50 μM cMyc(363-381), τ = 800 ms
■ 50 μM cMyc(363-381), τ = 200 ms
■ 50 μM cMyc(363-381) + 50 μM 10074-G5, τ = 800 ms
■ 50 μM cMyc(363-381) + 50 μM 10074-G5, τ = 200 ms
■ 50 μM 10074-G5, τ = 800 ms
■ 50 μM 10074-G5, τ = 200 ms
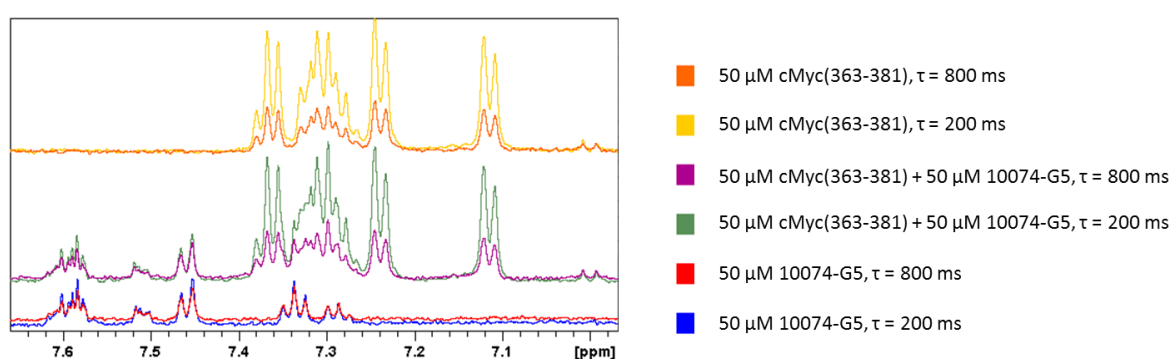
Figure 43. A comparison of ^{1}H CPMG-filtered NMR data for cMyc(363-381) and 10074-G5 at variable τ values.

■ 50 μM cMyc(402-412), τ = 800 ms
■ 50 μM cMyc(402-412), τ = 200 ms
■ 50 μM cMyc(402-412) + 50 μM 10058-F4, τ = 800 ms
■ 50 μM cMyc(402-412) + 50 μM 10058-F4, τ = 200 ms
■ 50 μM 10058-F4, τ = 800 ms
■ 50 μM 10058-F4, τ = 200 ms
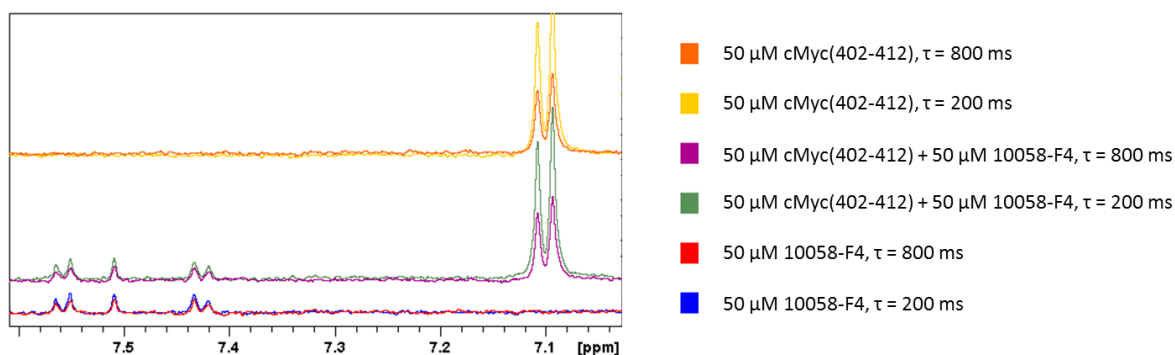
Figure 44. A comparison of ^{1}H CPMG-filtered NMR data for cMyc(402-412) and 10058-F4 at variable τ values.

■ 50 μM cMyc(402-412) + 50 μM 10058-F4
■ 50 μM cMyc(363-381) + 50 μM 10074-G5
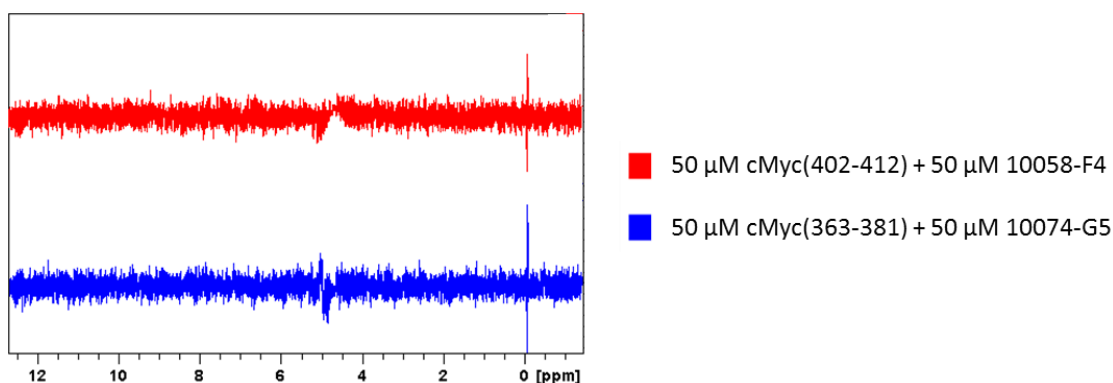
Figure 45. A comparison of ^{1}H STD NMR data in the presence of cMyc constructs and their respective literature compounds.

JY-3-094, a derivative of 10074-G5, was noted to produce CSPs for Phe374 and Phe375 aromatic and aliphatic protons in cMyc(363-381), as reported in earlier studies [155]. Other easily distinguishable CSPs were also observed in the methyl region of the peptide, representing Leu370 and Leu377 side chain Hδs, as determined by 2D ^{1}H-^{1}H TOCSY data. (**Figure 46**). The CSP data for Phe374, Phe375 aromatic protons was fitted to 1:1 binding model (SimFit) and yielded K_D of 5.5 and 5.6 mM, respectively (**Figure 47**).

■ 100 μM cMyc(363-381)   ■ 100 μM cMyc(363-381) + 4 mM JY-3-094
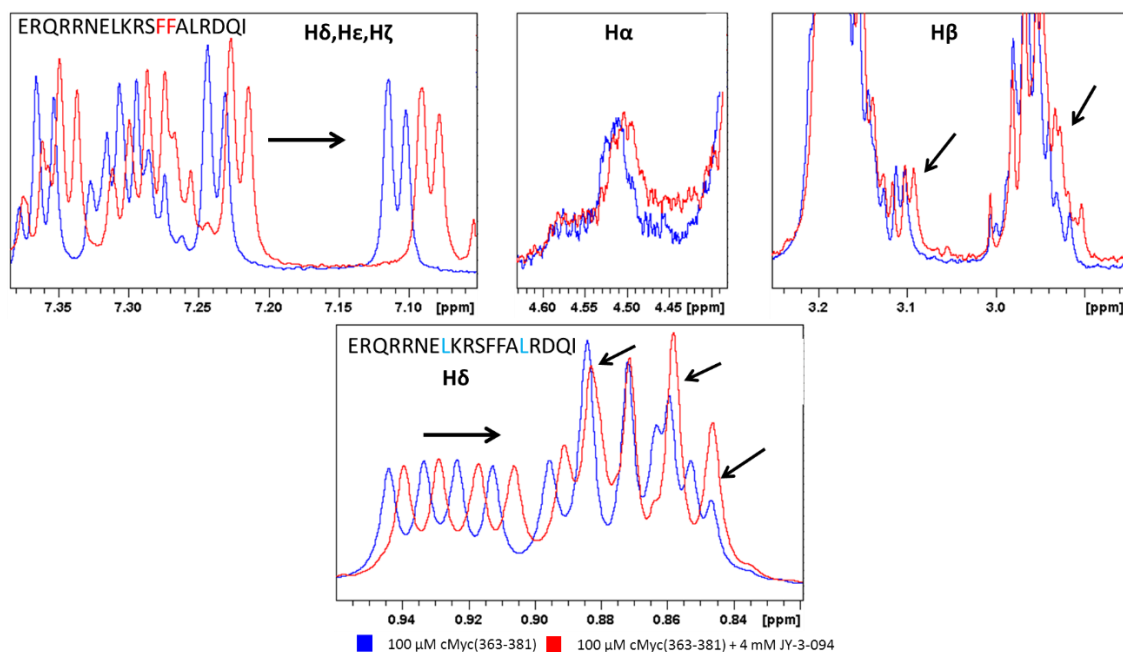
Figure 46. *A comparison of cMyc(363-381) ^{1}H spectra in the absence and presence of JY-3-094. The black arrows mark the affected amino acid protons and chemical shift direction.*
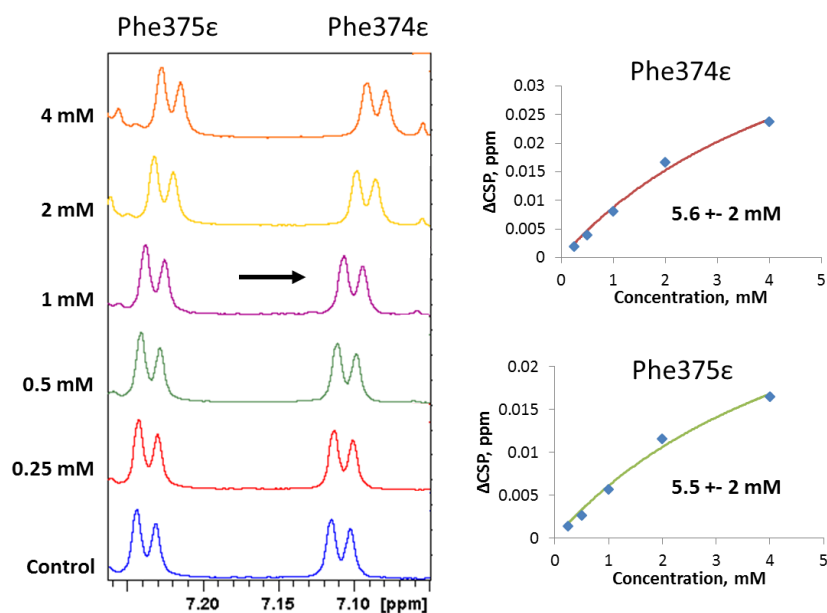
Figure 47. *The observed CSPs for Phe374ε and Phe375ε protons of cMyc(363-381) in the presence of JY-3-094 with determined K_D values. The black arrow indicates the vector of CSPs.*

The initial buffer solution that the experiments were performed in did not contain any salt during this exploratory phase. Therefore, the cMyc:ligand interactions were also assessed in the presence of 200 mM NaCl. The observed CSPs decreased which suggested that the interactions were partially charge-based (**Figure 48**).
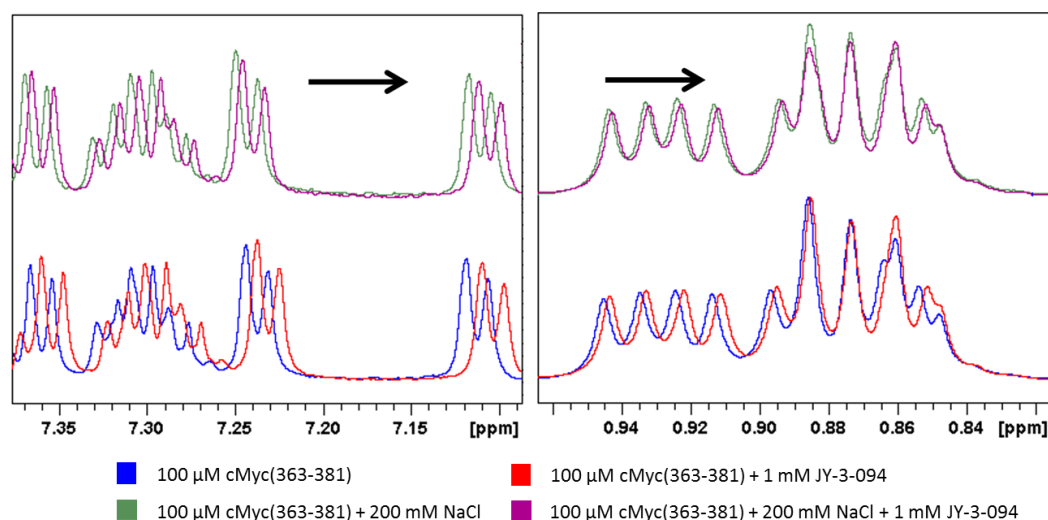
Figure 48. *A comparison of cMyc(363-381) ^{1}H spectra in the absence and presence of JY-3-094 at 0 and 200 mM of NaCl. The black arrows mark the affected amino acid protons and chemical shift direction.*

The charged interactions of JY-3-094 with cMyc(363-381) may be driven by carboxylic acid moiety (**Figure 42**). In addition to this, due to Phe signals being affected the most, aromatic π-π stacking contributions cannot be disregarded. Aromatic nitro groups on ligands have also been reported to affect the strength of aromatic stacking [156].

The follow-up ^{1}H CPMG-filtered data did not reveal any increased T_2 relaxation rate for JY-3-094 or cMyc(363-381) peptide (**Figure 49, Figure 50**). In addition to this, the wLOGSY data also did not indicate any specific protein:ligand interactions (**Figure 51**).
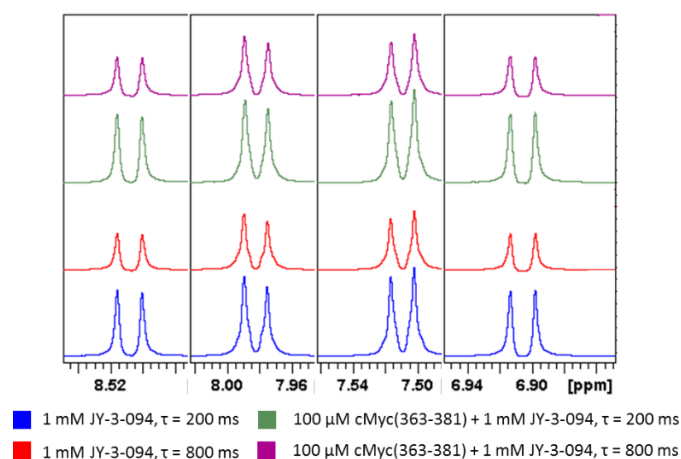
Figure 49. *A comparison of ^{1}H CPMG filtered JY-3-094 spectra in the absence and presence of cMyc(363-381) at different delays.*
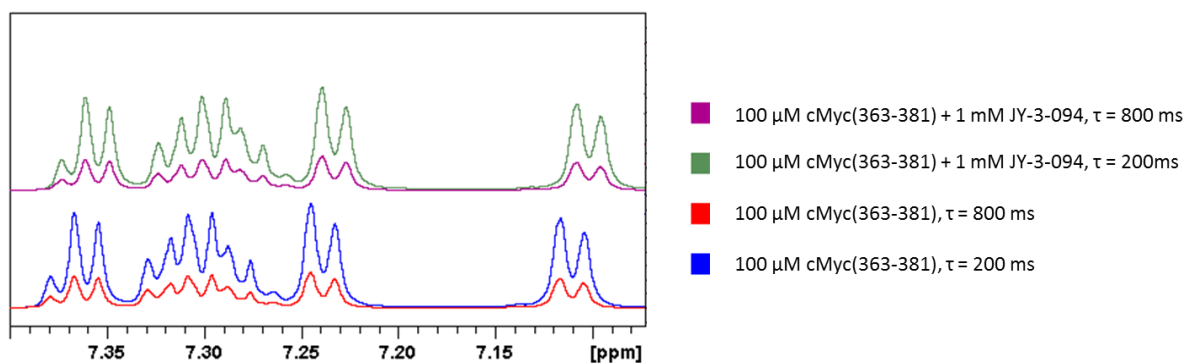
Figure 50. *A comparison of ^1H CPMG filtered cMyc(363-381) spectra in the absence and presence of JY-3-094 at different delays.*
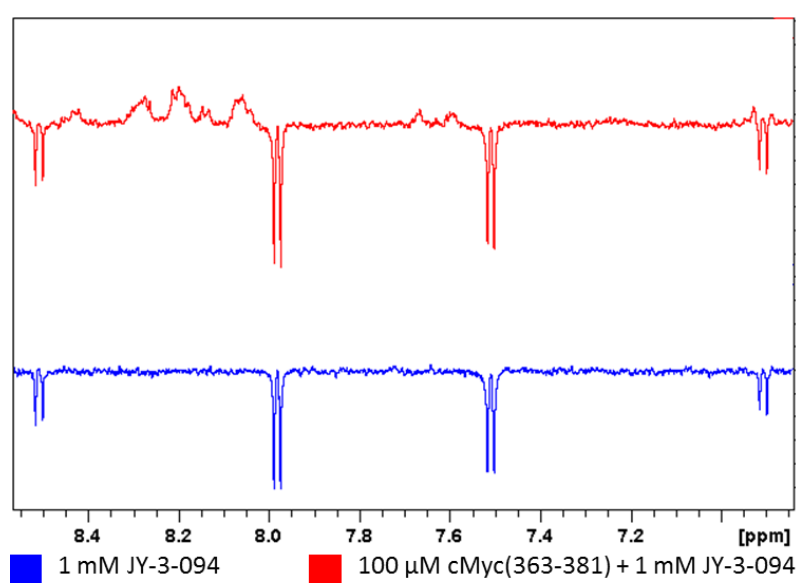
Figure 51. *A comparison of ^1H wLOGSY JY-3-094 spectra in the absence and presence of cMyc(363-381)*

Due to observed CSPs for cMyc(363-381), it was hypothesized that JY-3-094 may affect the diffusion rate of the peptide by potentially making it to fold into a more compact structure. The diffusion rate was measured by using methyl region of the peptide (0.8 – 1 ppm) and was not perturbed in the presence of the ligand (**Table 3**). The lack of diffusion rate changes suggested that the peptide did not undergo significant conformational changes upon interacting with JY-3-094.

Table 3. *The calculated diffusion rates for cMyc(363-381) in the presence and absence of JY-3-094.*

Mixture	Diffusion rate x 10^{-11}, m^2/s
100 µM cMyc(363-381)	1.98
100 µM cMyc(363-381) + 1 mM JY-3-094	2.00
1 mM cMyc(363-381)	1.96
1 mM cMyc(363-381) + 1 mM JY-3-094	1.96

3.5 Conclusions

The production of cMyc peptides as soluble fusion protein was shown to be prone to proteolysis. The proteolysis problem was solved by expressing the peptides as fusion proteins in inclusion bodies. However, the non-efficient CNBr chemical cleavage prevented the release of cMyc peptides from the fusion protein, potentially due to impurities of the sample.

The cMyc literature-reported ligands 10074-G5 and 10058-F4 exhibited non-satisfactory solubility profiles (< 100 μM), preventing the assessment of IDP:ligand interactions with NMR spectroscopy. The observed deviation of solubility from literature values suggested that the original literature-reported data may have been obtained using some specific, non-explicitly stated conditions.

JY-3-094, a more soluble derivative of 10074-G5, has been confirmed to interact with aromatic Phe amino acids of non-labelled cMyc(363-381), as described previously in the literature. Further studies have shown these interactions to potentially be charge-based and exhibit low mM affinity. The 1D ligand-observed NMR methods, however, did not indicate any cMyc(363-381):JY-3-094 interactions. It would be of interest to produce isotopically labelled cMyc peptides to better understand the cMyc(363-381) interaction mechanism with JY-3-094.

4 Protein tyrosine phosphatase 1B (PTP1B)

4.1 Introduction

The non-communicable diseases account for 41 million deaths every year, according to WHO. Of these, 9 million are caused by various types of cancer [157], [158]. Out of the reported oncogenic types, the most common ones include lung, breast, colorectal and pancreatic malignancies. In addition to this, 1.6 million deaths per year are caused directly or indirectly by diabetes. A direct connection between diabetes and obesity has also been established [159].

WHO reports that almost 2 billion adults in the world were overweight in 2016, out of which 13% were obese. Obesity directly affects the quality of life, the well-being of a person and produce a strain on the healthcare system. It is also has a direct relation to type 2 diabetes which affected over 400 million people in 2014. It has been related to such outcomes as blindness, kidney failure, heart attack and strokes [160], [161]. While type 1 diabetes is treated by insulin injections, type 2 diabetes is usually tackled with low molecular weight drugs, such as insulin sensitizers and insulin secretagogues [162].

The role of phosphate tyrosine kinases (PTKs) and phosphate tyrosine phosphatases (PTPs) in cell signalling and regulatory functions has been investigated and is known reasonably well. While PTKs are responsible for phosphorylating target proteins, PTPs dephosphorylate them. These post-translational modifications (PTMs) may cause activation and/or deactivation of various cellular signalling pathway cascades [163], [164]. It was also shown that by inhibiting protein kinases the amplitude of the cellular signal is affected more than the duration, whereas phosphatase inhibitors influenced the amplitude and duration of the signal similarly [165]. While PTKs have been established as potential oncogenic drug targets, PTPs were considered to be tumour growth suppressors, mainly due to their dephosphorylating capabilities. However, a mutated Src homology 2 (SH2)-containing protein-tyrosine phosphatase-2 (SHP2) was identified as an oncoprotein which promoted development of acute myelogenous leukaemia [166]. This example showed that PTPs should also be considered as potential therapeutic drug targets.

One of a few PTPs of particular interest is protein-tyrosine phosphatase non-receptor type 1 (PTPN1), or protein tyrosine phosphatase 1B (PTP1B). The enzyme contains an N-terminal

structured catalytic domain, a regulatory intrinsically disordered C-terminal domain, and a hydrophobic intrinsically disordered C-terminal domain (**Figure 52**).

Full length PTP1B (1-435)

Folded catalytic domain (1-301)

Hydrophobic C-terminal domain (400-435)

Regulatory intrinsically disordered domain (301-400)
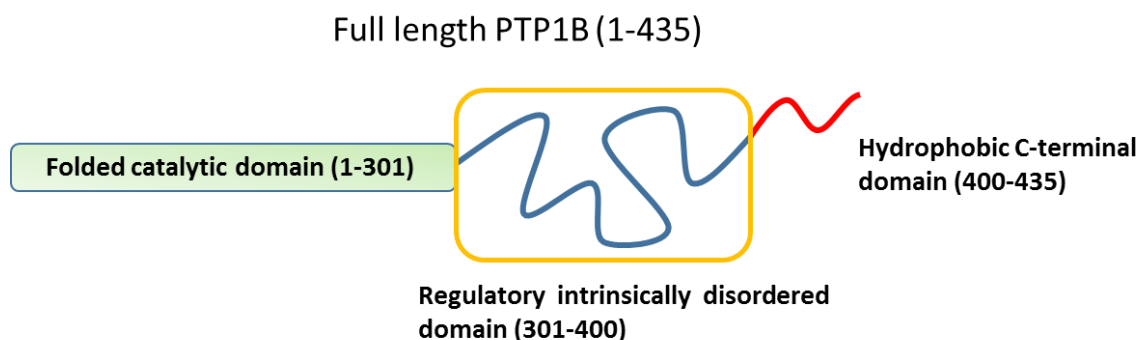
Figure 52. A structural representation of full length PTP1B.

PTP1B was shown to be involved in diabetes and obesity development in mice models, where the deletion of PTP1B-coding gene kept the mice healthy with an increased sensitivity to insulin even under a high-fat diet that would eventually lead to obesity [167]. PTP1B is also considered to be involved in mammary tumorigenesis, where it is overexpressed together with PTK HER2 and promotes tumorigenesis [168]. The direct relation between tumorigenesis and increased levels of PTP1B shows that PTPs can promote downstream signalling, which earlier was thought to be done only by kinases [169].

Numerous PTP1B inhibitors have been developed that target the structured domain of the protein [170]–[172]. Due to PTP's ability to hydrolyse phosphorylated Tyr (pTyr) residues, first inhibitors were designed to mimic these interactions. Multiple pTyr mimics were designed to contain a non-hydrolysable phosphonodifluoromethyl phenylalanine moiety. However, due to a highly charged nature, cell permeability of the compounds were a recurring problem. Approaches to tackle this problem included the usage of fatty acid chains or nitrofuryl moieties in order to shield the charge. Alternative pTyr mimetics have also been tested, such as carboxylic and sulfonic acids, but with limited success, mainly due to either selectivity or cell permeability problems [172]. Less negative imide derivatives exhibited increased cell permeability, but showed low *in vivo* efficacy. The development of non-charged molecules showed increased cell permeability and high *in vivo* efficacy, but with unclear inhibition targets. Multiple natural compounds and their derivatives were identified to exhibit good selectivity, bioavailability and *in vivo* efficacy in test models due to their unique chemical

scaffolds and may be developed further into more appealing drugs in terms of activity and permeability.

Despite multiple approaches towards producing specific, high affinity, low toxicity PTP1B inhibitors, only a few molecules have made it to the clinical trials, including ertiprotafib (**Figure 53**). However, ertiprotafib failed in phase 2 due to undesirable side effects and off-target interactions, including against IκB kinase β [173].

Several allosteric binders have also been identified using experimental and computational modelling tools [174], [175]. In the case of J64512, the compound binds to the ordered region of PTP1B that is ~20 Å away from the catalytic site and locks the catalytically essential WDP loop, as shown by X-ray and NMR data [176]. As an alternative approach, antisense oligonucleotides (ASOs) were used to suppress the expression of PTP1B. ISIS-113715 exhibited positive *in vivo* anti-diabetic effects in animal models and clinical trials were initiated in 2011, followed by a higher potency ASO ISIS-PTP1BRx with positive results in 2015 (**Figure 53**) [177].

The IDR at the C-terminus of PTP1B has not been targeted mainly due to its disordered nature. This presented difficulties when using typical biophysical approaches, such as X-ray crystallography, which was necessary to identify any potential mechanisms of action with a binder. It was suggested that the IDR could be involved in autoregulatory functions by interacting with the catalytic domain, as demonstrated in TCPTP case (~75% similarity of catalytic domains) [178]. Interestingly, it was identified that the truncation of PTP1B releases the protein to the cytosol from the endoplasmic reticulum membrane in human platelets and increases PTP1B's activity two-fold, thus confirming the importance of the C-terminal IDR [179], [180]. Currently, only a few compounds are known to bind to the C-terminal IDR of PTP1B which inhibit its enzymatic activity via distinct conformational changes (**Figure 53**). MSI-1436, or trodusquemine, is a potent inhibitor that targets the C-terminal region of PTP1B [105], [181]. MSI-1436 was shown to selectively bind to PTP1B with 2:1 stoichiometry and noncompetitively inhibit its activity by causing distinct conformational changes, as determined using FRET, CD, ITC and 2D NMR techniques. The proposed binding model suggests that the binding event involves α7 helix (285-298), along with residues 299, 310, 311 and the newly identified α9 helix (360–377) in the IDR. Moreover, the IDR of PTP1B exhibited reduced hydrodynamic radius in the presence of MSI-1436, as determined by DLR, suggesting that the C-terminal region of PTP1B obtained a more compact structure. Hydrodynamic radius

R_H may be an important parameter when exploring LMW interactions with IDPs if a formation of more compact structures is anticipated. However, no data from other drug discovery techniques are available for MSI-1436, including ligand-observed NMR, or SPR.

An MSI-1436 analogue DPM-1001 was developed later which exhibited higher enzymatic inhibitory levels towards PTP1B in the presence of Cu(II) ions (**Figure 53**). Several metal ions have previously been reported to inhibit PTP1B activity by producing reactive oxygen species [182]. However, the reported interaction mechanism of DPM-1001 exploited the sterol properties of the molecule and coordination of Cu(II) ions via the charged tail of DPM-1001 and His residues in the IDR of PTP1B [183]. In 2015, another analogue to MSI-1436 named claramine with a faster synthesis route and similar enzymatic inhibitory levels *in vitro* was developed [184]. However, no structural data was provided which would help to understand the exact binding mechanism. The previous example of MSI-1436 showed that direct observation of IDR:ligand interactions are possible and potential binders may be identified during screening campaigns.
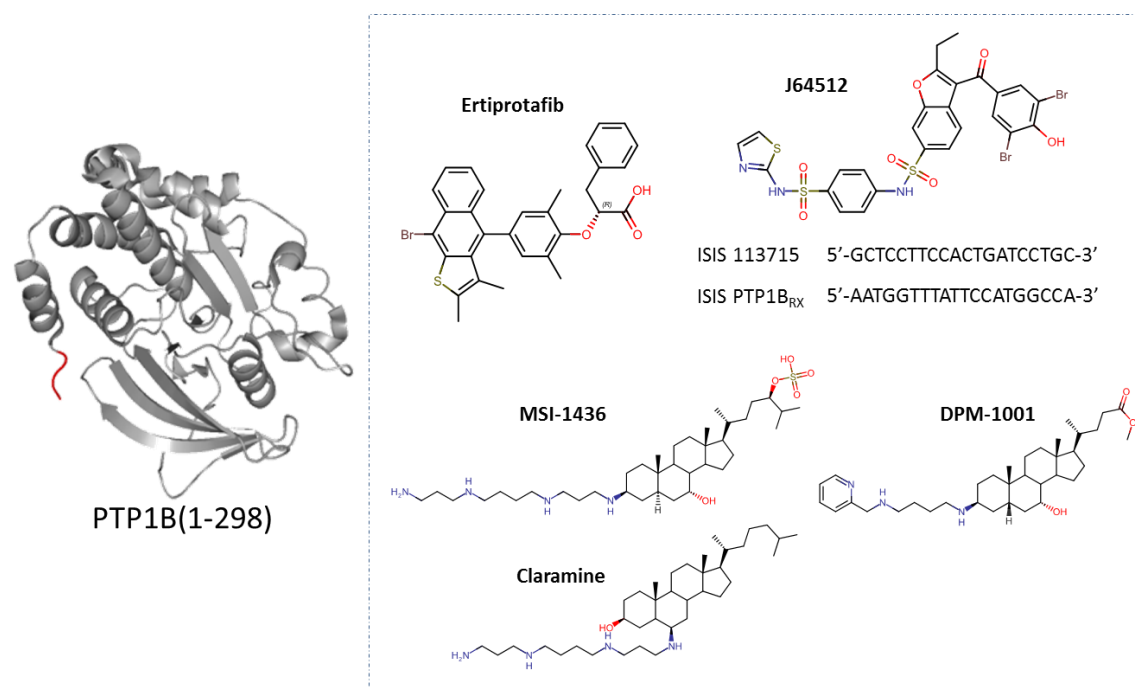
Figure 53. A structural representation of PTP1B(1-298) structure (PDB: 1SUG) with red marking indicating the start point of C-terminal IDR (Left). Chemical structures of PTP1B inhibitors (Right).

In this study, the folded domain only of PTP1B(1-301) and the folded domain with IDR (1-393) were successfully produced. This was followed by an exploratory study that investigated literature-reported ligands that interacted with the IDR of PTP1B using NMR, SPR and MST

techniques. Multiple fragment screening campaigns were also performed using NMR and SPR with a diverse set of fragment libraries in order to learn more about IDP:fragment interaction patterns with a specific focus on the C-terminal IDR of PTP1B.

4.2 Protein production and characterization

In order to investigate literature-reported ligands and perform fragment screening, multiple versions of PTP1B protein had to be produced in sufficient yields.

4.2.1 PTP1B(1-301)

The plasmid coding for the folded catalytic domain of human PTP1B(1-301) with N-terminal 6xHis-tag was kindly provided by N. Tonks (USA) in a T7-driven pET-28b expression vector. The protein was expressed in *E. coli* BL21(DE3) RIL CodonPlus cells with yields of ~18 mg per 1 L of LB media (**Figure 54**) (**§7.3.1, §7.3.3**). In addition to this, the activity assay of 6xHis-PTP1B(1-301) has shown that the enzyme is active and can be used for subsequent biophysical assays (**Figure 55**) (**§7.3.10**).
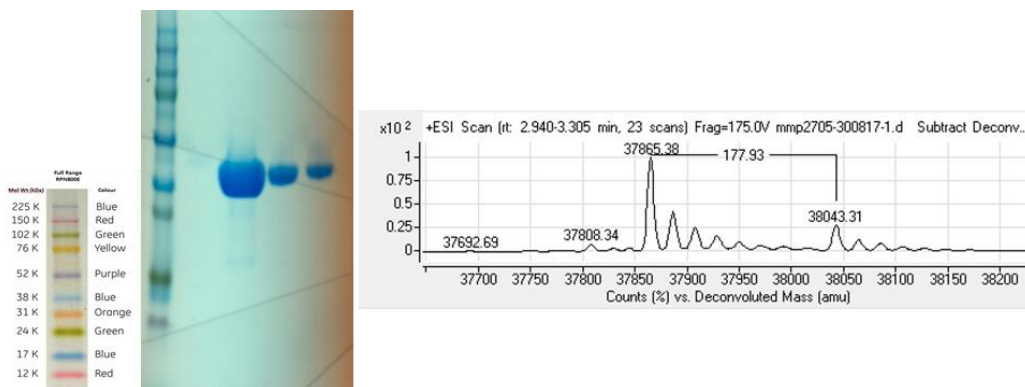
Figure 54. The QC of pure PTP1B(1-301) by SDS-PAGE and HPLC-MS.
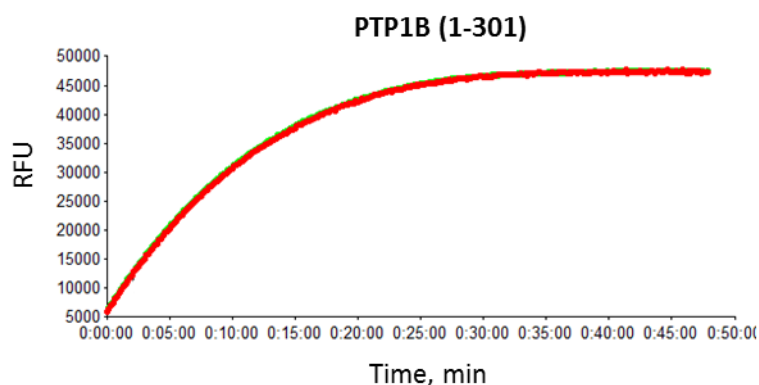
Figure 55. Time-resolved enzymatic activity of 6xHis-PTP1B(1-301) using DiFMUP as substrate. Red and green colors mark replicates.

After expressing the folded domain of PTP1B, the expression of PTP1B(1-393) with C-terminal IDR was attempted.

4.2.2 PTP1B(1-393)

The expression of 6xHis-PTP1B(1-393) in pET-28b vector in *E. coli* BL21(DE3) pLysS produced a truncated version of PTP1B with around 30-40 amino acid cleavage at the disordered C-terminus region which was confirmed by intact MS-MS and peptide mass fingerprint (PMF) studies (**Figure 56, Figure 57**) (**§7.3.1, §7.3.3, §7.3.9**).


```
        10         20         30         40         50         60
GHMASMEMEK EFEQIDKSGS WAAIYQDIRH EASDFPCRVA KLPKNKNRNR YRDVSPFDHS

        70         80         90        100        110        120
RIKLHQEDND YINASLIKME EAQRSYILTQ GPLPNTCGHF WEMVWEQKSR GVVMLNRVME

       130        140        150        160        170        180
KGSLKCAQYW PQKEEKEMIF EDTNLKLTLI SEDIKSYYTV RQLELENLTT QETREILHFH

       190        200        210        220        230        240
YTTWPDFGVP ESPASFLNFL FKVRESGSLS PEHGPVVVHC SAGIGRSGTF CLADTCLLLM

       250        260        270        280        290        300
DKRKDPSSVD IKKVLLEMRK FRMGLIQTAD QLRFSYLAVI EGAKFIMGDS SVQDQWKELS

       310        320        330        340        350        360
HEDLEPPPEH IPPPPRPPKR ILEPHNGKCR EFFPNHQWVK EETQEDKDCP IKEEKGSPLN

       370        380        390
AAPYGIESMS QDTEVRSRVV GGSLRGAQAA SPAKGEPS
```

charge	m/z	mass (Da)	
3	14077.0	42228.0	
4	10557.9	42227.6	
	avge	**42227.8**	Da
	stdev	**0.3**	

MW exp	MW calc	Dev.(Da)	Range
42227.800:	42227.696	0.104	1-364

Figure 56. MS-MS data of 6xHis-PTP1B(1-393). Yellow marking shows truncated part as deduced from MW.

```
  1  MEMEKEFEQI  DKSGSWAAIY  QDIRHEASDF  PCRVAKLPKN  KNRNRYRDVS
 51  PFDHSRIKLH  QEDNDYINAS  LIKMEEAQRS  YILTQGPLPN  TCGHFWEMVW
101  EQKSRGVVML  NRVMEKGSLK  CAQYWPQKEE  KEMIFEDTNL  KLTLISEDIK
151  SYYTVRQLEL  ENLTTQETRE  ILHFHYTTWP  DFGVPESPAS  FLNFLFKVRE
201  SGSLSPEHGP  VVVHCSAGIG  RSGTFCLADT  CLLLMDKRKD  PSSVDIKKVL
251  LEMRKFRMGL  IQTADQLRFS  YLAVIEGAKF  IMGDSSVQDQ  WKELSHEDLE
301  PPPEHIPPPP  RPPKRILEPH  NGKCREFFPN  HQWVKEETQE  DKDCPIKEEK
351  GSPLNAAPYG  IESMSQDTEV  RSRVVGGSLR  GAQAASPAKG  EPSLPEKDED
```

Figure 57. **The** *PMF results for 6xHis-PTP1B(1-393). Red marked letters represent the identified peptide sequences that were compared against a full length PTP1B(1-435).*

Globular proteins are generally more stable and protease-resistant during expression, whereas disordered proteins are more prone to proteolytic degradation due to their unfolded nature [185]. Another possible reason for truncated versions of proteins could be translational arrest during transcription and/or translation due to secondary RNA structure formation [186]. However, this phenomenon was not expected due to commercial codon and plasmid optimization procedures which should prevent these types of occurrences.

For confirmatory purposes, ten discrete *E. coli* BL21(DE3) pLysS colonies coding for 6xHis-PTP1B(1-393) in pET28b vector were picked after transformation and tested for expression of full length 6xHis-PTP1B(1-393) protein. The results concluded that the truncated protein was

consistently produced, as determined by Maxwell 16 Polyhistidine Protein Purification Kit (Promega) and SDS-PAGE (**Figure 58**) (**§7.3.4**).
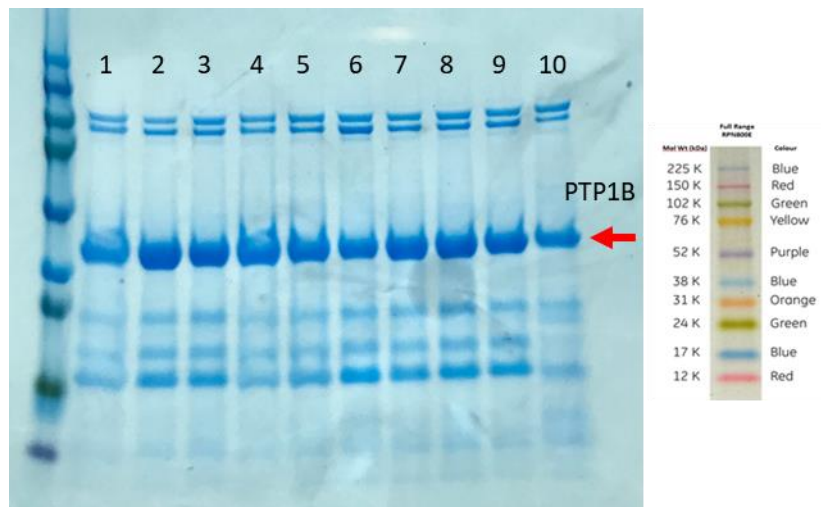
Figure 58. The evaluation of expression levels of truncated 6xHis-PTP1B(1-393) protein from different colonies.

It was reported that the 6xHis tag on the C-terminal side of PTP1B(1-393) has helped to prevent proteolytic degradation of the protein and obtain full length protein (N. Tonks, personal communication). Therefore, it was decided to change the location of 6xHis-tag from N-terminus to C-terminus.

The codon-optimized PTP1B(1-393)-6xHis construct in pJ411 expression vector was tested for expression in the *E. coli* BL21(DE3) pLysS cell line. This approach had only a partial success due to significant variations of expression levels between the discrete transformed colonies (**Figure 59**). The colony that was identified to produce full length PTP1B(1-393)-6xHis was used to produce target protein, yielding 3 mg per 1 L of LB media (**Figure 60**). All of the follow-up attempts to produce PTP1B(1-393)-6xHis were not successful due to significantly lower yields. It is possible that PTP1B(1-393)-6xHis exhibited toxicity towards *E. coli* which prevented reproducible expression of the protein. While 6xHis-PTP1B(1-301) was expressed in high yields previously, the variable and low expression levels of PTP1B(1-393)-6xHis were potentially caused by the IDR (**§4.2.1**).
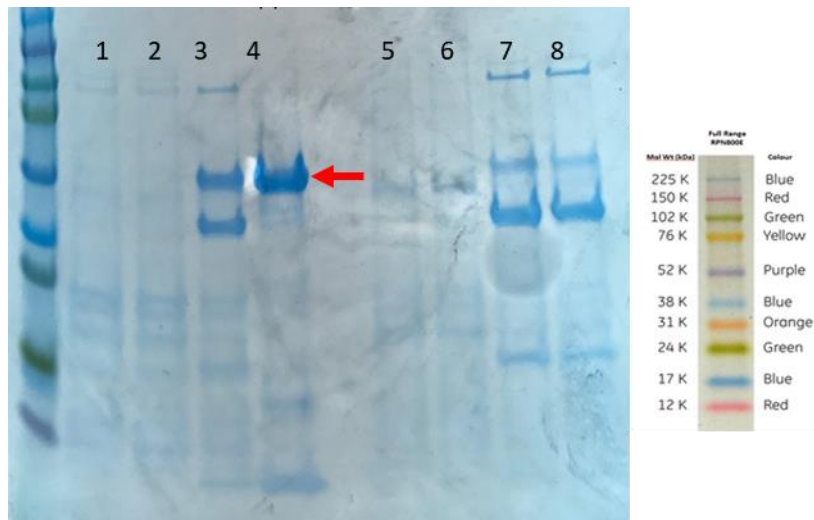
Figure 59. *The variability of PTP1B(1-393)-6xHis construct expression levels between different colonies with pJ411 expression vector in E. coli BL21(DE3) pLysS. Red marked colony produced full length PTP1B(1-393)-6xHis.*
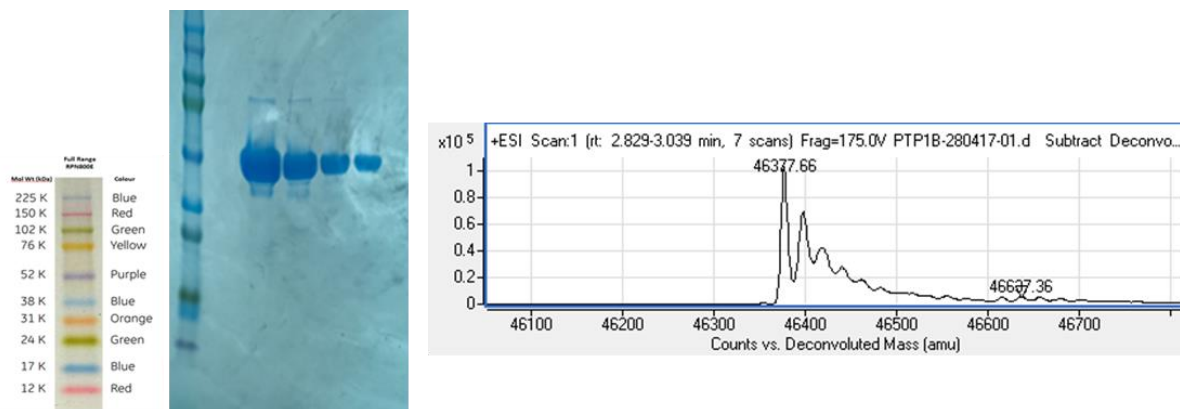
Figure 60. *The QC of PTP1B(1-393)-6xHis by SDS-PAGE and HPLC-MS.*

After additional literature research, it was decided to test various N-terminal tags in order to stabilize the expression of PTP1B(1-393)-6xHis [187]. While various tags are usually used to enhance the solubility and folding of expressed proteins, it was assumed that the tags may stabilize the expression of PTP1B(1-393)-6xHis. Multiple PTP1B(1-393)-6xHis constructs were produced using standard cloning methods (**§7.2.2**) and included the following cleavable N-terminal tags:

- SUMO-TEV-PTP1B(1-393)-6xHis in T7-driven pJ411 vector
- MBP-TEV-PTP1B(1-393)-6xHis in T7-driven pJ411 vector
- Trx-TEV-PTP1B(1-393)-6xHis in T7-driven pJ411 vector
- GST-TEV-PTP1B(1-393)-6xHis in T7-driven pJ411 vector
- GST-3C-PTP1B(1-393)-6xHis in Ptac-driven pGEX-6P-1 vector

It was noticed that *E. coli* BL21(DE3) pLysS cell line did not produce as much GST-tagged protein with pJ411 expression vectors as it did with pGEX vector (**Figure 61**). This observation indicated that the commercially available T7-driven pJ411 vector was not a good expression vector for PTP1B(1-393)-6xHis. The pGEX vectors operate with Ptac promoter which has been determined to be not as strong as T7 promoter in terms of total protein produced during expression [188]. The structure of the plasmid may also have had strong influence on the expression levels.
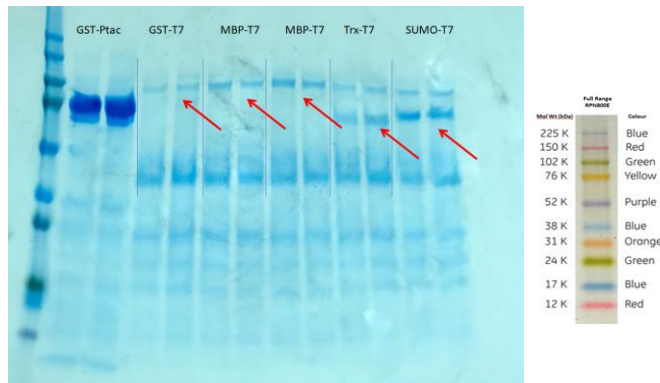
Figure 61. Test expression of PTP1B(1-393)-6xHis with different N-terminal tags.

Therefore, the protein-coding insert of GST-3C-PTP1B(1-393)-6xHis was subcloned into a T7-driven pET expression system pET-24a to further increase the expression levels. Moreover, it was also decided to test *E. coli* BL21(DE3) cell line without *pLysS* gene in order to potentially increase expression levels even more. Interestingly, it was observed that *E. coli* (BL21)DE3 produced more of GST-PTP1B(1-393)-6xHis than BL21(DE3) pLysS, possibly due to tighter T7 promoter control in the pLysS strain (**Figure 62**). The pLysS genotype-containing *E. coli* systems produce T7 lysozyme and are used to provide stricter T7 promoter control in order to prevent 'leaky expression' and potential toxicity to *E. coli* before induction [189].
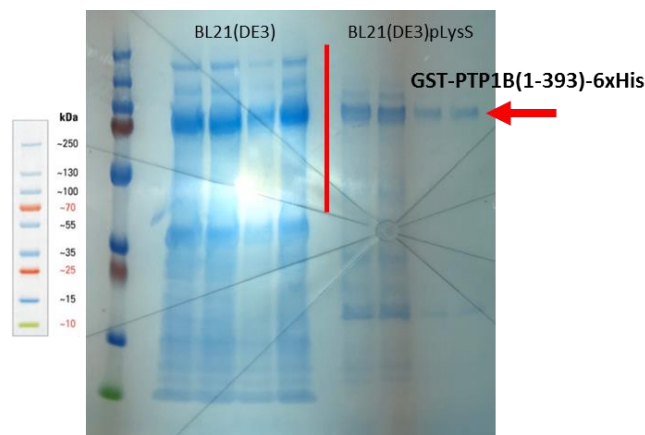
Figure 62. The evaluation of test expression of GST-3C-PTP1B(1-393)-6xHis (pET-24a) construct in different E. coli strains.

The expression of GST-3C-PTP1B(1-393)-6xHis (pET-24a) in *E. coli* BL21(DE3) yielded about 5 mg of full length PTP1B(1-393)-6xHis per 1 L of TB media, as determined by A_{280} (**§7.3.6**). The proper folding of the enzyme has also been confirmed with an enzymatic activity assay (**Figure 63**), (**§7.3.10**).
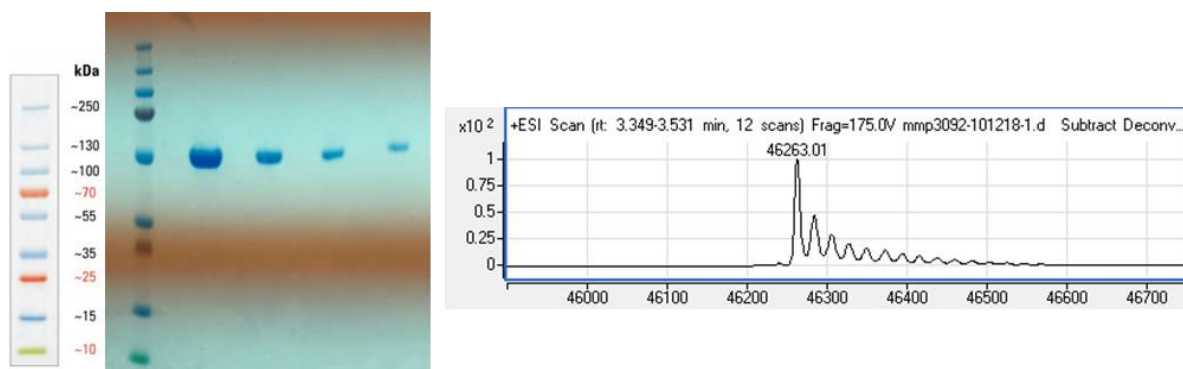
Figure 63. The QC of pure PTP1B(1-393)-6xHis by SDS-PAGE and HPLC-MS.
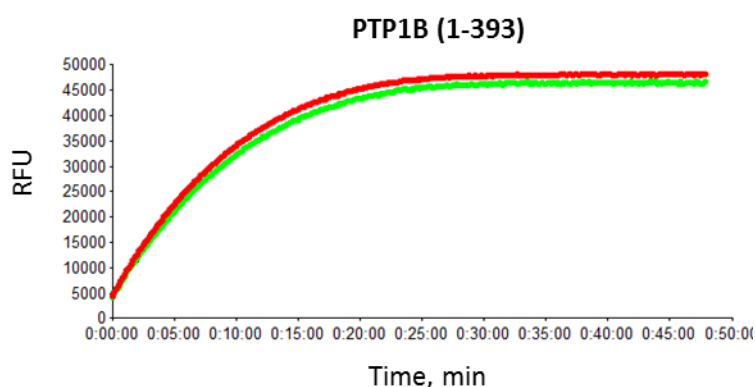
Figure 64. Time-resolved enzymatic activity of PTP1B(1-393)-6xHis. Red and green colors mark replicates.

Once 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis proteins were produced, literature-reported ligands were obtained and characterized using NMR.

4.3 Characterization of the ligands

As mentioned previously, several compounds have been reported to interact with the IDR of PTP1B, including MSI-1436 (**Figure 65**) (**§4.1**). While MSI-1436 interactions with the PTP1B(1-393) have been already characterized in-depth, it was still of interest to investigate the interactions further using 1D ligand-observed NMR methods, including affinity determination by NMR, because such data has not been reported yet.
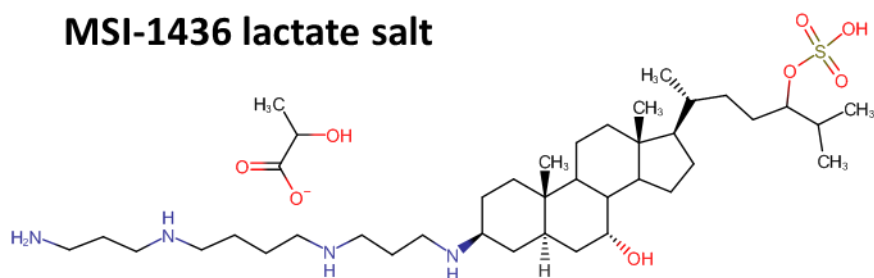
Figure 65. *The chemical structure of the investigated MSI-1436.*

The structural integrity of the acquired MSI-1436 was assessed by 1D ^{1}H NMR in d$_6$-DMSO (**§7.4.2**). The ligand was determined to be soluble up to 800 μM in 50 mM HEPES pH 6.9, 150 mM NaCl, 4% d$_6$-DMSO, 5% D$_2$O using 100 μM DSS at 0.0 ppm as a reference. However, it was noticed that MSI-1436 may form aggregates above 200 μM (**Figure 66**). The methyl peaks were noticed to broaden out instead of increasing in intensity during titration experiments which was consistent with the formation of soluble aggregates.
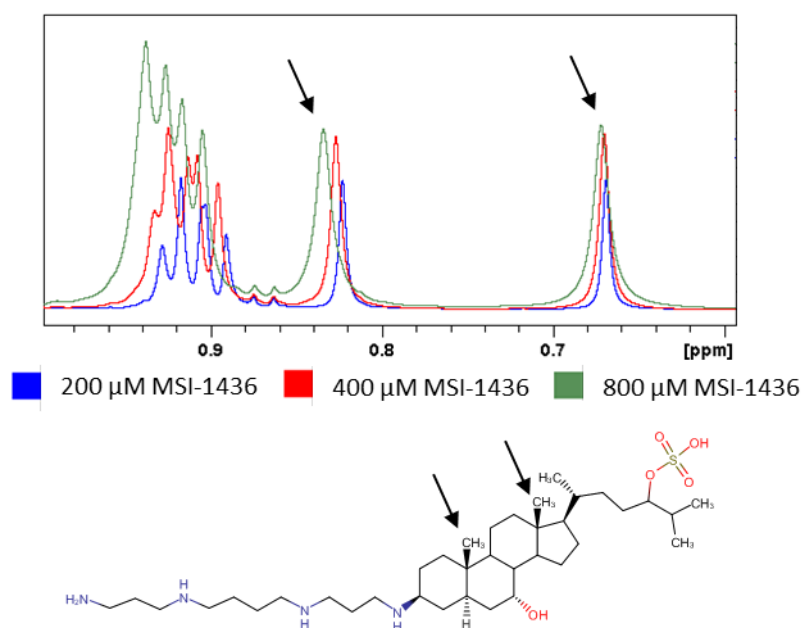
Figure 66. ^{1}H *NMR spectra of MSI-1436 at different concentrations. The black arrows mark methyl peaks in the structure.*

The structure of MSI-1436 is similar to a detergent molecule with a polar head and a non-polar tail. It is possible that at a certain concentration the compound is forming micelles with marked methyls getting buried in hydrophobic environment (**Figure 66**).

The compound was further investigated using 1D NMR methods in the presence of PTP1B(1-393)-6xHis.

76

4.4 Detection of ligand binding by NMR

The ^{1}H STD, wLOGSY and CPMG-filtered NMR techniques are typically used to identify potential binders to the target protein. Due to MSI-1436 not having any aromatic moieties, ^{1}H STD experiments were not performed due to possible excitation of aliphatic chains of the ligand during the experiment. Furthermore, the acquired wLOGSY data did not indicate any binding of the ligand due to severe overlay of protein- and ligand-specific signals. In addition to this, wLOGSY data suggested that MSI-1436 may form aggregates as the positive NOE signals were observed without any protein present in solution (**Figure 67**). The possibility of aggregates was also confirmed by ^{1}H CPMG-filtered experiments which showed the increased T_2 relaxation of MSI-1436 at higher ligand concentrations (**Figure 68**).
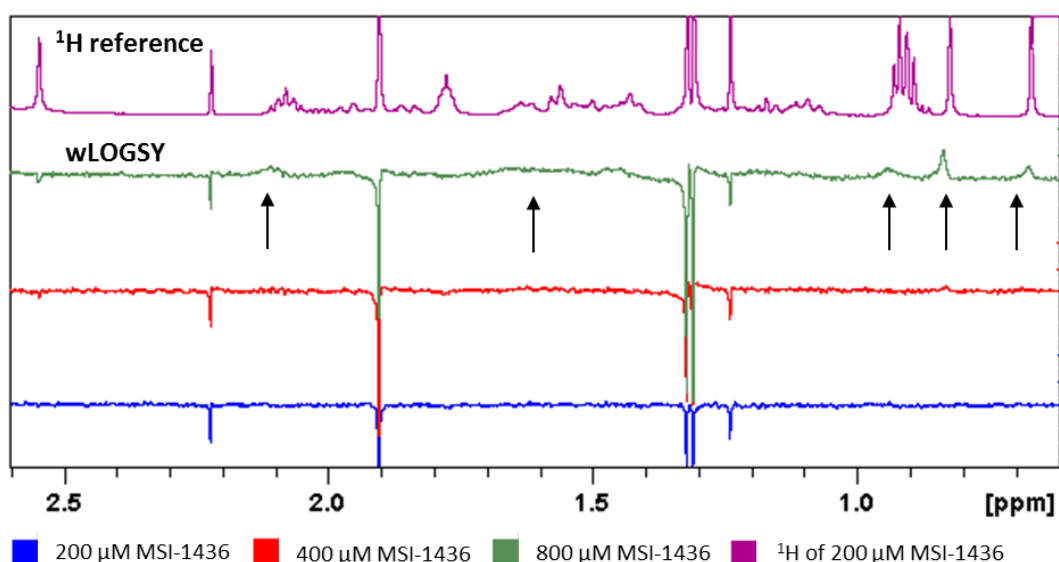
Figure 67. A comparison of wLOGSY spectra of MSI-1436 at different concentrations. The black arrows mark compound specific peaks.
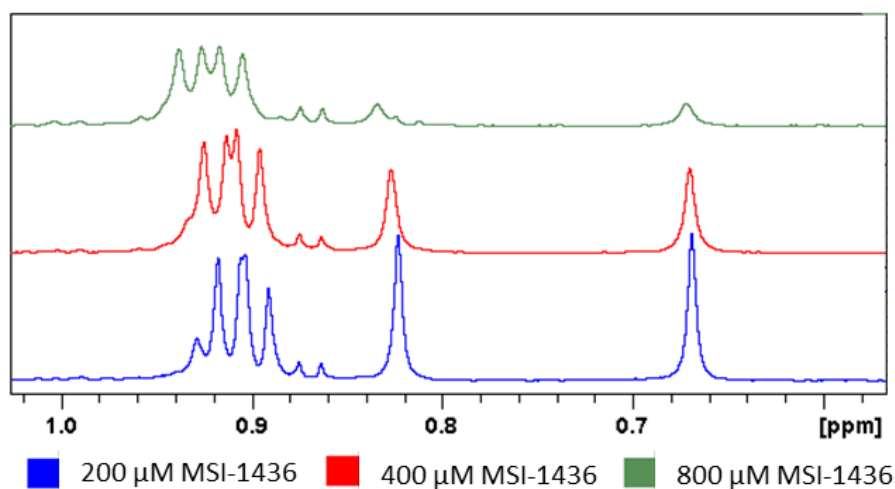
Figure 68. A comparison of ^{1}H CPMG-filtered spectra of MSI-1436 at different concentrations. τ = 200 ms

The previously mapped CSPs for IDR of PTP1B(1-393) were obtained in the presence of 450 µM of MSI-1436 [105]. It is possible that the reported CSPs represented interactions between the protein and the aggregated MSI-1436 rather than a monomeric form of the ligand.

It has been also noted that the methyl peaks of MSI-1436 in ^{1}H spectra broaden significantly in the presence of PTP1B(1-393)-6xHis which was consistent with protein:ligand interactions (**Figure 69**). In addition to this, the ^{1}H CPMG-filtered spectra also indicated broadening after 200 ms delay to some ligand-specific peaks which further confirmed the previously reported MSI-1436 interactions with PTP1B(1-393)-6xHis.
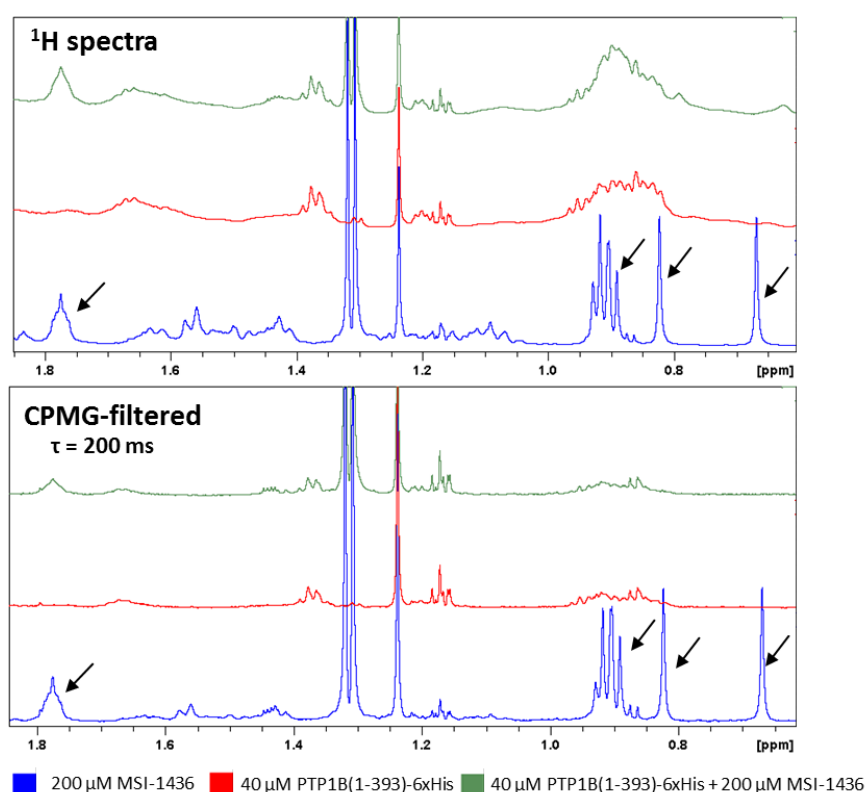
Figure 69. A comparison of ^{1}H and CPMG-filtered spectra for MSI-1436 in the presence and absence of PTP1B(1-393)-6xHis. The black arrows indicate MSI-1436 specific resonance peaks.

Usually, the indole NH signals of Trp are broadened out if they are buried in the hydrophobic core of the protein. However, the indole NH of W333 was observable in ^{1}H spectra due to it being in the solvent-exposed IDR of PTP1B(1-393)-6xHis. After addition of MSI-1436, slight CSPs were observed for W333 (**Figure 70**). The signal broadening may be explained by a slight ordering or folding of the IDR, causing W333 signal to decrease in intensity. This shows that Trp amino acid may be used to observe some conformational changes for IDRs in general.
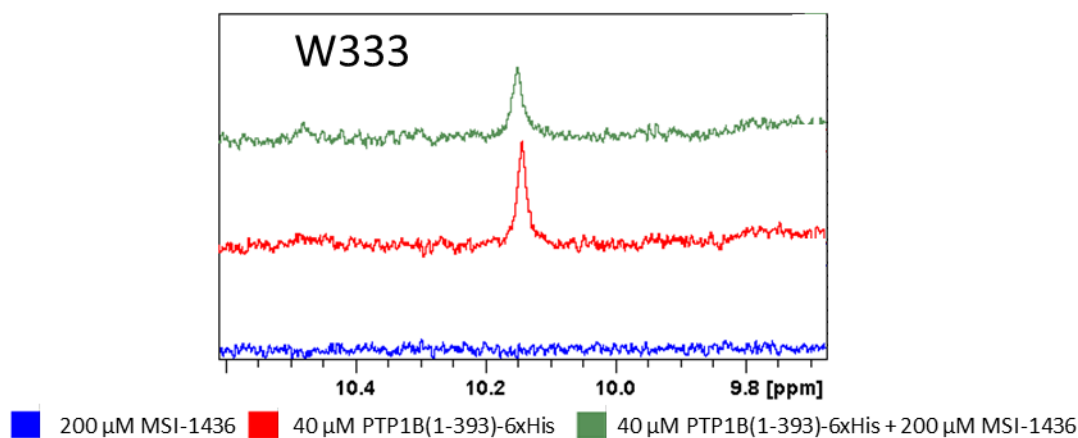
Figure 70. *A comparison of ¹H spectra of W333 of PTP1B(1-393)-6xHis in the presence and absence of MSI-1436.*

The previously reported K_D values for PTP1B:MSI-1436 interactions, ranging from 0.6 to 4 µM, agreed with the experimentally determined K_D value of 3.6 +/- 0.8 µM via ¹H NMR (**Figure 71**) [105]. ¹H NMR spectra was used to track methyl peak of MSI-1436 in the presence of 20 µM PTP1B(1-393)-6xHis and the resulting ΔCSP values were fitted to 1:1 binding model. It is also possible that the observed CSPs may be occurring due to micelle formations at increasing concentrations. However, previous experimental NMR DOSY data suggested that no aggregates of MSI-1436 are forming up to 10 µM [190].
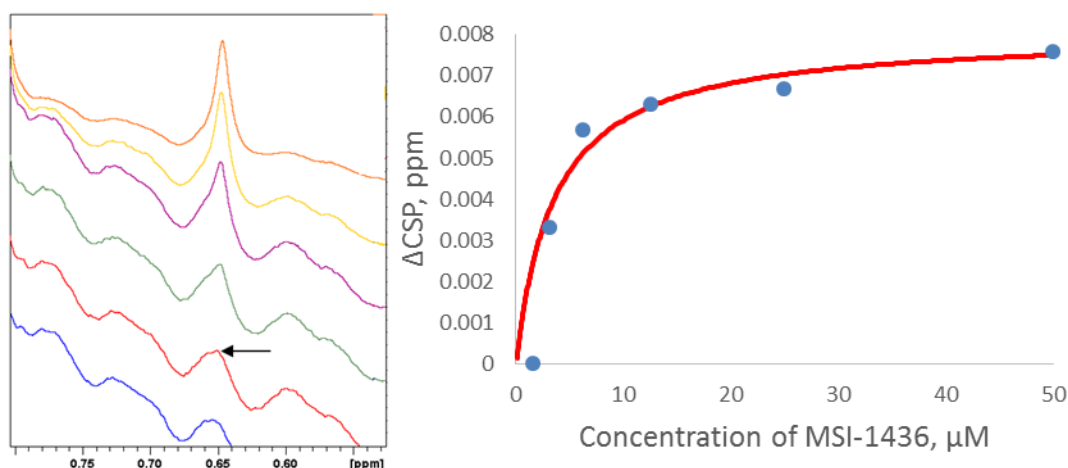
Figure 71. *The evaluation of MSI-1436-specific peak in the presence of 20 µM of PTP1B(1-393)-6xHis at different MSI-1436 concentrations. The fitting of ΔCSPs for the marked methyl peak of MSI-1436 using 1:1 binding model is also shown.*

It was then decided to investigate the reported folding phenomenon of the IDR of PTP1B in the presence of MSI-1436 using NMR DOSY. Methyl peaks in the aliphatic region (0.75 – 1 ppm) of PTP1B proteins were used to calculate the diffusion rate. The data has shown that hydration radii R_H for PTP1B(1-301) and PTP1B(1-393) were similar to previously reported values obtained via dynamic light scattering (DLS) technique (1-301: 24.1 Å and 1-393: 32.8 Å)

(**Table 4**) [105]. The deviations can potentially be explained by slightly different experimental conditions.

Table 4. *The calculated diffusion rate and R_H values for PTP1B constructs as determined by NMR DOSY.*

Protein	Diffusion rate, x 10^{-11} m^2/s	R_H, Å	R_H, Å [105]
6xHis-PTP1B(1-301)	8.2 +/- 0.31	27	24.1
PTP1B(1-393)-6xHis	6.44 +/- 0.12	35	32.8

After addition of 200 µM of MSI-1436, the diffusion rates of protein could not be determined unambiguously due to overlapping aliphatic peaks of protein and ligand using standard one-component fitting approach. The two-component fitting option, however, suggested no significant change in the diffusion rate (**Table 4, Table 5**). The two-component fitting option takes into account that there are two overlaying peaks on top of each other from two different species.

Table 5. *The calculated diffusion rate values for PTP1B(1-393)-6xHis as determined by NMR DOSY.*

Protein	Diffusion rate, x 10^{-11} m^2/s	
	One-component fitting	Two-component fitting
PTP1B(1-393)-6xHis	7.17 +/- 0.45	6.3 +/- 0.38

It has been shown previously that the calculated molecular mass of PTP1B(1-393) increased in the presence of MSI-1436 via SEC experiment [105]. NMR DOSY allows determination of diffusion rates using specific peaks rather than the protein as a whole. This may make any partial conformational changes difficult to detect, especially if the majority of the aliphatic peaks in PTP1B(1-393)-6xHis represent the already folded part of the protein.

Following the NMR experiments, SPR and MST approaches were also used to evaluate the observed PTP1B(1-393)-6xHis:MSI-1436 interactions.

4.5 Detection of ligand binding by SPR and MST

The presence of C-terminal 6xHis tag in PTP1B(1-393) allowed a non-covalent and oriented immobilization of PTP1B(1-393)-6xHis on Ni-NTA chip using Biacore T200 (**§7.4.4**). The multi-cycle studies have revealed that MSI-1436 interacted non-specifically with the reference surface because the observed responses were higher on reference surface than on the protein-immobilized surface. This caused the reference-subtracted curves to give a negative signal (**Figure 72**).
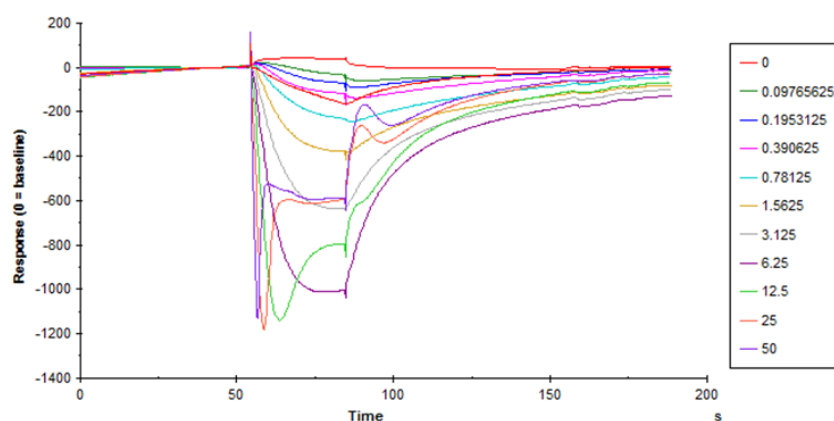
Figure 72. The kinetic dose-response profile of MSI-1436 interacting with immobilized PTP1B(1-393)-6xHis. Immobilization level: ~1000 RU. MSI-1436 concentration range: 0.01 to 50 μM as two-fold dilutions.

Moreover, the observed signals were significantly higher than the expected R_{max} of 15 RU for 1000 RU immobilization level with 1:1 binding model. The interactions of MSI-1436 with the reference surface at pH 7.4 were assumed to be charge-based due to the positively charged polyamine part of MSI-1436 and the negatively charged carboxymethylated dextran surface (**Figure 73**). It is also possible that such large responses are caused by occurring conformational changes of PTP1B(1-393) that SPR technique is also sensitive to (**§1.5.2.3**).
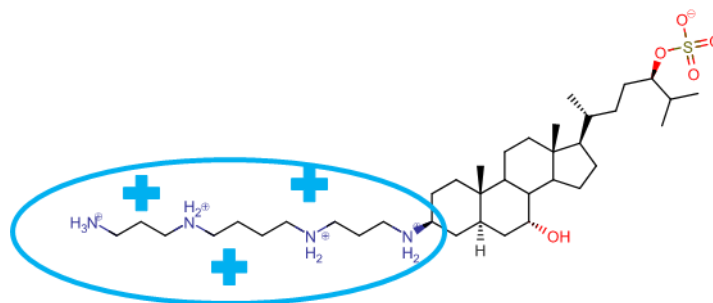
Figure 73. A potential charge distribution in MSI-1436 at pH 7.4.

The charged-based hypothesis was confirmed by injecting MSI-1436 over empty, non-derivatized surface and still observing binding kinetics. The increased salt concentration decreased the non-specific residual binding of the compound to the surface (**Figure 74**). It is possible to see that MSI-1436 has a residual binding at 40 μM in the presence of 150 mM NaCl. In contrast, no residual binding was observed in the presence of 500 mM NaCl.
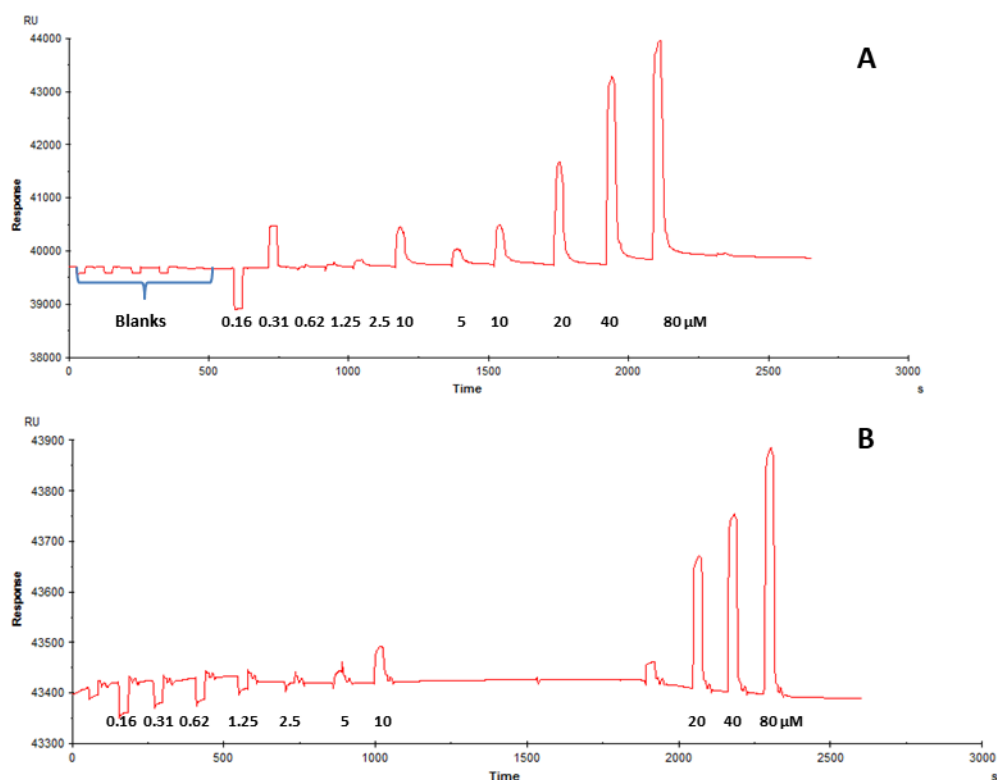
Figure 74. Absolute binding levels to empty surface of Ni-NTA chip in the presence of 150 mM (A) and 500 mM (B) of NaCl at denoted MSI-1436 concentrations.

In order to negate charge-based interactions, PTP1B(1-393)-6xHis may be immobilized using amine coupling, followed by the blocking of the carboxymethyl groups with ethanolamine or positively charged ethylenediamine. However, such immobilization approach may introduce non-oriented conformations and exhibit non-representative responses against MSI-1436.

As an orthogonal approach to NMR and SPR, MST was used to assess potential interactions between PTP1B(1-393)-6xHis and MSI-1436 (**§7.4.3**). The protein was labelled non-covalently via 6xHis tag with RED-Tris-NTA dye which was determined to have a K_D of 2.3 +/- 0.4 nM (**Figure 75**).
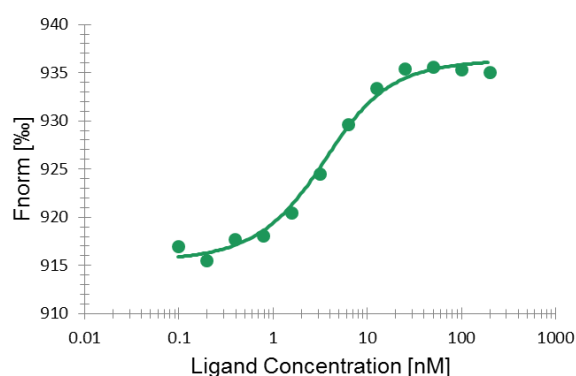
Figure 75. A graphical representation of MST data fit for RED-Tris-NTA dye interaction with PTP1B(1-393)-6xHis.

While further MST data was consistent with MSI-1436 binding to the protein, initial fluorescence data has shown that the initial fluorescence of the dye increased, as the ligand concentration increased (**Figure 76**). Previous NMR data suggested that aggregates may be forming at higher concentrations which may have caused the fluorescent dye to be shielded from the solvent and have its fluorescence profile increased. Due to non-saturating data, the K_D value could not be reliably determined.
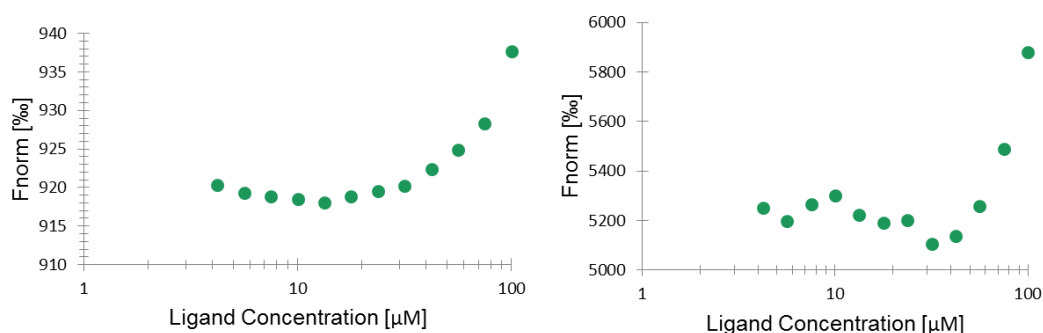
Figure 76. *Graphical representation of MST (**Left**) and initial fluorescence (**Right**) data for labelled protein in the presence of different concentrations of MSI-1436.*

According to the previously determined K_D, another MST experiment was performed with lower concentrations of MSI-1436. The results showed no obvious interactions between the labelled PTP1B(1-393)-6xHis and MSI-1436 (**Figure 77**).
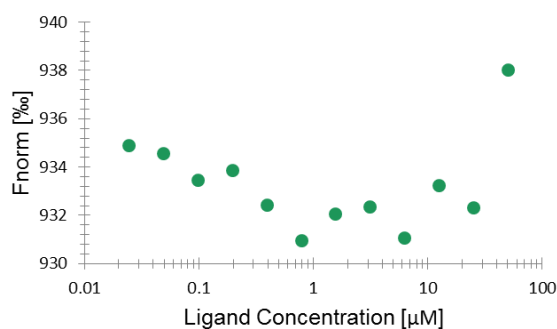
Figure 77. *A graphical representation of MST data for labelled PTP1B(1-393)-6xHis in the presence of different concentrations of MSI-1436.*

In conclusion, NMR, SPR and MST data showed that the knowledge about potential experimental artefacts is important, including aggregate formation or charged-based interactions. The investigation of MSI-1436:PTP1B(1-393) interactions was followed by several fragment screening campaigns in order to identify potential fragments that bind to the IDR of the protein.

4.6   Fragment screening via ^{19}F NMR

^{19}F CPMG-filtered NMR fragment screening campaign was used to identify fragments that interact with the disordered C-terminal region of PTP1B (**§7.4.2, §7.4.2.3**).

In recent years, ^{19}F-labelled molecules have been used to expand the current compound libraries with additional chemical space coverage. It has been shown that the substitutions of hydrogen to fluorine can affect such physicochemical properties as pKa and lipophilicity [191], [192]. In addition to this, fluorine can also participate in protein-ligand interactions which can affect the apparent affinity values [193]. The Carr-Purcell-Meiboom-Gill (CPMG) pulse scheme can be readily applied to be used in ^{19}F CPMG-filtered experiments which allows screening and identification of fluorinated binders in a relatively fast manner [194]. The inherent chemical shift anisotropy (CSA) in the bound state and the chemical shift difference between bound and free state contribute significantly to the observed line width of fluorinated fragments in ^{19}F CPMG-filtered NMR spectra [195]. This increased sensitivity allows using relatively low fragment concentration in order to detect binding events. Moreover, the observed spectral range for ^{19}F NMR is much wider than for ^{1}H NMR which enables the screening of cocktails containing larger number of fragments, thus saving experimental time. This allows a quick evaluation and prioritization of the ^{19}F labelled fragment binders which can be then further validated using orthogonal techniques.

Due to availability and simplicity, CF_3-labelled fragment library of 1664 compounds (Novartis) was used in ^{19}F CPMG-filtered fragment screening. The fluorine signals were observed as sharp singlets which may not have been the case with CF or CF_2 fragments due to possible additional coupling patterns and/or broadening of the signals. The PTP1B active site-directed inhibitor TCS401 was used in all mixtures to reduce potentially high hit rate [196] (**Figure 78**).
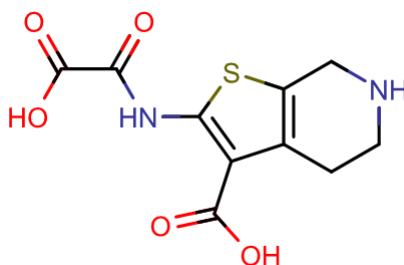
Figure 78. *The chemical structure of PTP1B active site-directed inhibitor TCS401.*

After obtaining reference ^{19}F CPMG-filtered spectra for fragment cocktails, the same experiment was performed in the presence of 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis for comparative purposes. A total of 210 fragments, or 13% of the library, were identified as potential hits against the common folded domain for both 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis. The potential hits exhibited more than 80% decrease in signal intensity after CPMG filter (τ = 320 ms) (**Figure 79**). In addition to this, 16 fragments were identified as potentially specific binders to PTP1B(1-393)-6xHis. It was noted, however, that the signal broadening for PTP1B(1-393)-6xHis-specific ligands was lower than for the folded region-targeting hits (30% vs 80%, respectively) (**Figure 80**). It is possible that the lesser broadening was observed due to weaker binding affinity of the fragments to the IDR region, local environment changes or conformational flexibility of the system. The screening results were followed up with single compound experiments to confirm the potential hits against the IDR of PTP1B(1-393)-6xHis.

19F CPMG

Potential hit against both targets

- 61.00   - 61.05   - 61.10   - 61.15   - 61.20   [ppm]

■ 30 µM of fragment mix only   ■ 30 µM of fragment mix only
■ 30 µM of fragment mix + PTP1B(1-301)   ■ 30 µM of fragment mix + PTP1B(1-393)
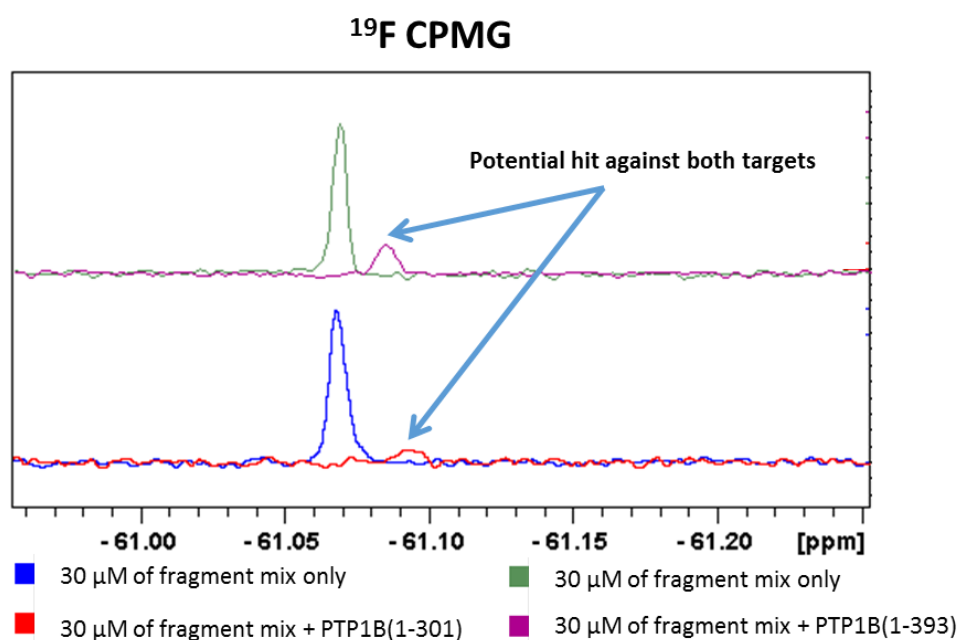
Figure 79. [19]F CPMG-filtered spectra for the identified hit against both PTP1B proteins. (τ = 320 ms)
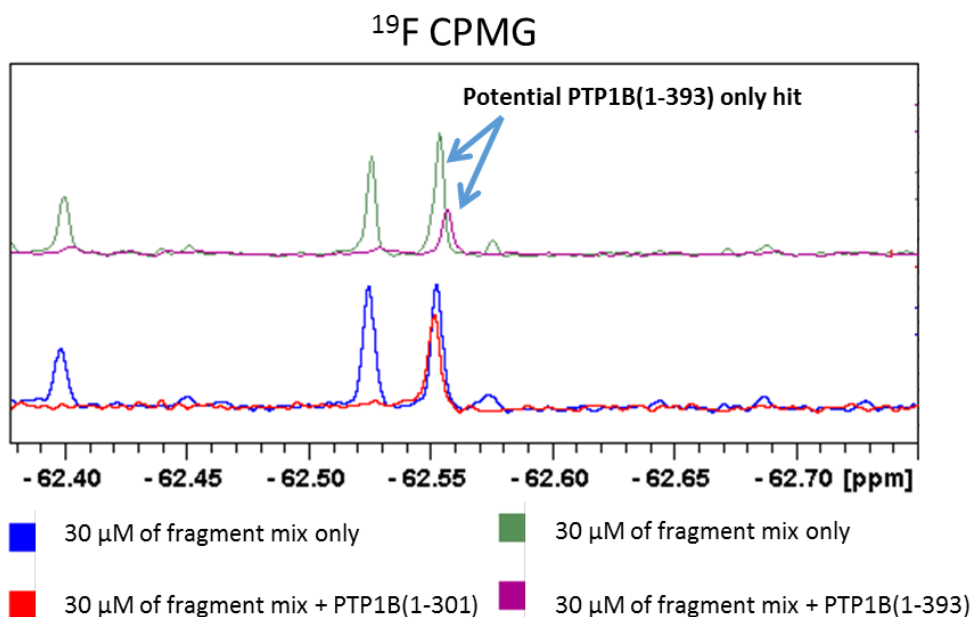
Figure 80. *^{19}F CPMG-filtered spectra for potential hit against PTP1B(1-393)-6xHis . (τ = 320 ms)*

The single compound titration experiments concluded that all of 16 potential hits towards PTP1B(1-393)-6xHis were artefacts because the relaxation patterns of the fragments were similar for both 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis constructs. However, one fragment (compound **V1**) exhibited slight CSPs and new peak appearance when in presence of both PTP1B constructs with more noticeable change for PTP1B(1-393)-6xHis. This was indicative of potential covalent binding events (**Figure 81, Left**). The ligand was identified to be a stereoisomer hence the observed multiple resonance peaks in the spectra.

A fresh batch of compound **V1** was prepared in order to account for possible degradation of the compound during the long-term storage in d$_6$-DMSO. Interestingly, the differences in peak shape and position were noted between the original reference spectra obtained previously and the fresh stock (**Figure 81**). The spectral differences were eventually attributed to the presence of DTT in the buffer solution. Initially, the screening campaign was performed in the presence of 0.5 mM TCEP which acted as a reducing agent (**Figure 81, Left**). After acquiring a fresh compound stock, DTT was used in buffer solutions instead of TCEP due to availability (**Figure 81, Right**). The ^{1}H NMR QC data suggested that some structural changes occurred to **V1** with DTT present, whereas no changes were observed in the presence of TCEP (**Figure 82**). The structure of **V1** contains a morpholine-2-one moiety where the electrophilic carbonyl carbon may be attacked by nucleophilic thiol of DTT causing a formation of thioester bond and thus readily perturbing the observed resonance signals.
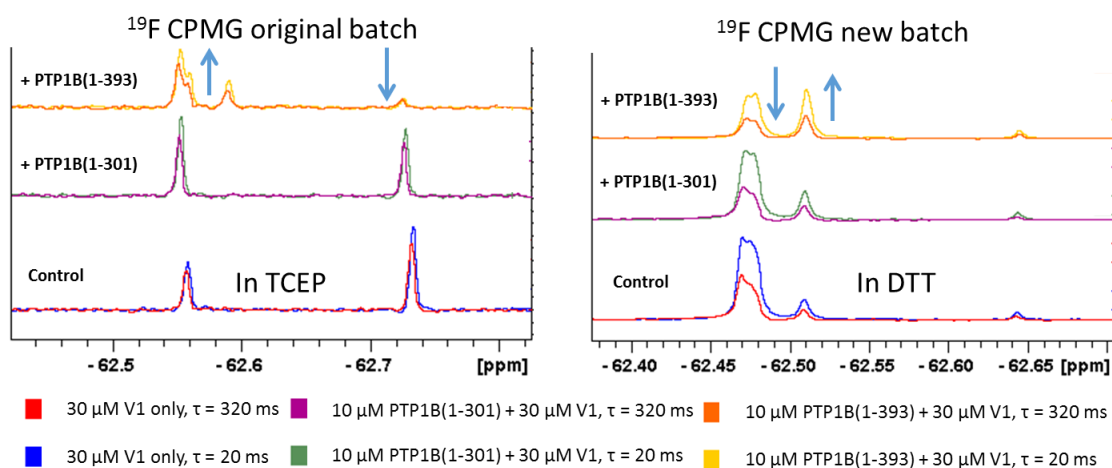
Figure 81. *A comparison of ¹⁹F CPMG-filtered resonance peaks of compound **V1** in the absence and presence of PTP1B constructs in TCEP and DTT-containing buffer solutions. The arrows mark the vector of compound **V1**-specific CSPs.*
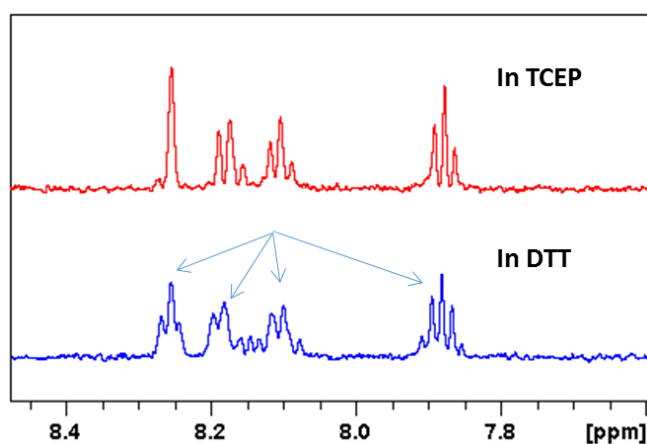
Figure 82. *¹H NMR spectra of compound **V1** in the presence of 1 mM TCEP or DTT. The arrows mark compound **V1**-specific peaks.*

It was also noted that the original 6xHis-PTP1B(1-301) protein stock was stored with 0.5 mM TCEP, whereas PTP1B(1-393)-6xHis construct contained 2 mM DTT. During sample preparation, PTP1B(1-393)-6xHis construct was diluted from 134 μM to 10 μM which decreased the concentration of DTT to ~150 μM (**Figure 83, Left)**. The residual DTT could still have reacted with 30 μM of compound **V1** and given false-positive results when comparing PTP1B(1-301) and (1-393) data even in the presence of 0.5 mM TCEP in the buffer solution. This exact effect of residual DTT has been observed in the very first ¹⁹F CPMG-filtered spectra, where compound **V1** had an extra peak with PTP1B(1-393) construct (**Figure 81, Left**) which was not observed in the presence of PTP1B(1-301).

In order to remove any data ambiguity, both PTP1B protein constructs were buffer-exchanged to replace any residual DTT with 1 mM TCEP. Subsequently, ¹⁹F CPMG-filtered experiments have confirmed CSPs for both PTP1B(1-301) and (1-393) constructs by exhibiting changed

peak intensity ratios of the resonances indicating that compound **V1** interacts with PTP1B(1-393)-6xHis construct more readily than with 6xHis-PTP1B(1-301) (**Figure 83, Right**).
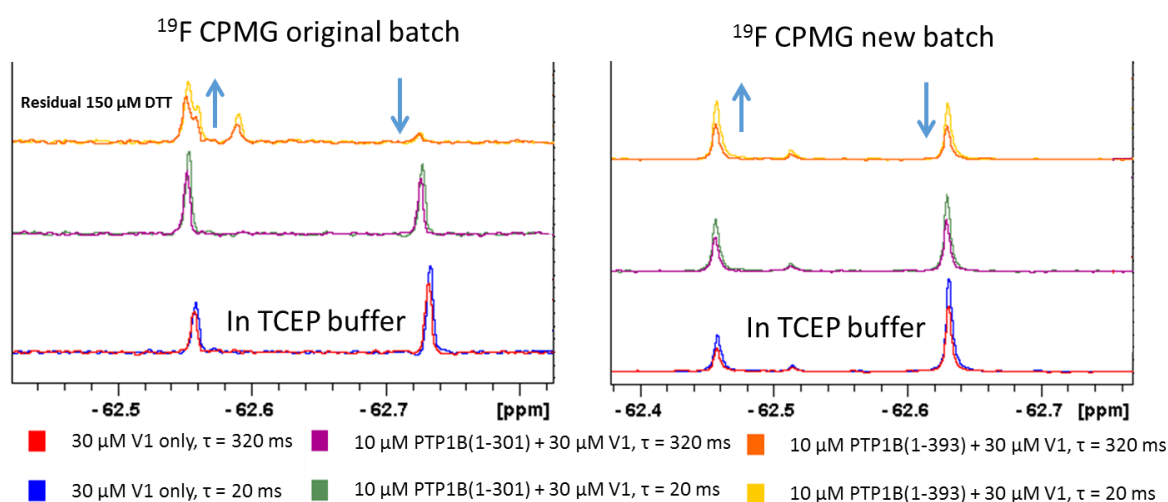
*Figure 83. A comparison of ¹⁹F CPMG-filtered resonance peaks of compound **V1** in the absence and presence of PTP1B constructs with 1 mM TCEP before (**Left**) and after (**Right**) buffer exchange of protein stocks. The arrows mark the vector of compound **V1**-specific CSPs.*

This was followed by protein-observed and ligand-observed ¹H CPMG filtered experiments which did not indicate any increased relaxation patterns for the IDR region of PTP1B(1-393)-6xHis when compared to reference data. The only observed changes were for compound **V1** itself due to covalent binding which were more evident for PTP1B(1-393)-6xHis construct (**Figure 84**).
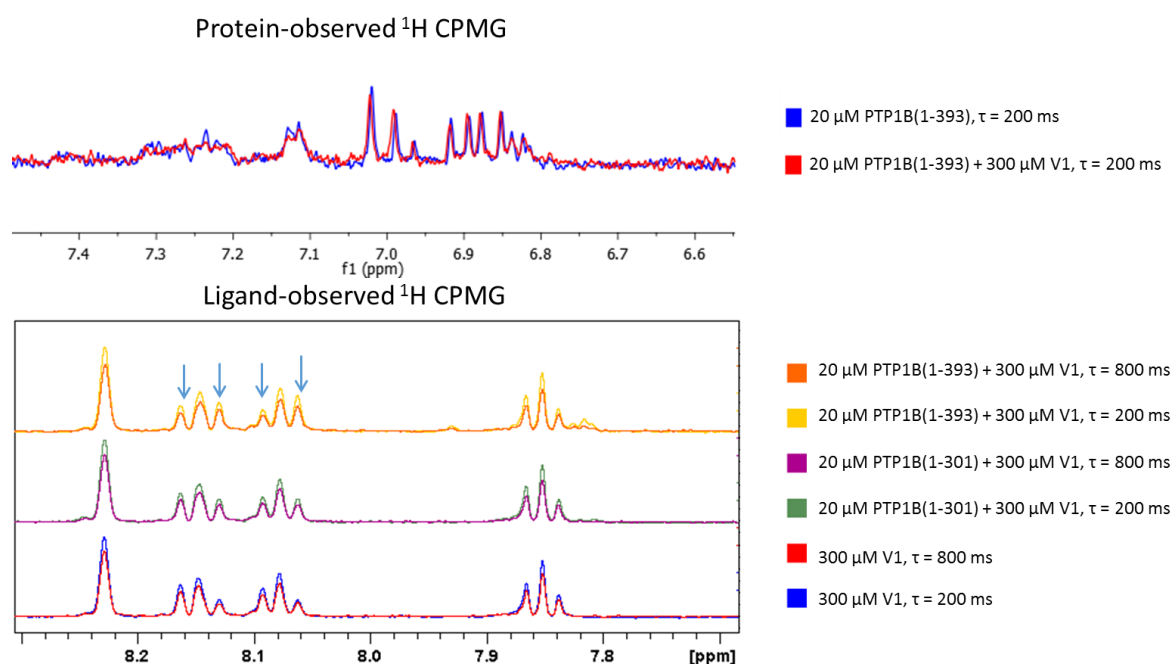
*Figure 84. Comparison of protein- and ligand-observed CPMG-filtered results in the presence and absence of compound **V1** with PTP1B(1-301)-6xHis and PTP1B(1-393)-6xHis protein. The arrows mark the perturbed peaks.*

Due to covalent binding events, it was hypothesized that the bound compound **V1** may affect the diffusion rate of PTP1B(1-393)-6xHis. However, only small diffusion rate changes were observed by NMR DOSY (**Figure 85**). This indicated that the protein hydration radius R_H has not significantly changed. In comparison, MSI-1436 is known to bind to the IDR of PTP1B(1-393)-6xHis and possibly affect its hydrodynamic radius which was not observed in previous experiments with NMR DOSY (**§4.4**). However, compound **V1** may be just too small to affect the IDR relaxation rates and the diffusion rate of the protein significantly enough to be observed.

Compound	Diffusion coefficient **D** x 10^{-11}, m²/s	Difference, %
Control	6.3	0.00
V1	6.48	2.8
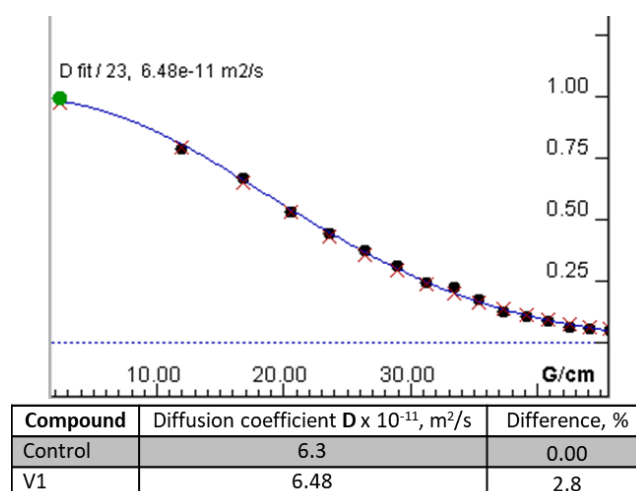
*Figure 85. Comparison of PTP1B(1-393)-6xHis diffusion rate values in the presence and absence of compound **V1**.*

After revisiting ^{19}F CPMG-filtered screening data, it was determined that compound **V1** did not exhibit an increased signal broadening in the presence PTP1B(1-393)-6xHis when compared with 6xHis-PTP1B(1-301) (**Figure 86**). As mentioned previously, the data was originally evaluated using ^{19}F CPMG-filtered spectra with 320 ms delay due to large amount of samples and lack of automated software for quick and accurate processing (**Figure 86, Right**). However, when ^{19}F CPMG-filtered data with 20 ms and 320 ms delay was examined, the relaxation profiles of compound **V1** looked very similar for both PTP1B constructs (**Figure 86, Left**). If 20 and 320 ms relaxation data were used during initial evaluation, the fragment **V1** would have been identified as a non-hit. This shows that a thorough data analysis is needed in order to avoid any false positives during the screening process, especially if the changes are small. The evaluation of large amounts of data could be accelerated by using automated evaluation tools, such as KNIME or MNova.
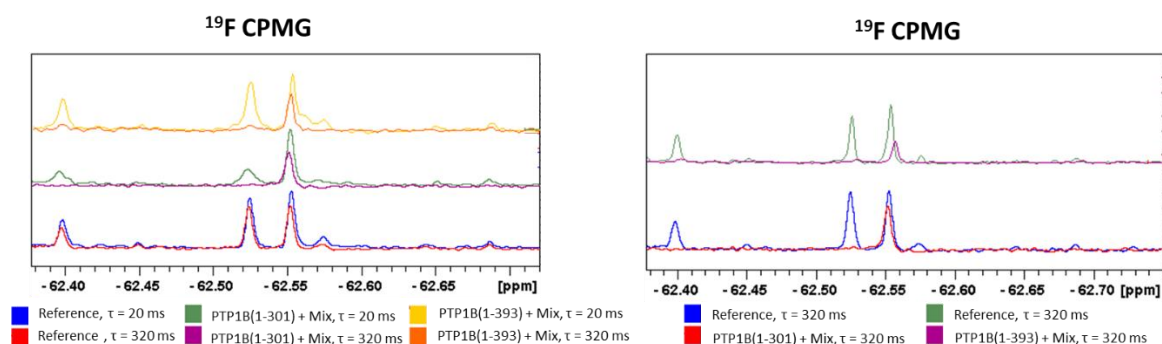
Figure 86. *Comparison of ^{19}F CPMG-filtered ligand-observed relaxation patterns of fragment mix in the presence and absence of 6xHis-PTP1B(1-301) and (1-393)-6xHis. **(Left):** τ = 20 and 320 ms. **(Right):** τ = 320 ms only.*

In order to confirm the potentially covalent interactions between compound **V1** and PTP1B constructs, HPLC-MS studies were performed (**§7.4.1, §7.4.1.1**). The obtained data indicated that 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis form multiple covalent adducts with compound **V1** after 18 hour incubation at RT in the presence and absence of 1 mM DTT or TCEP (**Figure 87, Figure 88**). It was determined that the interaction patterns in the presence of TCEP and without any reducing agent were similar, indicating that TCEP did not react with compound **V1**, whereas DTT readily formed covalent adducts with the compound. Compound **V1** was found to form 1:1 and 1:2 adducts with 6xHis-PTP1B(1-301) construct and 1:1, 1:2 and 1:3 adducts with PTP1B(1-393)-6xHis construct. 6xHis-PTP1B(1-301) has six Cys residues whereas PTP1B(1-393)-6xHis has eight. Out of those eight Cys residues, two are located in the disordered region of PTP1B(1-393)-6xHis (**Figure 89**). It is possible that at least one Cys in the disordered region of PTP1B(1-393)-6xHis has covalently reacted with compound **V1** due to the observed higher number of covalent adducts than for 6xHis-PTP1B(1-301).
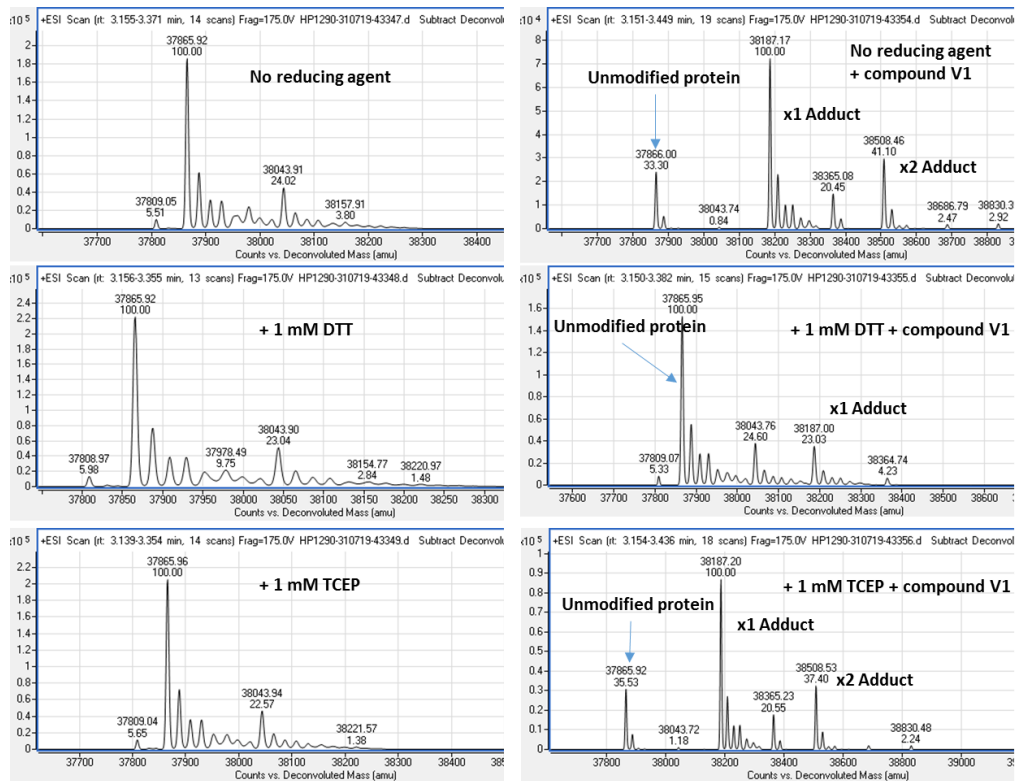
Figure 87. *A comparison of HPLC-MS spectra for 6xHis-PTP1B(1-301) in the absence and presence of DTT/TCEP and* **V1**.
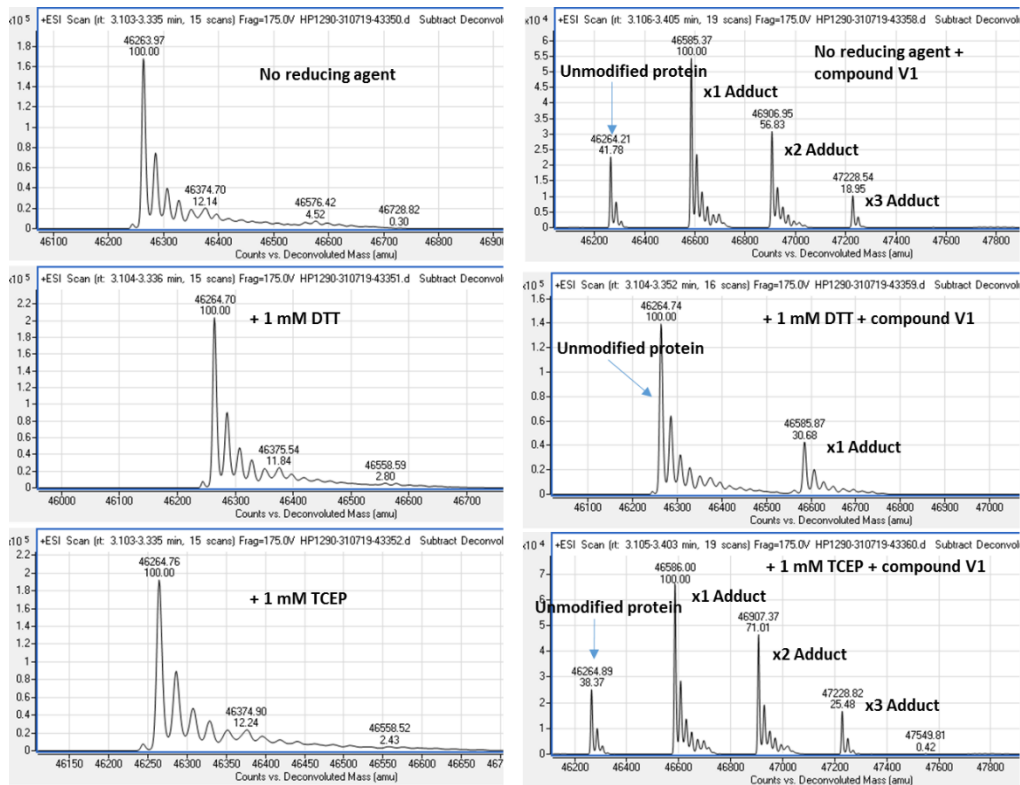
Figure 88. *A comparison of HPLC-MS spectra for PTP1B(1-393)-6xHis in the absence and presence of DTT/TCEP and* **V1**.

PTP1B(1-393)-6xHis

```
        10         20         30         40         50         60
GPGSMEMEKE FEQIDKSGSW AAIYQDIRHE ASDFPCRVAK LPKNKNRNRY RDVSPFDHSR

        70         80         90        100        110        120
IKLHQEDNDY INASLIKMEE AQRSYILTQG PLPNTCGHFW EMVWEQKSRG VVMLNRVMEK

       130        140        150        160        170        180
GSLKCAQYWP QKEEKEMIFE DTNLKLTLIS EDIKSYYTVR QLELENLTTQ ETREILHFHY

       190        200        210        220        230        240
TTWPDFGVPE SPASFLNFLF KVRESGSLSP EHGPVVVHCS AGIGRSGTFC LADTCLLLMD

       250        260        270        280        290        300
KRKDPSSVDI KKVLLEMRKF RMGLIQTADQ LRFSYLAVIE GAKFIMGDSS VQDQWKELSH
```

```
       310        320        330        340        350        360
EDLEPPPEHI PPPPRPPKRI LEPHNGKCRE FFPNHQWVKE ETQEDKDCPI KEEKGSPLNA

       370        380        390        400
APYGIESMSQ DTEVRSRVVG GSLRGAQAAS PAKGEPSHHH HHH
```

6xHis-PTP1B(1-301)

```
        10         20         30         40         50         60
MGSDKIHHHH HHENLYFQGH MASMEMEKEF EQIDKSGSWA AIYQDIRHEA SDFPCRVAKL

        70         80         90        100        110        120
PKNKNRNRYR DVSPFDHSRI KLHQEDNDYI NASLIKMEEA QRSYILTQGP LPNTCGHFWE

       130        140        150        160        170        180
MVWEQKSRGV VMLNRVMEKG SLKCAQYWPQ KEEKEMIFED TNLKLTLISE DIKSYYTVRQ

       190        200        210        220        230        240
LELENLTTQE TREILHFHYT TWPDFGVPES PASFLNFLFK VRESGSLSPE HGPVVVHCSA

       250        260        270        280        290        300
GIGRSGTFCL ADTCLLLMDK RKDPSSVDIK KVLLEMRKFR MGLIQTADQL RFSYLAVIEG

       310        320
AKFIMGDSSV QDQWKELSHE DLEPHN
```

*Figure 89. Sequences of PTP1B(1-393)-6xHis and 6xHis-PTP1B(1-301) with marked Cys (**yellow**) and disordered region (**red**).*

Additional MS/MS and/or NMR studies would be required to directly confirm the binding of compound **V1** to the disordered region of PTP1B(1-393)-6xHis.

In summary, the ^{19}F CPMG-filtered NMR screening process can be summarized with a simplified flow chart below (**Figure 90**).
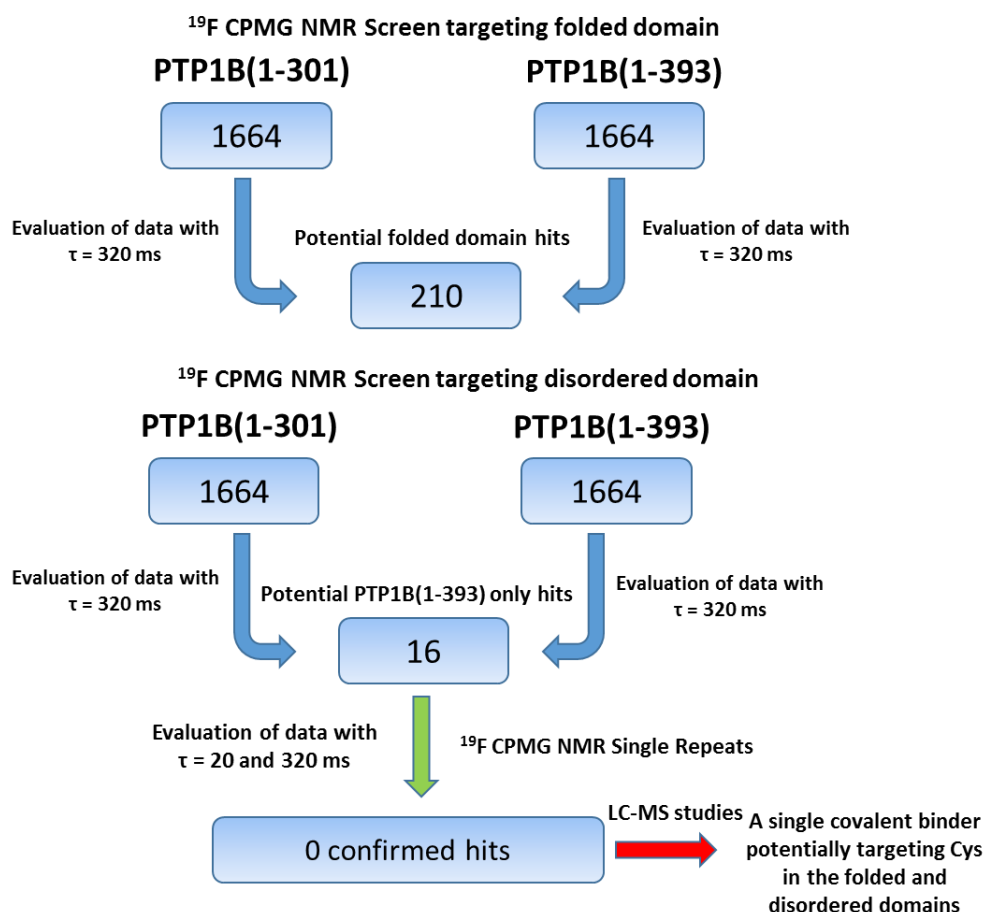
Figure 90. A flowchart of ^{19}F CPMG-filtered NMR screening against 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis.

In conclusion, no fragments were identified to bind to the IDR of PTP1B(1-393)-6xHis using ^{19}F CPMG-filtered experiments while 210 fragments were found to potentially interact with the folded domain of PTP1B. Due to focus of the screening towards the IDR of PTP1B, the hits towards the folded domain were not investigated further. It is possible that the affinity of the fragments to the IDR was not high enough to be detected, suggesting that such approaches may benefit from larger molecules providing more potential interaction possibilities.

Only due to serendipity, a covalent binder **V1** was identified which potentially interacts with Cys residues in the IDR of PTP1B(1-393). Due to previously shown importance of IDR to the activity of PTP1B with the example of MSI-1436, free Cys in the IDR may be targeted by covalent binders in order to force a specific, inactive conformation of the protein. In general, covalent binders have been less widely used in the field of drug discovery than non-covalent ligands due to their potential non-specificity and off-target interactions. However, there are multiple successful cases of targeted covalent inhibitors (TCIs) that have been developed to

specifically interact with targeted proteins [197]. An example of covalent fragment screening for tau K18 protein is described later (**§5.9**).

Following NMR fragment screening, an SPR-based fragment screening was initiated for both PTP1B constructs to identify IDR-specific hits.

4.7 Fragment screening via SPR

4.7.1 Assay development for positive control

In order to perform fragment screening using SPR, it was required to establish a positive control assay for PTP1B in order to evaluate its activity on the surface once immobilized. Suramin was used as a positive control for PTP1B which is a literature-reported inhibitor for phosphatases and other proteins [198], [199]. (**Figure 91**).
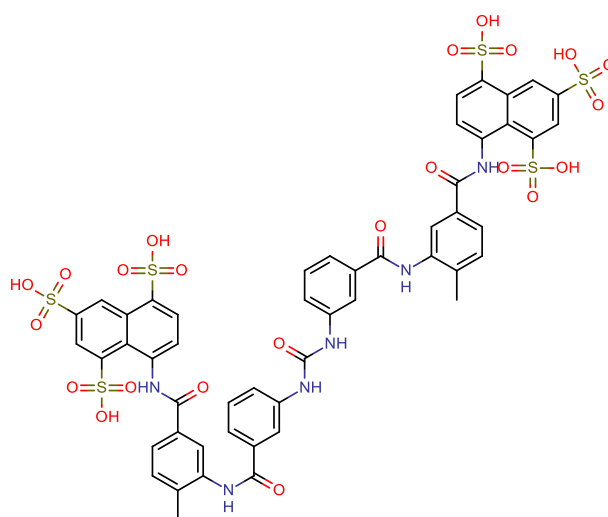
Figure 91. The chemical structure of suramin.

Due to the availability of material, standard amine-coupling approach on CM5 chip was used to immobilize both PTP1B proteins (**§7.4.4, §7.4.4.1**). The experimental data showed that suramin exhibited residual stickiness in the internal flow system of Biacore machine, as noted with blank injections, even after several washing steps with aqueous 50% DMSO solution (**Figure 92**). The residual stickiness of compounds may cause the subsequent signals to be artificially elevated and may prevent accurate evaluation of the data. Therefore, representative blank injections are required to objectively evaluate the interactions between the protein and a fragment. Since fragments have low MW, the expected responses are also low and any systematic perturbations can affect the observed results.
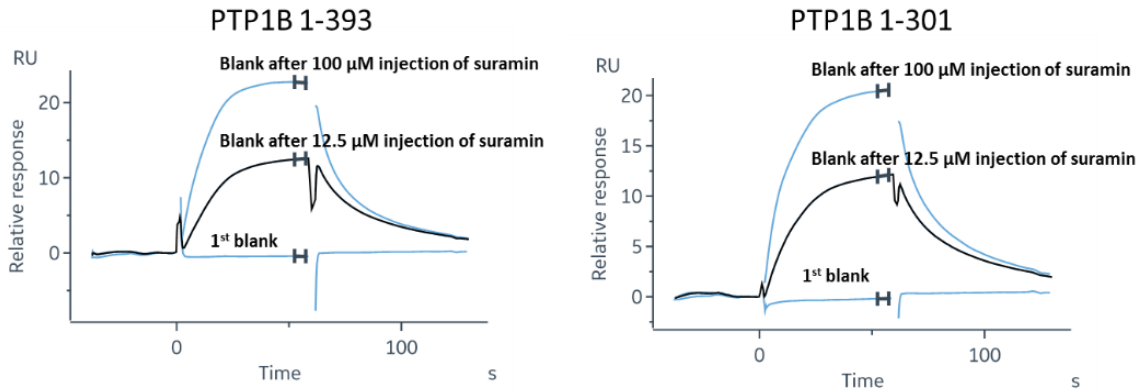
Figure 92. *A comparison of blanks against PTP1B(1-393)-6xHis and 6xHis-PTP1B(1-301) after suramin injections (Immobilization levels: PTP1B(1-393) ~7800 RU, PTP1B(1-301) ~5800 RU). The bolded marks in sensorgrams denote the report points.*

Due to suramin having six sulfate groups, it was predicted that using a high pH solution would help to remove it from the system. After changing the washing solution to 50 mM NaOH and including additional washing steps with carry-over controls, the blank injections showed significant signal reduction to acceptable levels (**Figure 93**).
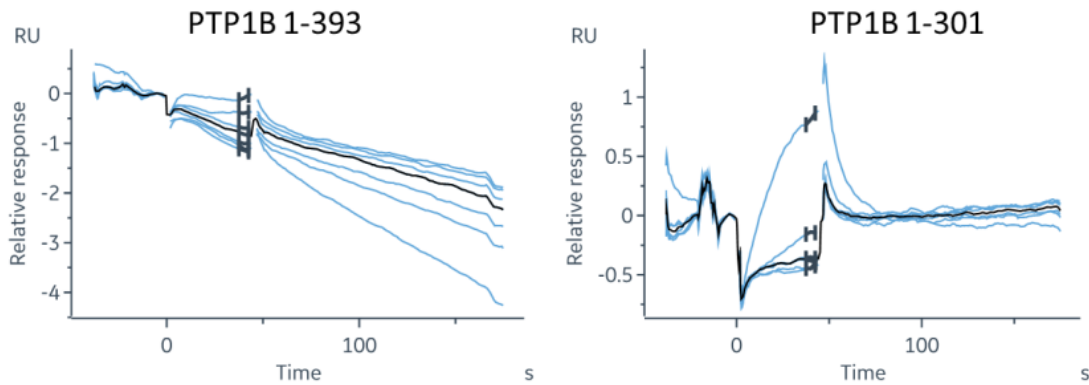
Figure 93. *A comparison of blank injections against PTP1B(1-393)-6xHis and 6xHis-PTP1B(1-301) with additional washing steps (Immobilization levels: PTP1B(1-393) ~6800 RU, PTP1B(1-301) ~6400 RU). The bolded marks in sensorgrams denote the report points.*

After resolving the residual stickiness of suramin, the interactions between PTP1B and suramin were assessed using multi-cycle kinetics. It was noted that the determined K_D values (39 and 42 µM) using 1:1 steady-state binding model were approximately 10 times higher than the previously reported K_i value of 4 µM [200] (**Figure 94**).
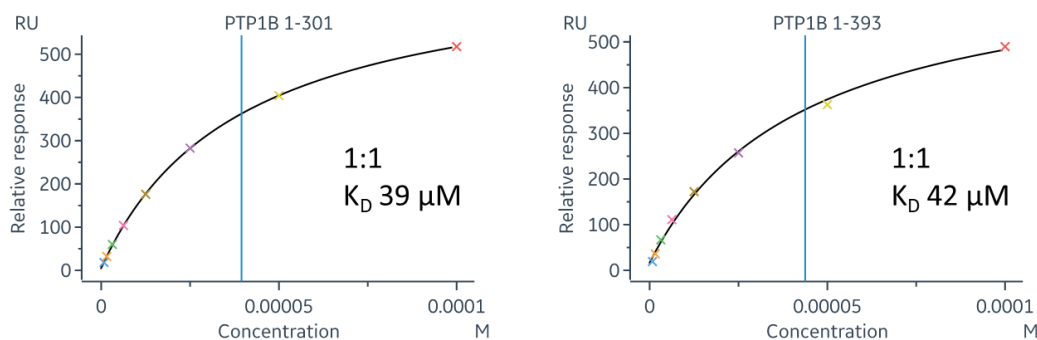
Figure 94. Steady state data fitting using 1:1 binding model for suramin binding behaviour to 6xHis-PTP1B(1-301) (*Left*) and PTP1B(1-393)-6xHis (*Right*).

This observation suggested that the 1:1 binding model may not be representative. The K_D and K_i values in case of PTP1B should be comparable because suramin was previously shown to be a competitive, active site-directed inhibitor via enzymatic assay. As a result, the K_i value calculated from IC_{50} can be directly compared to the experimentally measured K_D value. According to previous PTP1B studies with SPR, the parallel binding model, which includes terms for both specific, high affinity and non-specific, low affinity interactions, is a better representation of the system than 1:1 binding model. This is possibly because of the tendency of suramin to interact non-specifically with other proteins [201]. Suramin has six sulphate groups which carry significant net negative charge at experimental pH of 7.4 in addition to multiple hydrophobic regions. Therefore, non-specific charge-charge interactions are possible.

The fitting of experimental data using the parallel binding model resulted in comparable K_D values for high and low affinity terms which were also close to the literature K_i value of 4 μM (**Table 6**), (**Figure 95, Figure 96**).

Table 6. The determined K_D values for PTP1B(1-301) and PTP1B(1-393) interactions with suramin.

Protein	High affinity K_D, μM	Low affinity K_D, μM
6xHis-PTP1B(1-301)	1.0 +-/ 1.1	44 +/- 1.6
PTP1B(1-393)-6xHis	5.1 +/- 0.2	61 +/- 2.5

Typically, a concentration of 10 x K_D would be used to fully saturate the protein in control SPR experiments. However, this was not possible due to parallel binding events which would cause the high affinity interaction (primary binding) to be masked by the low affinity interaction (secondary binding) at 10 x K_D (**Figure 96**). It has been shown in previous studies that

numerous confirmed ligands do not saturate the binding site of target proteins before secondary binding effects develop [202]. This made the assessment of binding parameters, such as K_D, unreliable. Therefore, it was decided to use suramin at 7.5 µM for tracking the activity of both PTP1B constructs during the fragment screening assays. At such concentration, the majority of the signal would be represented by the high affinity term.
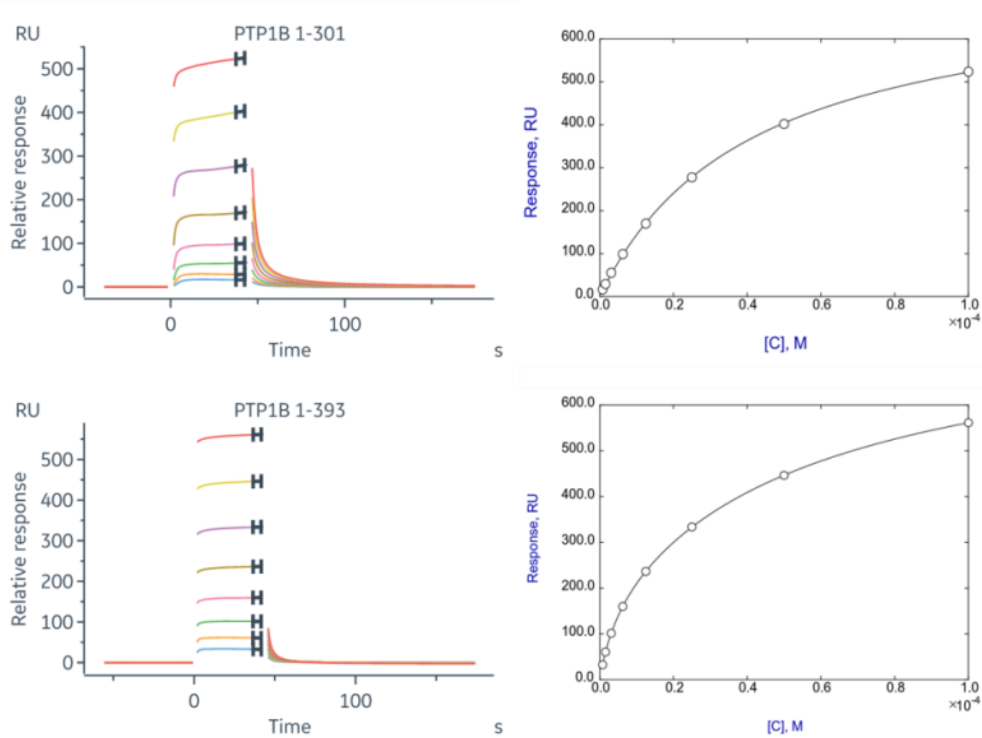
Figure 95. *The kinetic and steady state data plots for 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis with suramin fitted using parallel binding mode (SimFit). Immobilization level: PTP1B(1-301) ~6400 RU, PTP1B(1-393) ~9400 RU. The bolded marks in kinetic sensorgrams note the report points that were used to produce the steady-state data plots.*

$$R_{eq} = \frac{R_{max\,1} * [C]}{[C] + K_{D1}} + \frac{R_{max\,2} * [C]}{[C] + K_{D2}}$$

$$R_{eq} = \frac{R_{max\,2} * [C]}{[C] + K_{D2}} \quad \textbf{Low affinity site}$$

$$R_{eq} = \frac{R_{max\,1} * [C]}{[C] + K_{D1}} \quad \textbf{High affinity site}$$
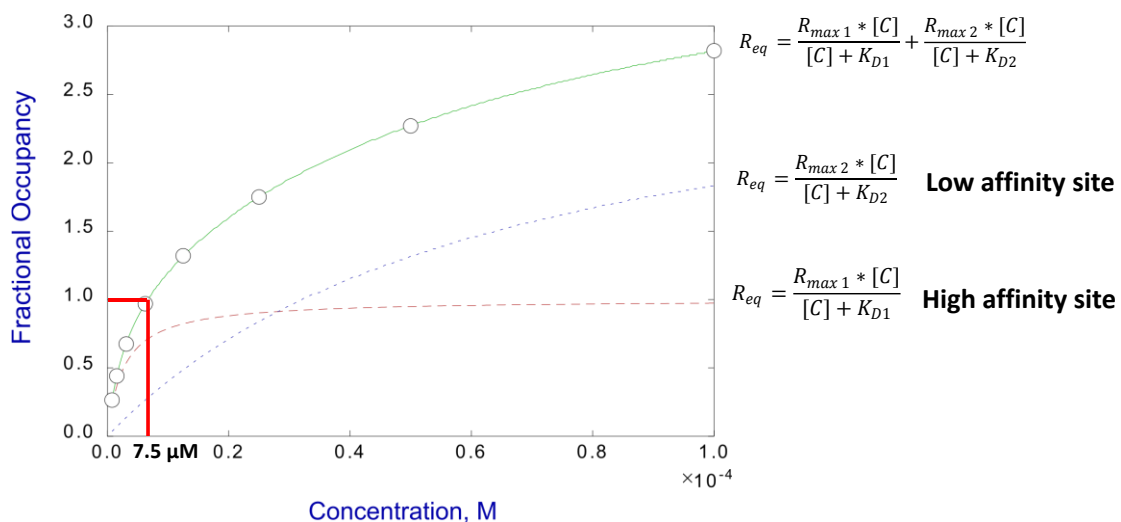
Figure 96. *A graphical representation of parallel binding fitted data for PTP1B:suramin interactions which indicates low and high affinity terms. (SimFit)*

The activity of immobilized protein was determined to be approximately 40% for 6xHis-PTP1B(1-301) and 60% for PTP1B(1-393)-6xHis. This was achieved by comparing suramin response levels to the calculated theoretical R_{max} values (**§7.4.4**). Both PTP1B constructs contain a folded domain which makes the immobilization of the protein onto the chip surface less random than for a completely disordered protein. PTP1B(1-393)-6xHis has a disordered C-terminal region with 7 Lys residues which may influence the preferred orientation of the protein during immobilization. The access of suramin to the binding site of PTP1B may get conformationally restricted more for 6xHis-PTP1B(1-301) than for PTP1B(1-393)-6xHis if PTP1B(1-393)-6xHis is immobilized via the IDR. This assumption was partially confirmed with differing activity levels of the immobilized proteins.

After establishing positive control assays with suramin, fragment screening was initiated by performing a clean screen (CS) assay to remove any sticky compounds that may artificially affect the observed response during subsequent injections.

4.7.2   Clean Screen (CS)

In general, fragments are expected to bind and dissociate quickly from the protein of interest due to relatively low affinities and fast kinetics. If the immobilized protein is known to be active on the surface and the immobilization levels are sufficiently high (R_{max} = 20-30 RU), some residual compound binding with reasonably slow dissociation behaviour can be tolerated (5-10 RU). However, this is only acceptable if the observed residual signal goes back to the pre-injection value before the end of the injection [203]. If the signal does not go back to the baseline after the end of the injection, there is a risk of drifting signal which may cause the automatic evaluation software to indicate the subsequent compound as a non-binder. In addition to this, there is also a risk of 'sticky' compounds occupying the binding pocket of interest. This would lower the observed response levels during the follow-up injections of compounds and any potential binders may be identified as false negatives. Therefore, the identified problematic or 'sticky' fragments should be removed from subsequent experiments [204], [205].

The automatic evaluation of CS data was used to quickly identify fragments with non-specific, residual or 'sticky' binding behaviour to empty and derivatized CM5 S Series chip surface (**Figure 97, Table 7**). The automatic evaluation software subtracted the baseline value of N_{i-1} from N_i and plotted the results into a graph with 10 RU threshold value. Two compounds were

identified to exhibit persistent artefactual binding towards empty CM5 chip surface. Of these, one stuck to all of the protein systems. For 6xHis-PTP1B(1-301), 8 out of 10 compounds were also identified for PTP1B(1-393)-6xHis. For PTP1B(1-393)-6xHis, all 9 compounds also exhibited very slow dissociation behavior for 6xHis-PTP1B(1-301).

Table 7. The number of 'sticky' compounds identified for empty surface and target proteins.

Surface derivatization	Sticky compounds	Surface-specific sticky compounds
Empty	2	1
6xHis-PTP1B(1-301)	10	2
PTP1B(1-393)-6xHis	9	0
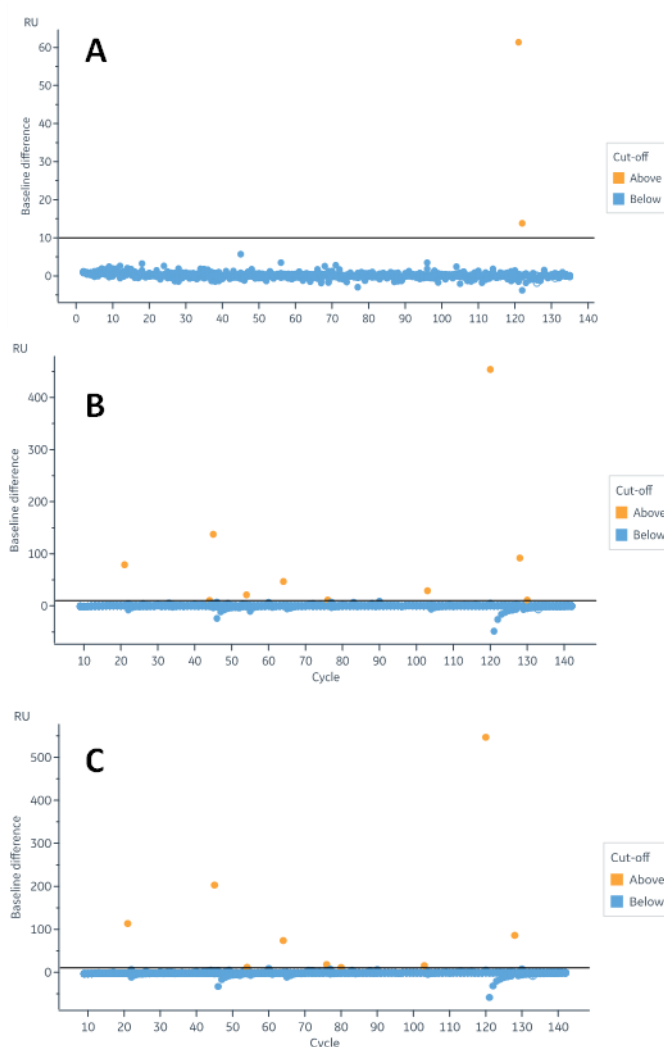
Figure 97. The representation of N_i - N_{i-1} plots for different derivatizations of the chip surface with threshold value of 'residual binding' set to 10 RU. A. Empty Surface (Immobilization level: 0 RU). B. 6xHis-PTP1B(1-301) (Immobilization level: 5800 RU). C. PTP1B(1-393)-6xHis (Immobilization level: 6600 RU).

The identified fragments were then removed from the subsequent Binding Level Screen (BLS) step.

4.7.3 Binding Level Screen (BLS)

The BLS data was used to prioritize potential fragment hits at a single 500 µM concentration that did not exhibit any anomalous or artefactual binding behavior markers, such as slope, slow dissociation or R> R_{max} (**Figure 98, Figure 99**). The slope marker indicates secondary binding events, whereas slow dissociation marker suggests non-desirable residual stickiness to the protein since fragments are expected to bind and dissociate quickly. In addition to this, response levels higher than theoretical R_{max} indicate superstoichiometric, or higher than 1:1, interactions. However, it has been reported that R to R_{max} values up to 3:1 could still be acceptable [203]. After the automated hit evaluation and prioritization procedure, 10% of the hits without anomalous binding behaviour markers (107 compounds for each protein construct system) were taken further to the Affinity Screen (AS) step. The percentage of the library selected can be adjusted accordingly to the size of the library, if necessary.
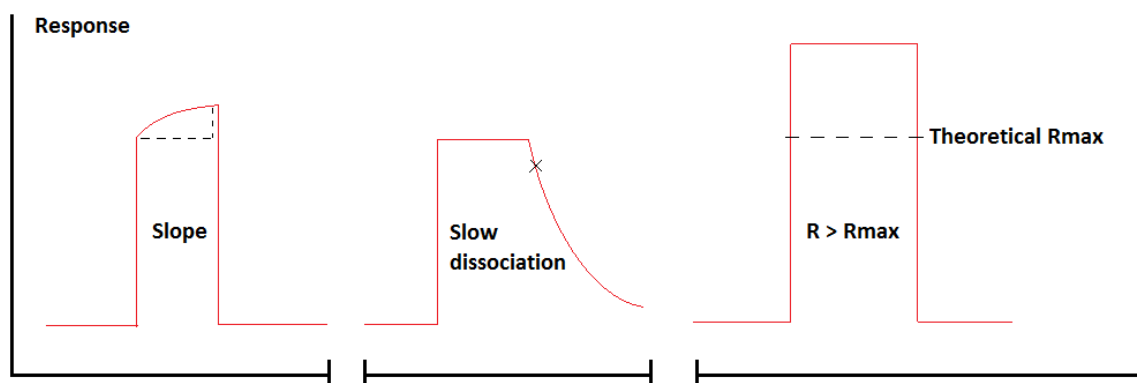
Figure 98. A representation of artefactual binding behaviour markers.
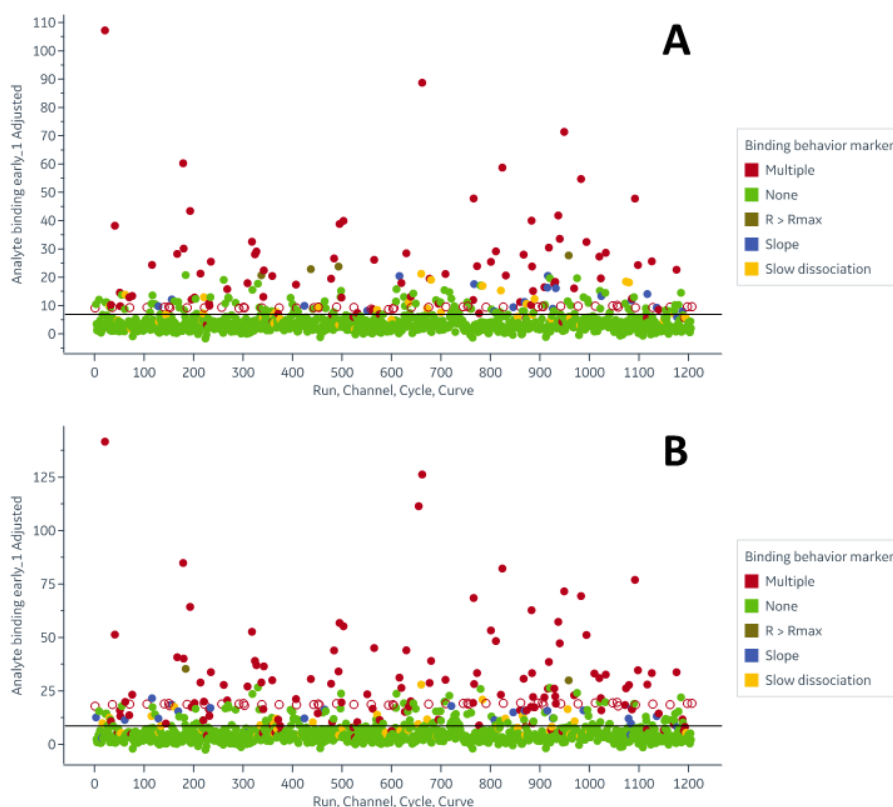
Figure 99. *The identification of well-behaved fragments using BLS approach. The samples exhibiting artefactual binding behaviour markers were flagged with specific colours.* **A.** *6xHis-PTP1B(1-301) (Immobilization level: 6300 RU).* **B.** *PTP1B(1-393)-6xHis (Immobilization level: 10000 RU).*

Following the comparison of automatically selected BLS hits for each PTP1B construct, 23 compounds were identified as binding uniquely to either PTP1B(1-301) or PTP1B(1-393) (**Table 8**). This observation suggested that some fragments may be targeting the IDR.

Table 8. Potentially unique hits for PTP1B proteins from automated BLS data evaluation.

Protein	Potentially unique hits
6xHis-PTP1B(1-301)	23 out of 107 when compared to 1-393 hits
PTP1B(1-393)-6xHis	23 out of 107 when compared to 1-301 hits

All 107 automatically prioritized hits for each PTP1B construct were taken further to the Affinity Screen (AS) step.

4.7.4   Affinity Screen (AS)

The fragments were injected at six different concentrations onto derivatized surfaces in order to generate a dose-response curve and evaluate the affinity of the ligands to the immobilized protein targets. The titration experiments identified multiple potential hits that exhibited saturating or near saturating steady-state curves against both PTP1B constructs, out of which some were potentially unique (**Table 9**).

101

Protein	Hits from AS	Potentially unique hits
6xHis-PTP1B(1-301)	26	19
PTP1B(1-393)-6xHis	30	23

The fragment hits for 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis were further prioritized which yielded 11 potentially unique hits for PTP1B(1-393)-6xHis and 11 potentially unique hits for 6xHis-PTP1B(1-301) (**Figure 100, Figure 101, Figure 102, Figure 103**). A total of seven common PTP1B hits were also identified (**Figure 104**).
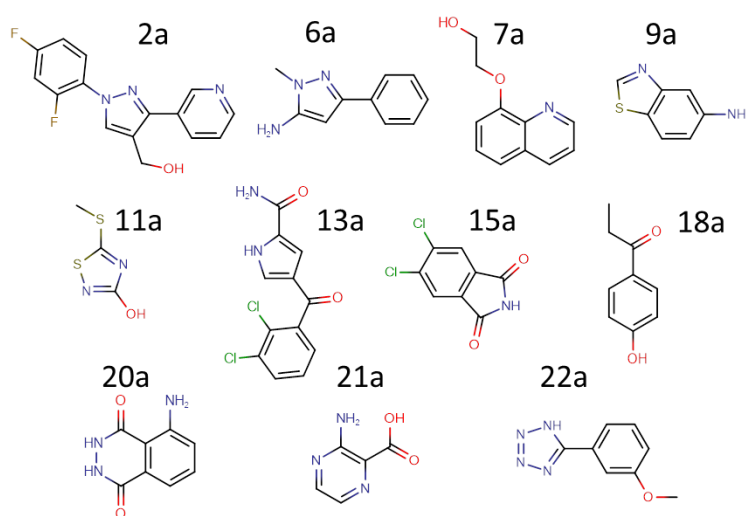
Figure 100. *The prioritized putatively unique hits for PTP1B(1-393)-6xHis, as identified during AS step.*
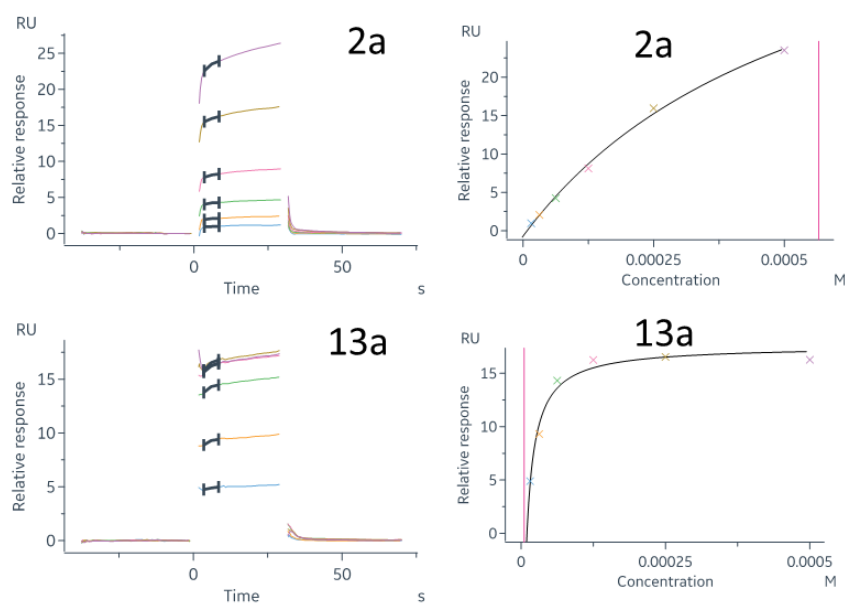
Figure 101. *Representative SPR steady-state fitted plots for PTP1B(1-393)-6xHis. The coloured vertical line represents the calculated K_D values according to the steady-state model. Immobilization level: ~8000 RU. The bolded marks in kinetic sensorgrams denote the report points that were used to produce the steady-state data plots.*
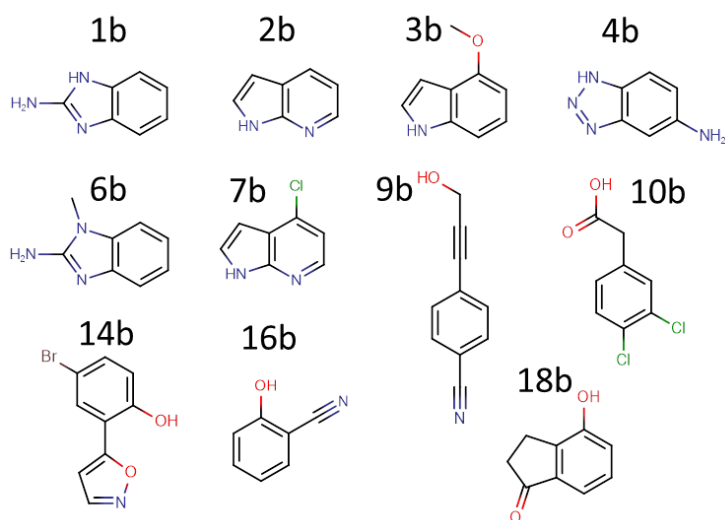
Figure 102. The prioritized putatively unique hits for 6xHis-PTP1B(1-301), as identified via AS step.
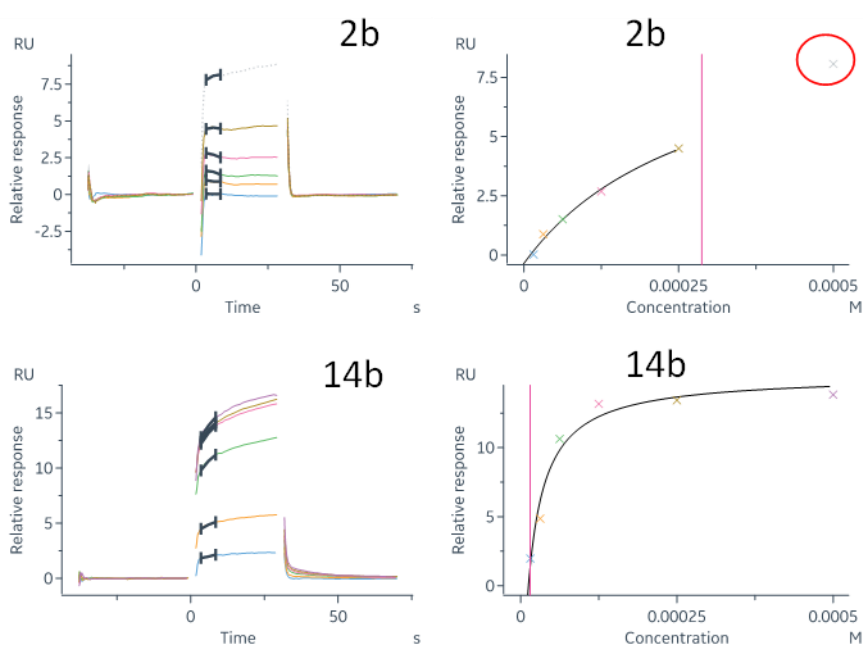
Figure 103. Representative SPR steady-state fitted plots for 6xHis-PTP1B(1-301). The coloured vertical line represents the calculated K_D values according to the steady-state model. Immobilization level: ~7600 RU. The red circle marks excluded data points. The bolded marks in kinetic sensorgrams denote the report points that were used to produce the steady-state data plots.
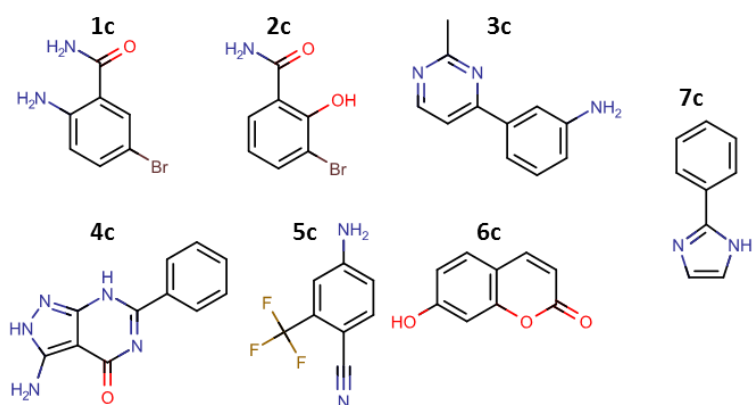
Figure 104. Common hits for both 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis that were identified via AS step.

It was also noticed that compound **13a** for PTP1B(1-393)-6xHis and **14b** for 6xHis-PTP1B(1-301) reached saturating steady-state profile during titration experiments. Therefore, it was possible to determine the surface activity and compare that to the activity determined from suramin titration assay (**§4.7.1, §7.4.4**). It was calculated that the fragment binding levels indicate 35% activity of the surface for both PTP1B constructs if 1:1 binding interactions are assumed (**Table 10**). Previously, suramin response levels suggested a 40% activity for 6xHis-PTP1B(1-301) and 60% for PTP1B(1-393)-6xHis. Considering possible deviations in the theoretical interaction model, the activity levels were comparable for 6xHis-PTP1B(1-301) construct. However, the determined activity levels for PTP1B(1-393)-6xHis differed almost by a factor of two. This implied that the parallel binding model may not be entirely representative for PTP1B(1-393)-6xHis:suramin interactions.

Table 10. The determined surface activity levels for immobilized PTP1B constructs using different compounds.

Compound	Surface activity, %	
	6xHis-PTP1B(1-301)	PTP1B(1-393)-6xHis
Suramin	40	60
13a	-	35
14b	35	-

As mentioned previously, the AS data evaluation was based on indicative saturating dose-response behavior. It is important to note that some of the hit compounds exhibited non-specific binding at concentrations higher than 250 µM which was compound-specific (**Figure 103**). Therefore, in some cases the data from concentrations above 250 µM were excluded in order to account for secondary binding events at higher concentrations.

Due to time and material constraints, the identified PTP1B construct-specific hits could not be tested against another construct using SPR in order to confirm the uniqueness of the fragments. In addition to this, it was decided to re-inspect the previously obtained BLS data to evaluate the automated evaluation workflow.

4.7.5 Manual evaluation of BLS

After the manual inspection of BLS data, it was concluded that the automatic hit prioritization workflow can produce ambiguous data in situations where common or unique hits for similar targets are to be identified. While single-target fragment screening campaigns benefit from automatic prioritization and offer fast data evaluation, comparative multiple-target fragment screening results may require manual inspection in order to reach an accurate result.

Initially, the automated BLS data evaluation suggested that out of the 107 compounds selected for each PTP1B construct, 23 compounds were unique for each PTP1B construct (**Table 8**). However, after the manual comparison of the BLS response signals it was concluded that all 23 potentially unique compounds for 6xHis-PTP1B(1-301) construct also produced a comparable signal against PTP1B(1-393)-6xHis, and vice versa (**Table 11, Table 12**).

Subsequently, it was also concluded that the majority of the hits against 6xHis-PTP1B(1-301) were not automatically selected for PTP1B(1-393)-6xHis to be taken to the AS step because of artefactual binding behavior markers (**Table 11, Figure 98**). This observation may be attributed to the slightly higher immobilization levels of PTP1B(1-393)-6xHis due to its higher MW in order to achieve similar theoretical R_{max} values. The derivatized surfaces with higher immobilization density can exhibit non-desirable behavior, such as aggregation or protein-protein interactions on the surface.

Table 11. The comparison of 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis BLS responses for automatically prioritized 'unique' PTP1B(1-301) BLS hits. *Artefactual binding behaviour marker. †Hits below threshold value (<8.69 RU).

6xHis-PTP1B(1-301) Unique BLS Hit	6xHis-PTP1B(1-301) Unique BLS Hit Signal, RU	PTP1B(1-393)-6xHis BLS Signal, RU
1	7.04	7.422†
2	19.06	27.83*
3	8.04	13.31*
4	8.70	11.43*
5	7.11	8.64*
6	7.12	10.79*
7	11.48	17.57*
8	12.92	15.88*

9	10.65	13.24*
10	11.52	16.45*
11	11.53	16.31*
12	10.47	12.63*
13	6.84	9.69*
14	7.08	7.41†
15	12.96	17.07*
16	12.43	17.95*
17	9.84	15.21*
18	8.88	12.09*
19	10.74	18.4*
20	20.76	35.33*
21	7.6	7.42†
22	11.27	15.47*
23	8.75	10.34*

In contrast, the majority of PTP1B(1-393)-6xHis hits were not selected for 6xHis-PTP1B(1-301) due to the response signals being only marginally lower than the automatically determined threshold line (**Table 12, Figure 105**).

Table 12. *The comparison of 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis BLS responses for 1-393 'unique' BLS hits.* **Artefactual binding behaviour marker †Hits below threshold value (<6.9 RU)*

PTP1B(1-393)-6xHis Unique BLS Hit	PTP1B(1-393)-6xHis Unique BLS Hit Signal, RU	6xHis-PTP1B(1-301) BLS Signal, RU
1	8.88	6.08†
2	9.08	6.63†
3	8.73	5.53†
4	9.74	6.61†
5	10.21	6.65†
6	8.76	6.31†
7	8.83	6.27†
8	11.06	7.58*
9	11.01	6.08†
10	9.87	6.81†
11	9.52	5.94†
12	9.24	6.65†
13	9.01	6.28†
14	8.69	6.1†
15	9.15	5.98†
16	9.09	6.12†
17	9.79	6.65†
18	9.04	6.19†
19	9.23	6.5†
20	10.13	5.94†
21	9.56	6.75†

| 22 | 8.95 | 6.07† |
| 23 | 8.89 | 5.37† |

1-301

1-393
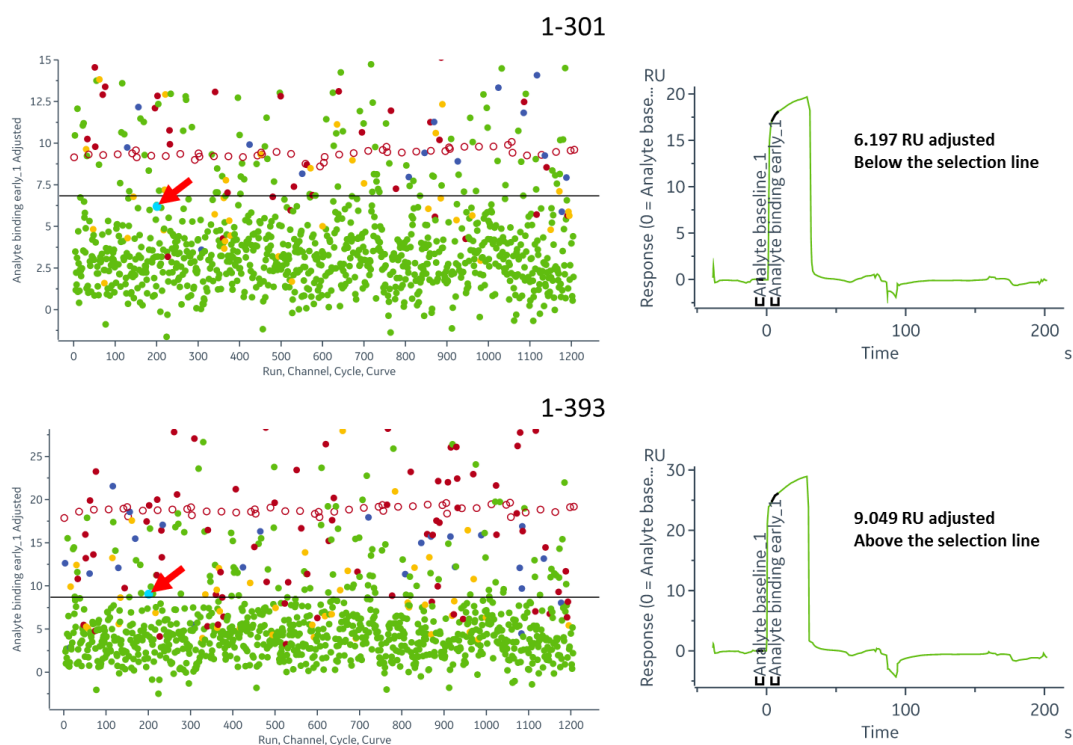
Figure 105. *The representation of BLS plots for both PTP1B constructs indicating possible occurrence of false positives/negatives during data analysis. Binding behaviour markers:* **green** *– no artefactual binding behaviour,* **yellow** *– slow dissociation,* **blue** *– slope,* **brown** *– R>R$_{max}$,* **red** *– multiple artefactual binding behaviours,* **open-red** *– measured surface activity.*

A comparative manual inspection of BLS data between both PTP1B constructs was also performed in order to identify any potentially specific fragments binding to a single PTP1B construct. The initial BLS responses had to be normalized to account for different immobilization levels used during BLS step which would affect theoretical R$_{max}$ values (**Table 13**). Therefore, the reported BLS signals for fragments with PTP1B(1-301)-6xHis surface were adjusted by a factor of 1.29 in order to match the expected response levels with the same immobilization of PTP1B(1-393)-6xHis. It was assumed that the actual activity of both PTP1B constructs are similar, as suggested previously (**Table 10**).

Table 13. *The difference between R$_{max}$ values for each PTP1B-derivatized surface used in BLS.*

Immobilized protein	BLS immobilization level, RU	Theoretical R$_{max}$ for 150 Da fragment, RU	R$_{max}$ ratio to each other
6xHis-PTP1B(1-301)	6300	24.9	0.78
PTP1B(1-393)-6xHis	10000	32	1.29

In total, 22 potentially specific fragments were identified for PTP1B(1-393)-6xHis (**Figure 106**). A fragment was considered to be a hit if the normalized signal observed for PTP1B(1-393)-6xHis was >50% higher than for 6xHis-PTP1B(1-301) to account for possible concentration differences due to the manually prepared samples. Signals of ≤ 1 RU were disregarded as they were comparable to the baseline noise and blank injections.
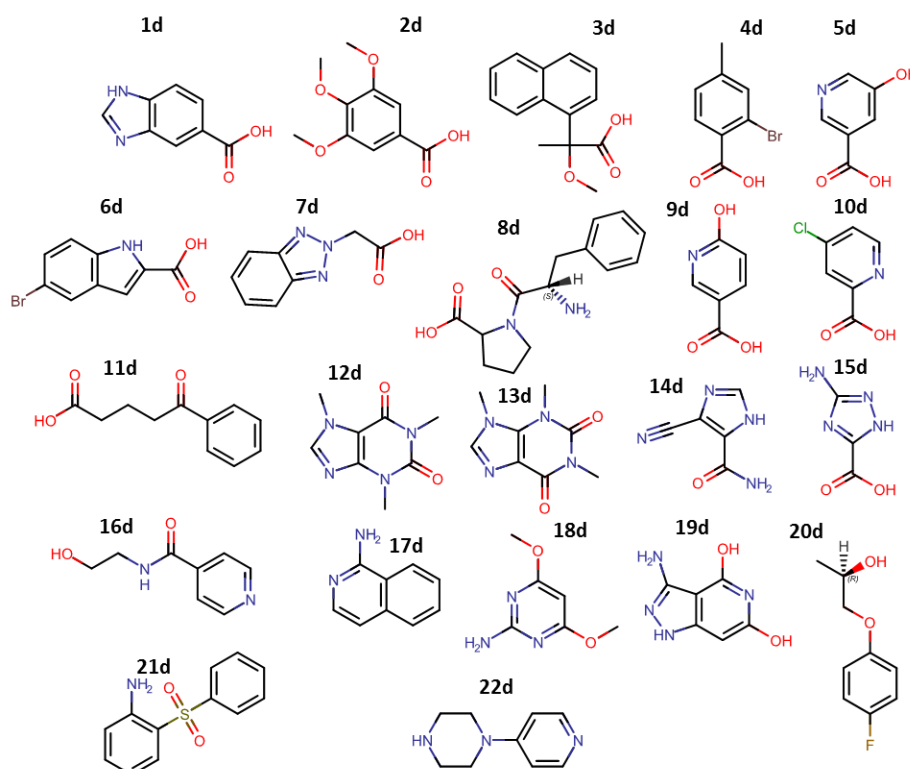
Figure 106. The identified hits for PTP1B(1-393)-6xHis only determined after manual inspection of BLS data.

The identified hits would need further characterization against both PTP1B constructs via SPR in order to evaluate the exact target specificity of the compounds.

4.8 Conclusions

PTP1B was used as a model system that represented a folded protein with an IDR. While the folded domain of PTP1B(1-301) was produced efficiently, several difficulties were noted during the production of PTP1B(1-393), including the proteolytic degradation of the IDR and low yields. The proteolysis of PTP1B(1-393) was solved with N-terminal GST and C-terminal 6xHis tag design, whereas the yields were increased by using a stronger T7 promoter.

The literature-reported interactions between MSI-1436 and PTP1B(1-393) were confirmed via 1D NMR with low micromolar affinities, whereas MST detected potentially ambiguous interactions. In addition to this, potential aggregate formation of MSI-1436 was observed

above 200 μM in aqueous solution by 1D NMR, possibly due to its chemical similarity to a detergent molecule.

While NMR DOSY determined the hydrodynamic radii R_H of PTP1B(1-301) and PTP1B(1-393) to be similar to the literature-reported values, the technique did not detect any change in R_H values in the presence of MSI-1436 potentially due to overlaying signals in the methyl region and not being able to observe protein as a whole during the experiment. It is also possible that the conformational changes may have to be significantly larger to be detectable by NMR DOSY.

The SPR technique was shown to be prone to charge-charge interactions as MSI-1436 was observed to interact non-specifically with the dextran surface, possibly due to excess charge on the ligand. The charge-charge interactions were partially confirmed by using higher salt conditions. These results suggest that the assay would further need specific experimental design. The SPR-based fragment screening campaign against both PTP1B constructs identified several potent fragments binding to the folded region of PTP1B, whereas fragments binding to the IDR still need confirmatory experiments. In addition to this, automatic workflow evaluation for fragment screening by SPR was identified to be error-prone when it comes to comparing results for two similar targets. Therefore, a manual inspection may be required for more accurate results.

^{19}F NMR CPMG-filtered approach was used to screen CF_3 fragment library against PTP1B(1-301) and PTPB(1-393) to detect IDR-directed fragments. The results indicated multiple non-covalent hits against the folded domain but none for the IDR. Only due to serendipity, a covalent binder **V1**, potentially reacting with Cys, was identified for both PTP1B(1-301) and PTP1B(1-393), with potential interactions with the IDR of PTP1B(1-393).

5 Tau K18

5.1 Introduction

The neurodegenerative disorders that are caused by filamentous accumulations of microtubule-associated tau protein in brain cells are called tauopathies. As knowledge about pathologies, diagnostic approaches and potential treatments for neurodegenerative diseases increased, tauopathies were further classified as primary or secondary, depending on whether tau protein is solely or together with other proteins contributing to certain pathology. Primary tauopathies include such disorders as Pick's disease, frontotemporal dementia and Parkinsonism linked to chromosome 17 (FTDP-17), and progressive supranuclear palsy [206]–[208]. Secondary disorders include such disorders as Alzheimer's disease (AD), Lewy body disorders and prion disease [209]–[211]. It is estimated that around 30 million people around the world are affected by tauopathies. According to the WHO, AD on its own accounts for up to 60% of 50 million cases of dementia worldwide, with 10 million new cases every year [212].

It has been noted that tau aggregates found in brain tissue exhibited increased levels of phosphorylation than typically observed and mostly contained double helical stacks, or paired helical filaments (PHFs) [213]. There are 79 potential phosphate acceptor residues in tau protein, out of which around 30 have been reported to be phosphorylated previously [214]. The negatively charged phosphates are considered to neutralize the positive charge of specific regions in the protein and thus, decrease microtubule:tau interactions and promote tau-tau association. The increased propensity for tau protein to form aggregates can also be caused by specific mutations in the microtubule-associated protein tau (MAPT) gene. These mutations can decrease the affinity of tau protein to the microtubules and thus increase the unbound population of tau. The increased unbound population of tau may become more prone to phosphorylation which in turn may promote aggregate formation [215].

Tau protein is expressed as six different isoforms in humans due to alternative splicing events [216]. These isoforms differ in the number and structural arrangement of four main functional domains: amino terminal region [217], proline rich region [218], microtubule binding domain [219] and carboxy terminal region [220] (**Figure 107**). The microtubule binding domain contains either 3 or 4 imperfect repeat (R) regions which have been determined to directly contribute to the microtubule binding affinity in tau. The different number of repeat regions in tau isoforms is produced again due to splicing events, as mentioned previously. It has been

determined that two hexapeptide motifs PHF6(VQIVYK) and PHF6*(VQIINK), located in 2R and 3R regions of microtubule-binding domain, play an important role in the formation process of PHFs and are found at the core of formed fibrils [221], [222]. The aggregation has also been noted to be enhanced by oxidation of Cys residues and interactions of tau with polyanionic molecules *in vitro,* such as heparin [223], [224]. In addition to this, single mutations in the PHF6 region that directly disrupt β-structures have been shown to reduce tau's ability to form aggregates [225]. These studies and observations have revealed important insights about tau aggregation and provided structural basis for developing aggregation inhibitors.

Multiple kinases and phosphatases, such as GSK-3, CDK5 and PP2A, have also been targeted in order to prevent or reduce the hyperphosphorylation levels of tau [226]. However, even though *in vitro* models have exhibited promising data, specific target inhibition without side effects has been difficult due to the targeted enzymes having multiple substrates.

Direct targeting of tau protein may provide an alternative approach. Because of the importance of PHF6 and PHF6* regions in fibril formation, tau K18 (4R repeat) and tau K19 (3R repeat) constructs have been used to track and evaluate the aggregation inhibition by LMW ligands. Tau K18 and K19 are found in the isoforms of full length tau (**Figure 107**).
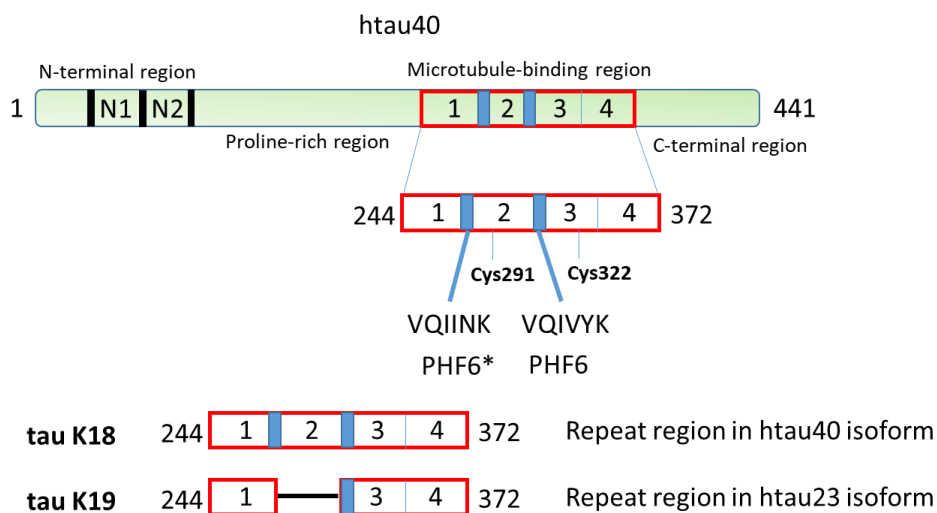
Figure 107. A structural representation of htau40 isoform which includes both N-terminal regions and 4R repeat regions. A zoomed-in part in the microtubule-binding region shows the location of PHF6 and PHF6 sequences. Tau K18 and K19 constructs are also represented that exist in different isoforms*

To date, multiple molecules have been identified as modulators of the tau aggregation process and state, including polyphenols, rhodanines, benzothiazoles and phenothiazines, using Thioflavin T (ThT) or Thioflavin S (ThS) assays [227]–[238] (**Figure 108**). However, the

interaction mechanisms between tau and the majority of the ligands has not been fully understood mainly due to the fact that monomeric tau is an IDP and the oligomeric/fibril species are difficult to study with standard biophysical tools. In addition to this, some ligands may interact only with monomers, oligomers or fibrils which in turn may modulate the further evolution of the complex towards insoluble fibril formation [233].
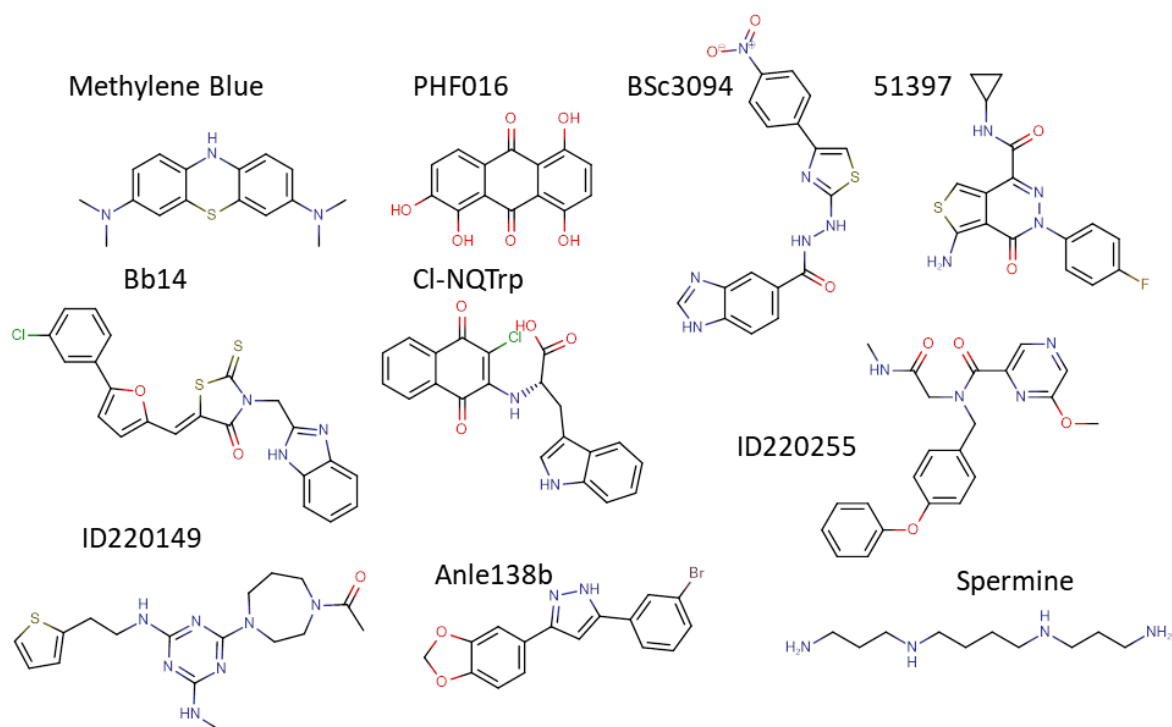
Figure 108. The chemical structures of multiple tau aggregation inhibitors.

Some structural insight, however, has been provided with the investigation of BSc3094 molecule which has been shown to interact with tau K18 [236]. ^{1}H NMR STD data was consistent with binding of the ligand to tau K18 with micromolar affinity, whereas SPR data indicated slow dissociation kinetics. In addition to this, Methylene Blue (MB) and compound **51397** have been shown to oxidize Cys residues in tau K18 and K19 constructs [234], [239]. However, the aggregation process was only inhibited for tau K18 which has two Cys residues, but not for K19 which contained a single Cys residue. MB interactions with Cys residues have been noted in another investigation with full length tau protein using NMR and SAXS approaches [240]. MST was used to directly measure binding affinities of some ligands to the preformed tau fibrils, such as thiazine red, a molecular tracer T807 and lansoprazole [241]. Another recent publication employed MST as well by showing that MB interacted with

monomeric tau protein with nanomolar affinities, possibly involving Cys residues [242]. These examples suggest that direct observation of IDP:ligand interactions is possible.

In this study, multiple constructs of monomeric tau K18 were produced. This was followed by an exploratory study that investigated literature-reported ligand interactions with monomeric tau K18 using biophysical NMR, MST, SPR and HPLC-MS techniques. Limited structure-activity relationship (SAR) studies were also performed with synthesized near neighbours (NNs) of the identified binder to further explore binding interactions. Finally, non-covalent and covalent fragment screenings were performed to identify potentially novel binders to tau K18 using NMR, SPR and HPLC-MS techniques.

5.2 Protein production and characterization

In order to investigate literature-reported ligands and perform non-covalent and covalent fragment screening, multiple versions of tau K18 protein had to be designed and produced.

As mentioned previously, tau protein forms toxic paired helical filaments (PHFs) under certain conditions, such as hyperphosphorylation (**§5.1**). After additional analysis, the microtubule binding regions tau K18 and K19 were found at the core of the fibrils and multiple approaches have been used to target this region with LMW ligands. Two Cys residues (C291,C322) for tau K18 were also identified in the fibril region that form disulphide bonds which are considered to promote PHF formation. In this project, in order to avoid any potential tau K18 instability and change of structure over time, it was decided to mutate both C291 and C322 to the serine double mutants C291S, C322S while maintaining similar properties of the protein. Serine residues retain hydrophilic properties of cysteine while not possessing any redox abilities of thiols.

The first construct for 6xHis-tau K18(C291S,C322S) was ordered in pJ411 T7-driven expression vector (**pDV1**) with 5'-NdeI and 3'-XhoI cloning restriction sites in order to enable molecular cloning approach, if necessary. 6xHis tag allowed affinity purification and labelling techniques for biophysical assays. The initial expression attempts of **pDV1** showed low level expression of target protein, as determined by HPLC-MS and SDS-PAGE using Maxwell 16 Polyhistidine Protein Purification Kit (**§7.3.1, §7.3.3, §7.3.4, §7.4.1**) (**Figure 109**).
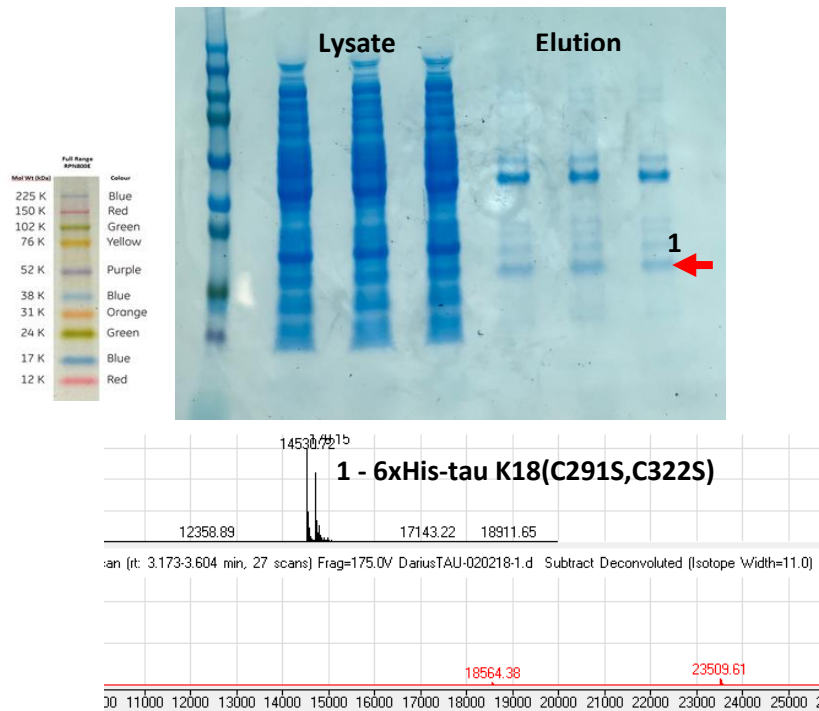
Figure 109. Expression evaluation of 6xHis-tau K18(C291S,C322S) construct in pJ411 vector system by SDS-PAGE and HPLC-MS.

In order to increase expression levels, 6xHis-tau K18(C291S,C322S) coding sequence was subcloned to a T7-driven pET vector system, pET-24a (**pDV2**). To achieve this, 6xHis-tau K18(C291S,C322S) coding plasmid (pRT2069 or **pDV1**) and pET-24a vector-containing plasmid (pRT2044, in-house) was digested, purified and ligated (**§7.2.2**) (**Figure 110**).
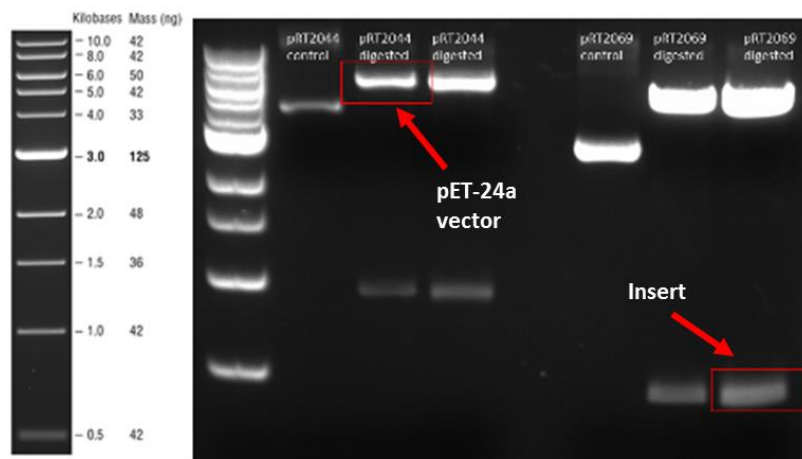
Figure 110. The plasmids containing pET-24a vector (pRT2044) and 6xHis-tau K18(C291S,C322S) (pRT2069) were digested and run on 1% agarose gel. The red-marked parts were excised for ligation.

Subsequent expression and purification attempts showed that 6xHis-tau K18(C291S,C322S) was produced in significantly larger amounts with pET-24a vector (**pDV2**) compared to pJ411 (**pDV1**) (**§7.3.1, §7.3.3**)(**Figure 109, Figure 111**). The optimized **pDV2** construct yielded 32 mg of 6xHis-tau K18(C291S,C322S) per 1 L of LB media. Multiple expression vectors are available

to be used for recombinant protein expression in various *E. coli* cell lines with each having its own specific elements, such as promoters and growth conditions [243]. The pET vectors are designed to use T7 RNA polymerase to drive the protein expression in high yields [244]. However, protein-coding DNA sequence or protein itself can be a problem for expression due to variety of reasons, such as secondary structure formation *in vivo*, toxicity, abrupt translation termination or protein aggregation. This may cause impaired protein production or even host cell death. The pJ411 expression vector with 6xHis-tau K18(C291S,C322S) insert could have exhibited one or several mentioned problems during expression which may have impaired the observed expression levels [245]. Similar inefficient expression of protein in pJ411 expression vectors was observed for PTP1B (**§4.2.2**).

The concentration of the expressed protein was determined using Pierce BCA protein assay due to low molar extinction coefficient of tau K18 (ε = 1490 1/M*cm) (**Figure 112**) (**§7.3.7**).
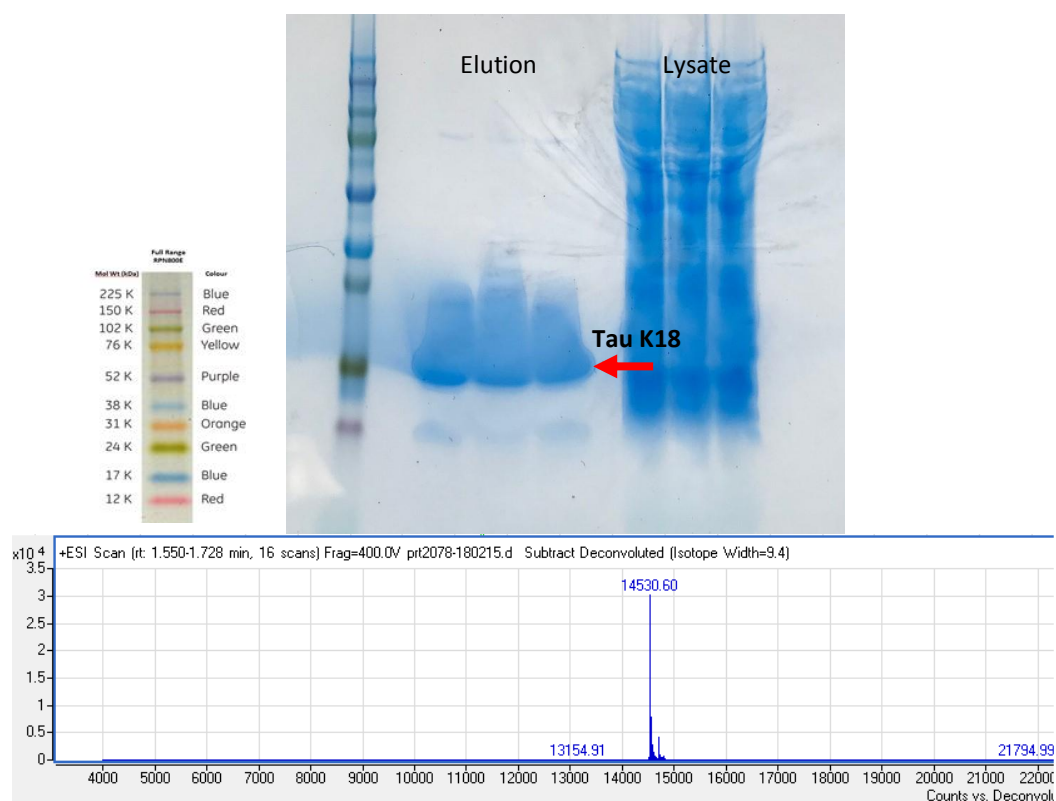
***Figure 111**. Expression evaluation of 6xHis-tau K18 (C291S,C322S) in pET-24a vector (**pDV2**) system by SDS-PAGE and HPLC-MS.*
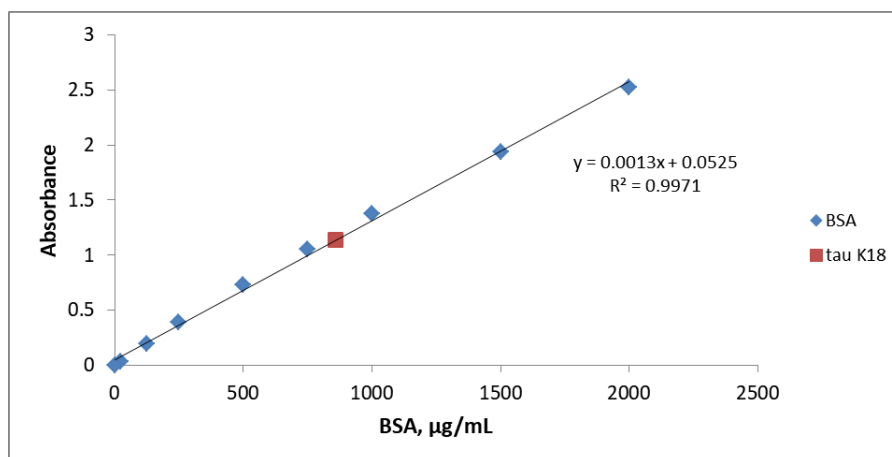
Figure 112. An example of concentration determination for 6xHis-tau K18(C291S,C322S) using BCA assay.

After successful expression of 6xHis-tau K18(C291S,C322S), the following native and single Cys-containing tau K18 constructs were produced in order to enable covalent fragment screening:

- 6xHis-tau K18(C291, C322S) (**pDV3**)
- 6xHis-tau K18(C291S,C322) (**pDV4**)
- 6xHis-tau K18(C291,C322) (**pDV5**)

The mutations of Ser to Cys (S291C and S322C) were achieved by using site-directed mutagenesis (SDM) with **pDV2** plasmid as template and appropriate primers (**§7.2.3**). It was noticed that SDM of S291C did not produce any correct sequence-containing clones when the S322C mutation was already present in the sequence (**pDV4**). It is possible that with the S322C mutation present, the original S291C primer did not bind sufficiently enough to the template causing SDM to fail (**Primer A, §7.2.3**). However, a single nucleotide mutation is not expected to influence DNA plasmid structure to the point of causing failed PCR, especially since the same primer worked previously to introduce a single S291C mutation. The problem was solved by extending the S291C primer by 4 nucleotides at both 5' and 3' sides (**Primer C, §7.2.3**). This was followed by successful expression and purification of all Cys mutants with comparable yields (**Figure 113**). It is possible that PCR attempt for S291C mutation was not successful due to non-optimal experimental design from the beginning, including temperature and timing, rather than problems with the primer.
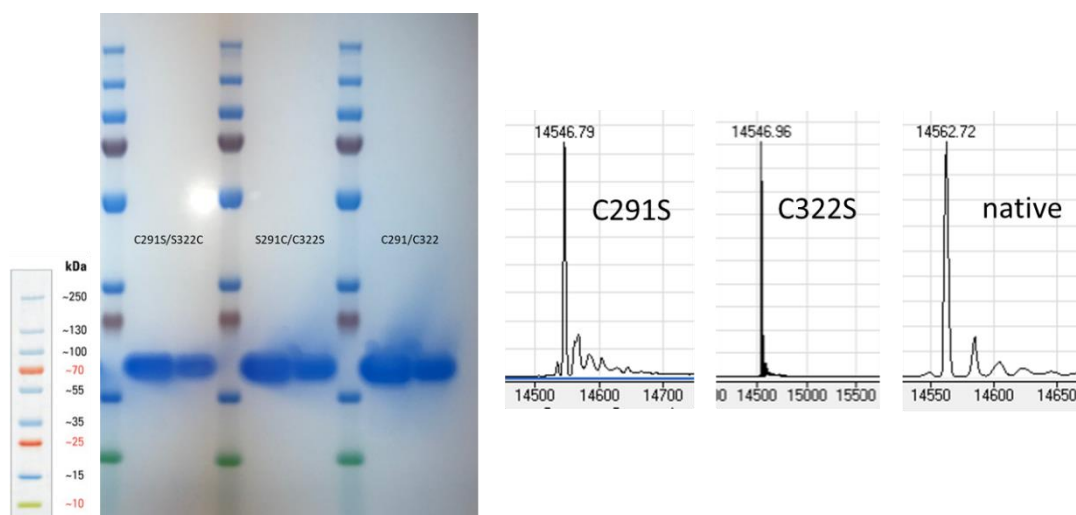
Figure 113. The expression evaluation of 6xHis-tau K18 Cys mutants by SDS-PAGE and HPLC-MS.

The same expression and purification protocols were applied to express the following isotopically labelled tau K18 constructs for 2D and 3D NMR experiments (**§7.3.2, §7.3.3.4**):

- ^{15}N-labelled (>99% labelling efficiency)
    - 6xHis-tau K18(C291S,C322S) (**pDV2**)
    - 6xHis-tau K18(C291,C322) (**pDV5**)
- ^{15}N,^{13}C labelled (>99% labelling efficiency)
    - 6xHis-tau K18(C291S,C322S) (**pDV2**)

Due to a low number of aromatic residues in tau K18 (Y310 and F346), it was decided to also label Y310 with ^{19}F derivative of Tyr to enable ^{19}F NMR approach (**§7.3.2, §7.3.3.4**). Amino acid specific ^{19}F labelling has been successfully used to study structure, binding and aggregation of IDPs, such as α-synuclein [246]. Specific labelling of aromatic amino acids was achieved by using glyphosate during expression which suppresses the synthesis of aromatic amino acids by inhibiting shikimate synthesis pathway. By manually introducing labelled aromatic amino acids into growth media, the cells can incorporate the labelled amino acids into the sequence in an efficient manner. This procedure was used to achieve specific ^{19}F labelling, namely for 3-fluoro-Y310 (**Figure 114**). The incorporation efficiency was determined by MS peak integration to be ~66% which was sufficient to observe the signal in ^{19}F NMR. The labelling efficiency may potentially be improved by increasing incubation time after addition of glyphosate and/or adding a larger amount of ^{19}F-labelled amino acid. This would ensure that the shikimate pathway is completely inhibited together with more ^{19}F-labelled amino acid available for specific labelling.
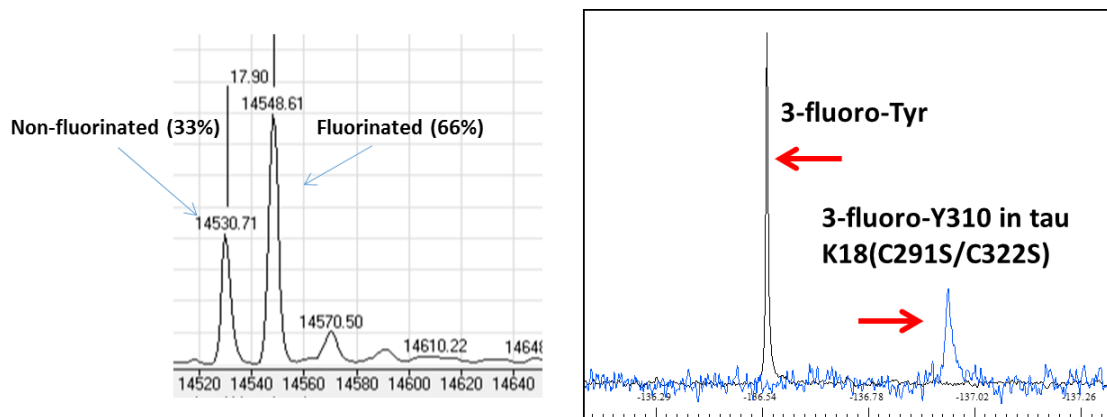
Figure 114. The expression evaluation of 6xHis-tau K18(C291S,C322S)(3-fluoro-Y310) by HPLC-MS and ^{19}F NMR.

For SPR technique, several immobilization approaches are available. While most of SPR experiments involve covalently immobilized proteins on the carboxymethylated dextran matrix, maintaining their functionality, the IDPs, such as tau K18, are prone to being immobilized in a non-functional conformation. Therefore, oriented immobilization can be a preferred immobilization method to maintain the functional profile of IDPs. While 6xHis tag is used for immobilizing proteins on Ni-NTA-modified surfaces, the surfaces may not be stable enough over extended period of time to perform SPR experiments, such as fragment screening. To solve this problem, a more stable approach is available. Biotin is a known cofactor that is involved in many metabolic processes in humans and many other organisms. It has been shown that interactions between biotin and streptavidin or avidin are one of the strongest non-covalent interactions known with K_D of ~10^{-15} M [247]. These specific biotin-streptavidin interactions have been exploited to yield many applications, including affinity-based separation, immunoassays and immobilization with defined orientation [248], [249].

It was discovered that *E. coli* contains an endogenous enzyme BirA which allows specific biotinylation of target proteins via AviTag sequence *in vitro* [250]. After expression and purification of AviTag-tau K18(C291S,C322S) in pET-24a expression vector (**pDV6**), the target protein was biotinylated with 1:1 ratio *in vitro* by BirA enzyme with >95% efficiency (in-house) and used in subsequent SPR assays (**Figure 115**) (**§7.3.2, §7.3.3.4**).
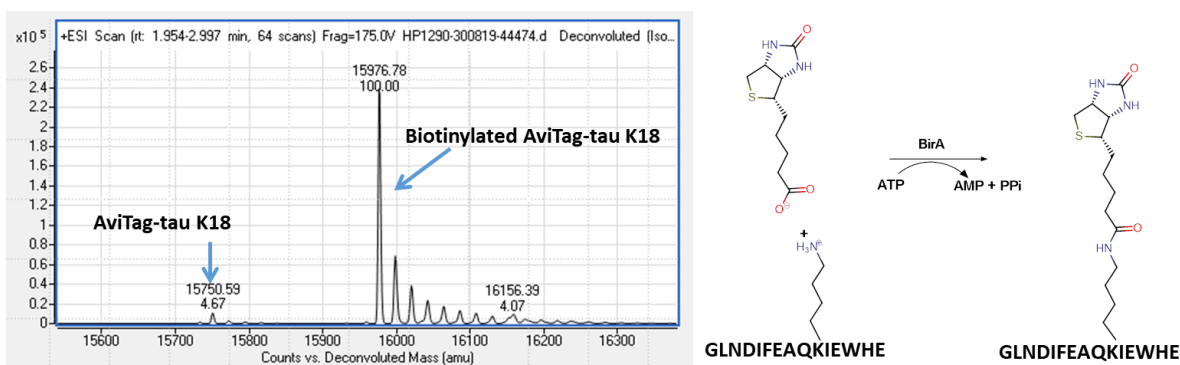
Figure 115. The evaluation of biotinylation efficiency of AviTag-tau K18(C291S, C322S) by BirA after 2 hours via HPLC-MS.

In summary, the following tau K18 constructs were expressed in *E. coli* BL21(DE3):

- 6xHis-tau K18(C291S,C322S) (**pDV2**)
    - Unlabelled
    - Uniformly ^{15}N labelled
    - Uniformly ^{13}C,[15]N labelled
    - Specific labelling of Y310 with 3-fluoro-Tyr
- Unlabelled 6xHis-tau K18(C291,C322S) (**pDV3**)
- Unlabelled 6xHis-tau K18(C291S,C322) (**pDV4**)
- Native 6xHis-tau K18 (C291,C322) (**pDV5**)
    - Unlabelled
    - Uniformly ^{15}N labelled
- Unlabelled biotinylated AviTag-tau K18(C291S, C322S) (**pDV6**)

Once multiple tau K18 constructs were available, literature-reported tau K18 ligands were acquired and characterized by NMR.

5.3   Characterization of tau K18 ligands

As mentioned previously, a number of compounds have been reported in the literature to modulate the formation and integrity of tau monomers, oligomers and/or fibrils (**§5.1**). However, the majority of ligand efficacy data has been obtained using cellular or phenotypic experiments, including directly observing tau fibril formation. In case of BSc3094, 1D ^{1}H NMR-STD was used to determine affinity to monomeric native tau K18(C291,C322) construct while SPR approach indicated strong interactions due to slow dissociation [236]. In addition to this, MB was found to interact with full length native tau and tau K18(C291,C322) with 86 and 125

nM affinity, respectively via MST approach. However, no other biophysical characterization of tau K18:ligand interaction mode was available.

Ten literature ligands were kindly provided by Servier (FRA) in powdered form and were prepared in d$_6$-DMSO (**Figure 116**).
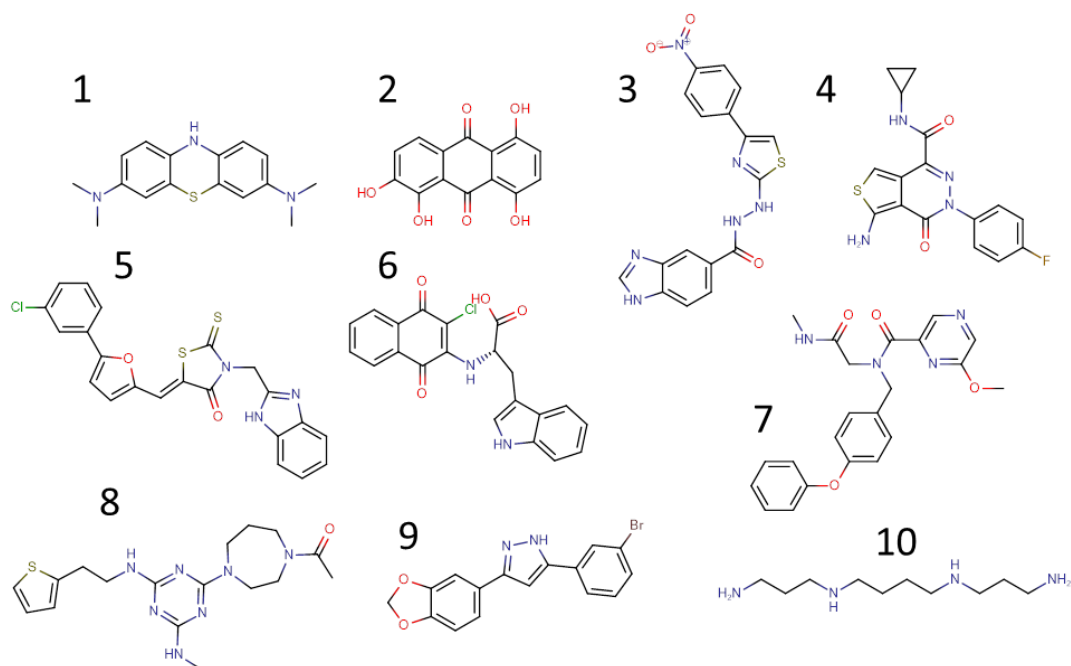
Nr	Literature ID	Solubility in µM @ 50 mM NaPi pH 6.6, 25 mM NaCl, 5% D$_2$O, 298 K				
		125	250	500	1000	2000
1	Methylene Blue [234]	🟩	🟩	🟩	🟥	
2	PHF016 [235]	🟥				
3	BSc3094 [231]	🟩	🟥			
4	51397 [237]	🟩	🟥			
5	Bb14 [231]	🟩	🟥			
6	Cl-NQTrp [238]	🟩	🟩	🟩	🟩	🟩
7	ID220255 [231]	🟩	🟩	🟥		
8	ID220149 [231]	🟩	🟩	🟩	🟩	🟩
9	Anle138b [229]	🟥				
10	Spermine [230]	🟩	🟩	🟩	🟩	🟩

Figure 116. Chemical structures and solubility profiles of literature-reported ligands for tau K18.

The structural integrity of the compounds was confirmed by 1D ^{1}H NMR in d$_6$-DMSO (**§7.4.2**). Due to the design of NMR experiments and expected low affinities of the compounds, the ligands were expected to be soluble at least at 125 µM in aqueous solution of 50 mM NaPi pH 6.6, 25 mM NaCl, 5% D$_2$O, 298 K, using 50 µM DSS at 0.0 ppm as a reference. Solubility experiments showed that eight out of 10 compounds were soluble at or above 125 µM (**Figure 116**).

The soluble compounds were then further investigated using protein-observed 2D NMR experiments to assess any binding interactions with monomeric ^{15}N-labelled 6xHis-tau K18(C291S,C322S).

5.4   Detection of ligand binding to tau K18 via protein-observed NMR

The evaluation of protein-ligand interactions by 2D ^{1}H-^{15}N HSQC NMR approach is considered to be a "golden standard" in drug discovery field due to technique's ability to identify the amino acid residues that are perturbed by protein:ligand interactions. 2D SOFAST-HMQC (SF-HMQC) technique makes this approach more applicable for screening purposes for small proteins due to fast data acquisition and retained data quality when compared to typical 2D HSQC technique [251].

The dose-response titration experiments showed that only compound **6** (**Cl-NQTrp**) produced significant CSPs for ^{15}N-labelled 6xHis-tau K18(C291S,C322S) (**Figure 117**). Literature-reported assignments were used to partially assign amino acid cross peaks where unambiguous assignments could be made [252]. The identified amino acids involved in interactions agreed with previously literature-reported binding mode for compound **6** as it interacts with PHF6 region [238] (**Figure 107**).

1H-15N SF-HMQC
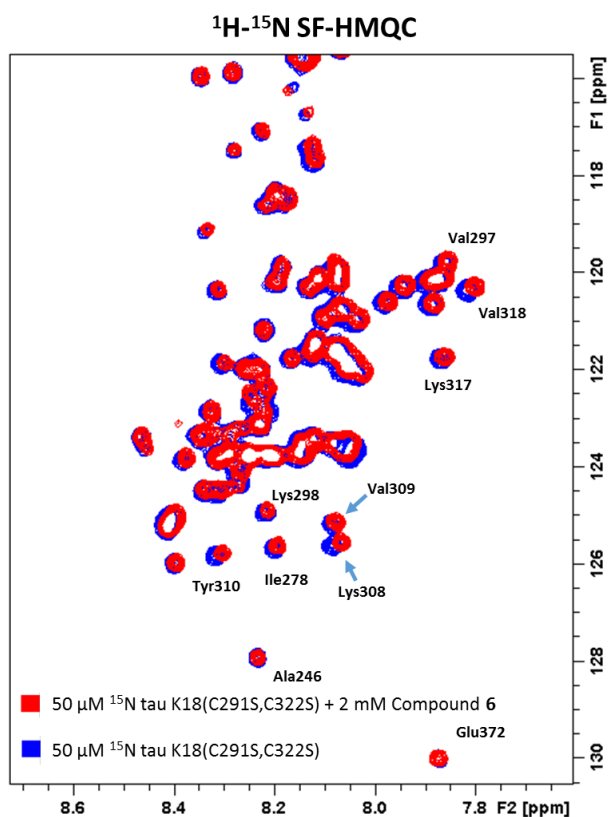
Figure 117. Observed CSPs for 15N-labelled 6x-His-tau K18(C291S,C322S) in the absence and presence of compound 6.

Several naphthoquinone-tryptophan derivatives have been previously shown to inhibit aggregation-prone proteins, such as Aβ and α-synuclein [253]. Compound **6** contains a carboxylic acid group and exists as a carboxylate at neutral pH. 6xHis-tau K18(C291S,C322S) has a theoretical pI of ~9.8 which results in a highly positive charge in neutral pH. It is possible that charge-charge interactions form between protein and compound **6** which disrupt fibril formation [238]. It has been also reported that negatively charged heparin and poly-Glu molecules also produce significant CSPs for tau K18 [254]. The possible charge-charge interactions were assessed with control titration experiments that included acetic acid, ethanolamine and L-Trp. The results concluded that 6xHis-tau K18(C291S,C322S) was not affected in the same manner as with compound **6** (**Figure 118**). The CSPs caused by acetic acid or ethanolamine were of similar level but in opposite vectors and could be attributed to the titration of ionizable groups present in the protein.
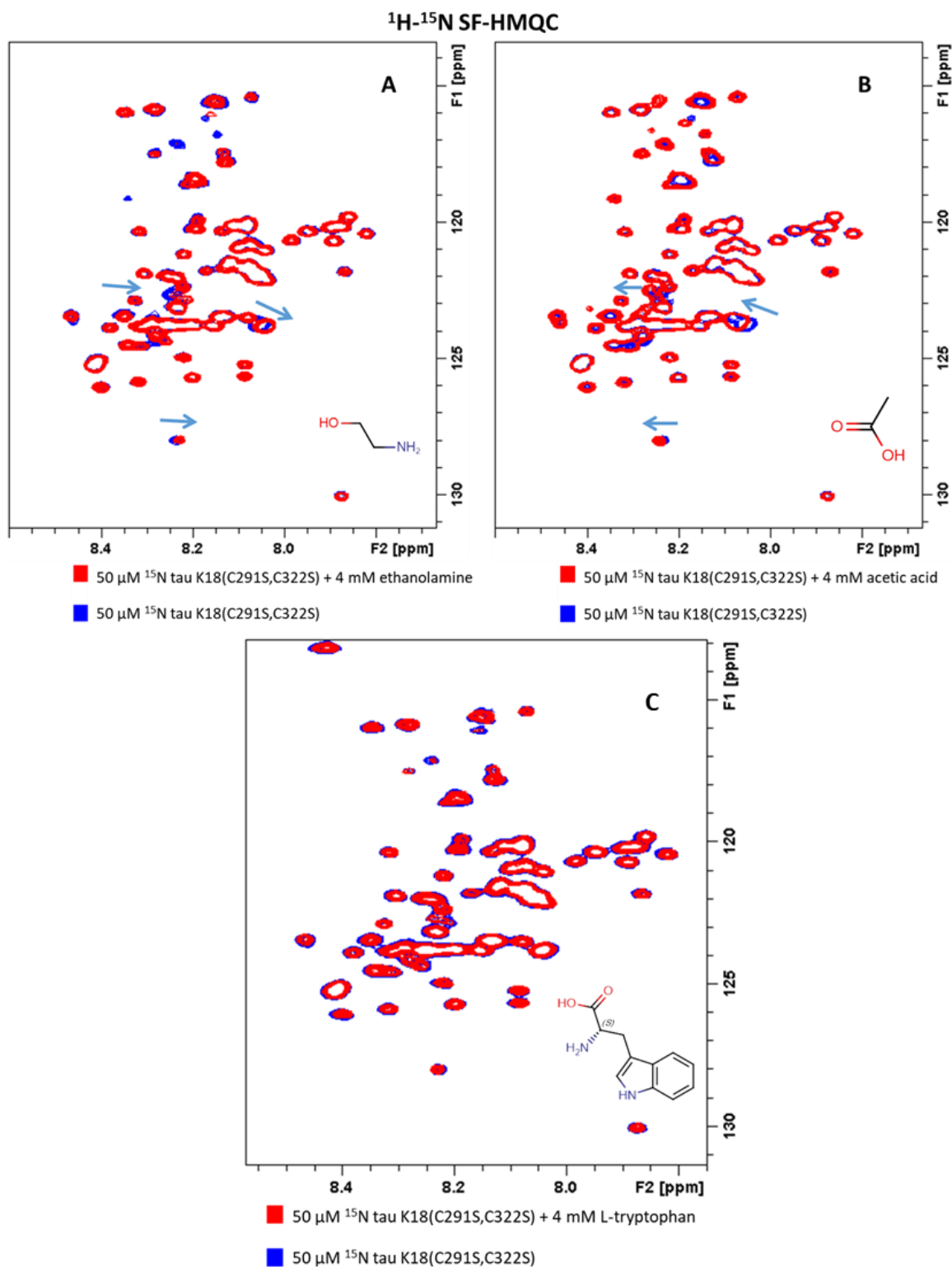
Figure 118. *Observed CSPs for ^{15}N-labelled 6xHis-tau K18(C291S,C322S) in the presence of ethanolamine (**A**), acetic acid (**B**) and L-Trp (**C**). The blue arrows mark the direction of cross peak shifts.*

Compound **6** was further investigated with 1D NMR techniques in order to characterize the interaction profile against tau K18(C291S,C322S).

5.5   Detection of ligand binding to tau K18 via ligand-observed NMR

2D NMR experimental results were followed up by 1D ligand-observed NMR techniques, such as ^{1}H STD, wLOGSY and CPMG-filtered (**§7.4.2**). Only compound **6** was tested using 1D NMR due to it being the only ligand that produced CSPs in 2D SF-HMQC spectra (**Figure 117**).

While initial ^{1}H STD and CPMG-filtered data did not indicate any interactions, wLOGSY data was consistent with IDP:ligand interactions between 6xHis-tau K18(C291S,C322S) and compound **6** (**Figure 119**). The STD technique depends on signal transfer through the structure of the protein onto the bound ligand, whereas wLOGSY signal depends on signal transfer via water molecules [85], [255]. Due to 6xHis-tau K18(C291S,C322S) being an IDP, the signal transfer onto the bound ligand may not be possible to detect in STD experiments due to lack of protein structure, whereas signal transfer via water molecules may be possible. The CPMG-filtered experimental results depend on T_2 relaxation rate of the ligand which is affected when a ligand interacts with protein [87]. Compound **6** was noted to exhibit an increased T_2 relaxation rate, where compound-specific signals broaden significantly after CPMG filter of 50 and 100 ms even without protein present (**Figure 119**). It has been previously observed that some compounds form aggregates in solution and NMR data can potentially indicate this phenomenon [132], [256]. It has also been shown that detergent-like molecules can recover signals of the fast-relaxing molecules in CPMG-filtered experiments by disrupting the soluble aggregates. For compound **6**, however, the addition of Tween-20 did not show any signal recovery for compound-specific signals (**Figure 120**). This suggested that the increased T_2 relaxation rate may not be entirely attributed to the aggregate formation, but rather to internal compound dynamics.
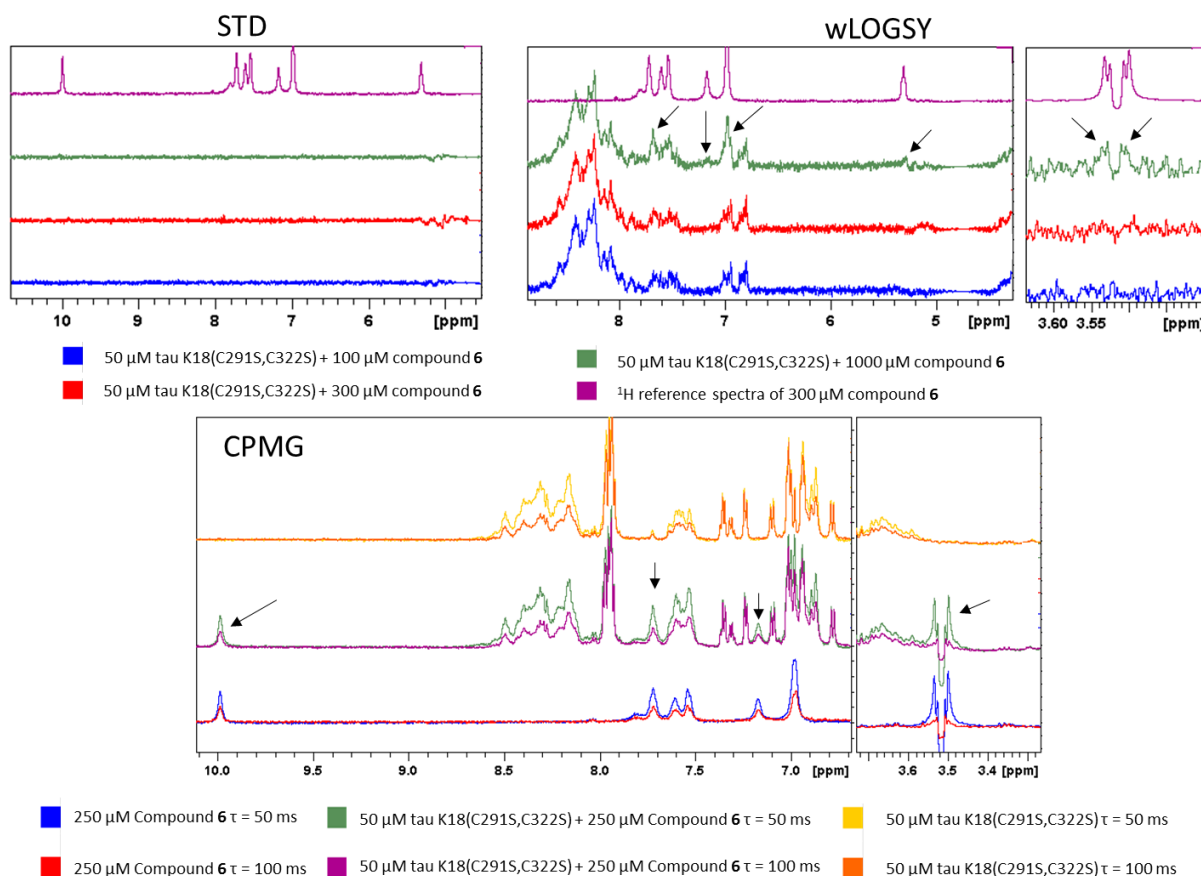
Figure 119. *Comparison of 1D ligand-observed NMR spectra in the presence and absence of 6xHis-tau K18(C291S,C322S) and compound **6**. STD and CPMG-filtered data do not indicate any IDP:ligand interactions, whereas wLOGSY does. The black arrows mark ligand-specific signals.*
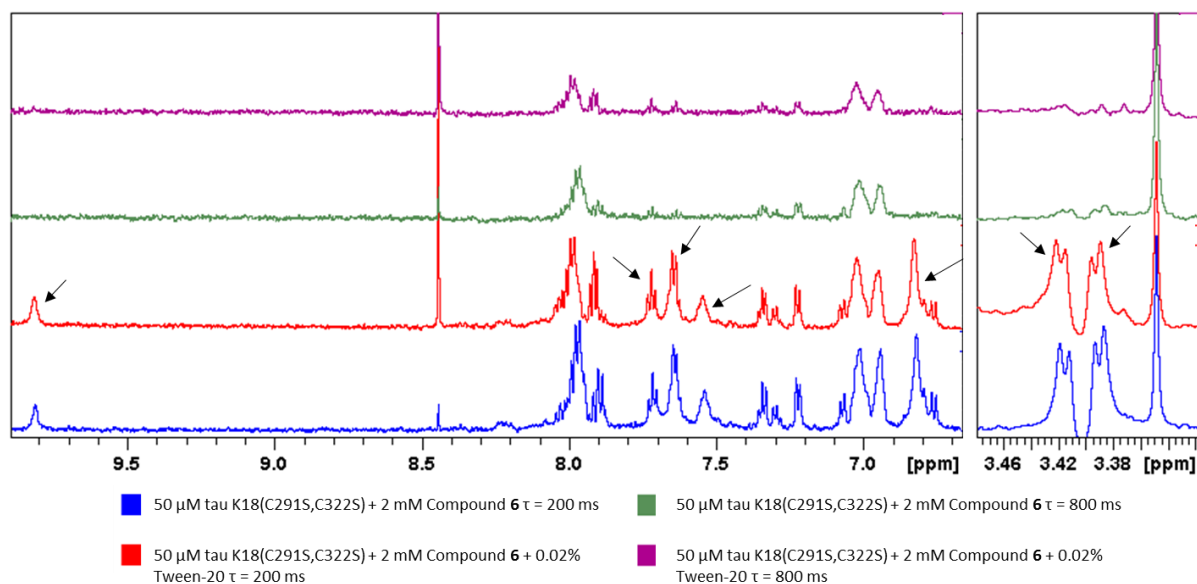
Figure 120. *1D ¹H CPMG-filtered NMR spectra of compound **6** with 6xHis-tau K18(C291S,C322S) in the presence and absence of Tween-20 at different delays. The black arrows mark ligand-specific signals.*

Due to different experimental design of SPR and MST techniques, it was possible to assess all ten literature-reported compounds against tau K18(C291S,C322S) at lower than 125 µM

concentrations in order to identify any potential binders that NMR approach may have missed.

5.6   Detection of ligand binding to tau K18 via SPR and MST

The literature-reported ligand interactions with tau K18(C291S,C322S) were assessed by SPR technique using biotinylated tau K18(C291S,C322S) construct (**pDV6**) (**§7.4.4**). Biotinylated AviTag-tau K18(C291S,C322S) was immobilized on SA chip and titration experiments were performed using single-cycle approach on Biacore T200. The concentrations used were based on publications and previous NMR solubility experiments. It was determined that compound **1**, compound **2** and compound **6** were interacting with biotinylated AviTag-tau K18(C291S,C322S) in dose-dependent manner (**Figure 121**). Compound **1** exhibited saturation profile to the protein with K_D of 39 µM (1:1 binding model). It was also noted that the sensorgrams were not stable during injection time suggesting specific events on the surface of the chip, such as conformational changes, rebinding of the analyte or precipitation of the compound. It was decided to remove the response of 300 µM compound **1** to the protein due to unstable sensorgram shape. Compound **2** was determined to interact with AviTag-tau K18(C291S,C322S) with a non-saturating profile as the steady-state responses did not reach saturation which was limited by compound solubility and slow dissociation behaviour. The same observation was noted for compound **6** in addition to a possible formation of aggregates or non-specific binding at concentrations above 2 mM.
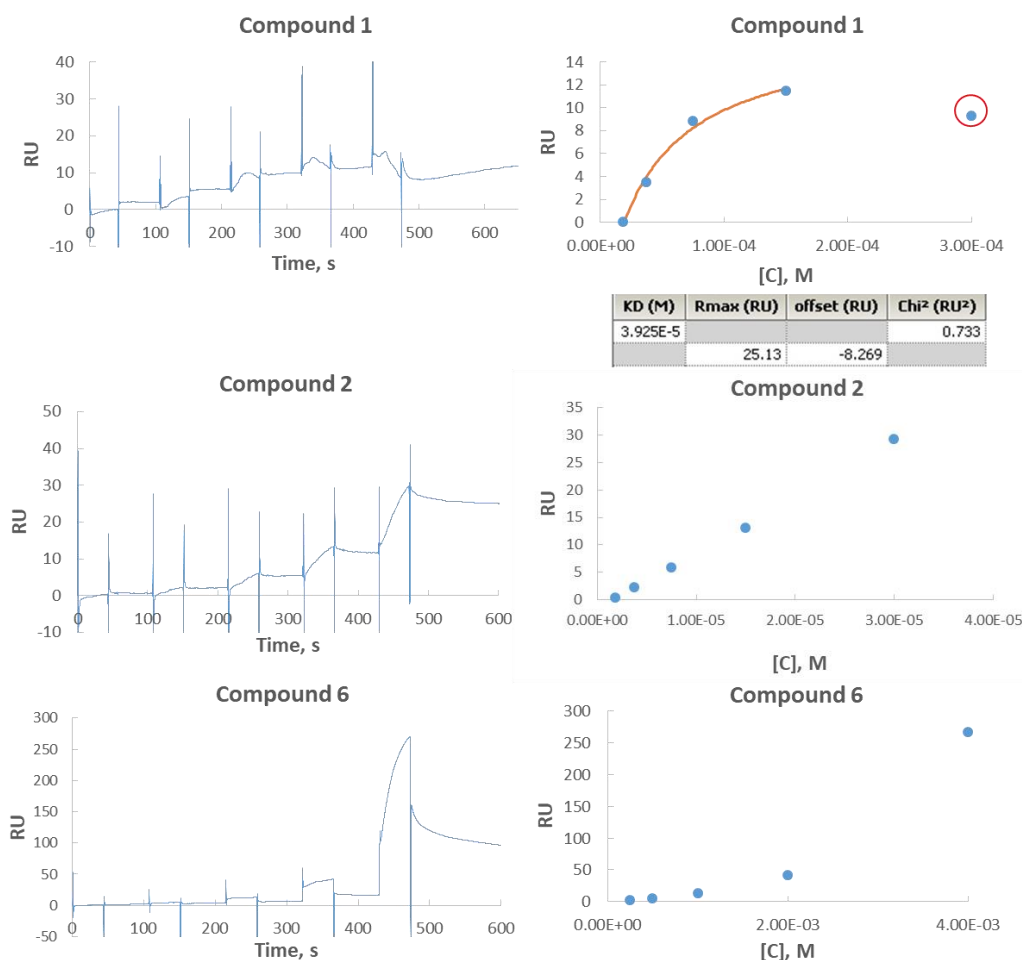
Figure 121. *Binding curves and steady-state affinity graphs for compounds 1, 2 and 6. Immobilization level: ~1000 RU of biotinylated AviTag-tau K18(C291S,C322S). Buffer solution: 50 mM HEPES pH 7.4, 150 mM NaCl, 0.05% Tween-20, 4% DMSO.*

It is possible that compound **6** may have affected the conformation of tau K18(C291S,C322S) or bound to the protein in non-stoichiometric manner as the concentration of the ligand was increased. The SPR angle shifts and the observed responses have been previously shown to be sensitive to changes of pH, ionic strength and conformation of the immobilized protein [257], [258]. While immobilized folded proteins can maintain their rigidity and interact with ligands as expected, the structure of IDPs may change upon interaction with ligand.

As an orthogonal method to NMR and SPR, MST was used to assess potential IDP:ligand interactions with literature-reported molecules (**§7.4.3**). Due to the presence of 6xHis tag in 6xHis-tau K18(C291S,C322S), the protein was labelled with RED-Tris-NTA dye which was determined to have the K_D of 14 +/- 2 nM to the protein (**Figure 122**).
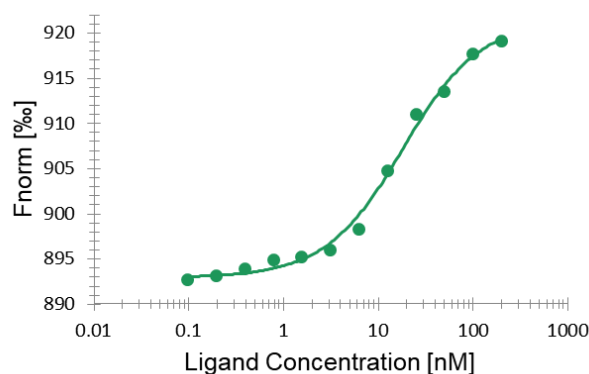
Figure 122. A graphical representation of MST data fit for RED-Tris-NTA dye interaction with 6xHis-tau K18(C291S,C322S).

The follow-up MST titration data suggested that compound **6** interacted with dye-labelled 6xHis-tau K18(C291S,C322S) in dose-dependent manner, as previously observed by NMR and SPR (**Figure 123, A**). However, additional control experiments showed that MST data of the dye was also affected by compound **6** without any protein present (**Figure 123, A**). In addition to this, initial fluorescence was also observed to be affected as concentration of compound **6** increased, indicating possible interactions between the dye and compound **6** (**Figure 123, B**). Previous investigation has reported that tryptophan-like molecules can quench the fluorescence of multiple organic dyes [259]. Compound **6** contains a Trp moiety which may have affected the observed fluorescence data, as shown with experiments in the presence of dye only. However, a larger fluorescence amplitude for labelled tau K18 (C291S,C322S):compound **6** complex than for dye:compound **6** complex indicated that there may be some interaction specificity and potential conformational changes to the protein in the presence of compound **6**. The affinity of compound **6** could not be reliably determined because the MST and initial fluorescence data did not reach saturation.
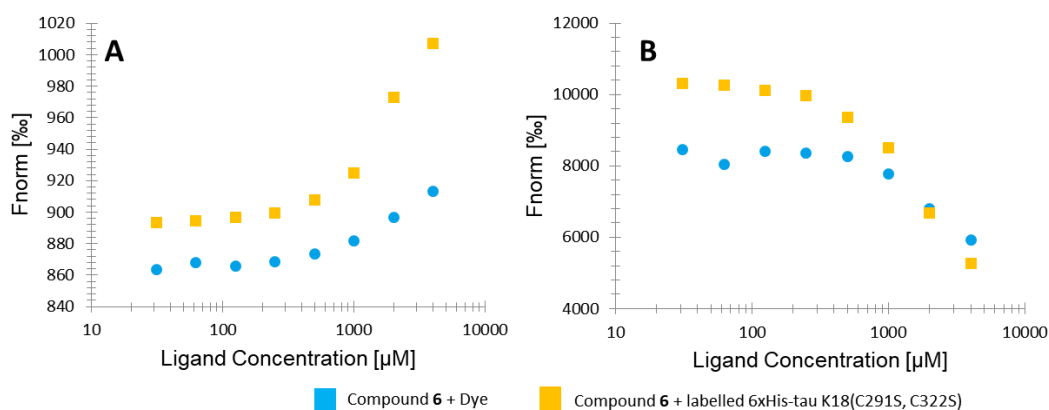
*Figure 123. A graphical representation of MST (**A**) and initial fluorescence (**B**) data for compound **6** in the presence and absence of 6xHis-tau K18(C291S,C322S).*

Compound **5** was also identified to interact with the labelled protein with K_D of 73 +/- 23 μM (1:1 binding model). The initial fluorescence was also observed to be affected by the compound. Interestingly, the initial fluorescence increased rather than decreased, which was in contrast to the data of compound **6** (**Figure 123, Figure 124**). Due to flexibility of IDPs, it is possible that 6xHis-tau K18(C291S,C322S) assumes a specific conformation which directly affects the RED-Tris-NTA dye. The dye then may become less or more shielded from the solvent which changes the fluorescence profile of the dye.
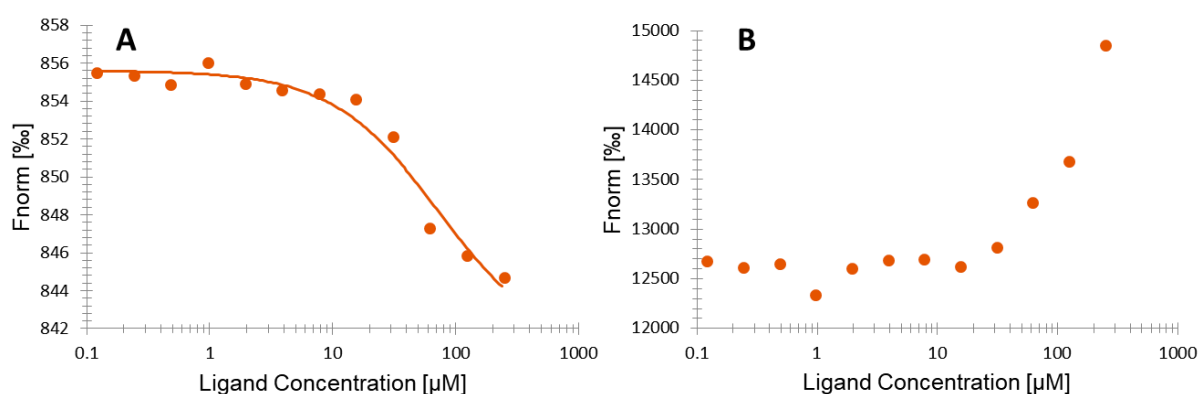
Figure 124. *A graphical representation of MST (**A**) and initial fluorescence (**B**) data for compound **5** in the presence and absence of 6xHis-tau K18(C291S,C322S).*

The overlap of intrinsic fluorescence of compound **1** and RED-Tris-NTA dye at λ = 600-650 nm prevented the direct labelling of the protein with the dye. However, this observation enabled a label-free approach. 6xHis-tau K18(C291S,C322S) was titrated from 5.9 to 2000 nM onto 100 nM of compound **1** and EC_{50} of 50 +/- 4 nM was determined with Hill coefficient of 4 (**Figure 125**). The data could not be fitted to a standard 1:1 model suggesting that multiple molecules may interact with the protein at the same time. Previously, it has been reported that compound **1** interacted with full length native tau and native tau K18(C291,C322) with nM affinity [242]. Additional molecular modelling suggested that the ligand could be binding near C291 and C322 residues. The tau K18 protein in this project had its cysteines mutated to serines (C291S,C322S) for stability purposes. The MST results indicated that the interactions between 6xHis-tau K18(C291S,C322S) and compound **1** were not largely affected by changing Cys to Ser, assuming that the previously proposed interaction mode model was correct.

It is possible that by labelling 6xHis-tau K18(C291S,C322S) randomly with covalent dyes, the results may be different due to the dye being potentially closer to the binding site of the ligands.

By comparing NMR, SPR and MST results it can be concluded that different biophysical techniques can provide different results and it is important to identify what possible artefacts may occur. While compound **1** did not produce any CSPs in NMR up to 500 µM (**§5.4**), SPR and MST indicated K_D in µM and nM range. If the calculated K_D values were in nM-µM range, protein-observed 2D NMR would have detected such potent binders to 6xHis-tau K18(C291S,C322S). It is possible that not all biophysical techniques detect true binding events for monomeric tau K18(C291S,C322S) or IDPs in general due to specific experimental conditions required, including labelling with a dye or immobilization on the surface.

5.7   Detailed characterization of compound 6 binding to tau K18

Following the protein- and ligand-observed NMR results, the interactions between compound **6** and tau K18(C291S,C322S) were further characterized using additional NMR and other biophysical methods.

Compound **6** contains a quinone group which is a potential Michael acceptor. Michael acceptors are known to react with nucleophilic groups present in proteins and form covalent adducts [260]. As mentioned previously, 6xHis-tau K18(C291S,C322S) had its two native Cys residues changed to Ser (C291S,C322S) for stability purposes. The HPLC-MS studies were performed in order to evaluate any possible covalent adduct formation. The studies suggested that no covalent adducts were present after 4 hours of incubation of 6xHis-tau

K18(C291S,C322S) and compound **6** at RT (**Figure 126**). In addition to this, due to previous reports of compound **6** preventing aggregation of native tau K18 which contained two Cys, compound **6** was also incubated in the presence of DTT. The HPLC-MS and NMR data was also consistent with no reactivity of the compound to thiol groups (**Figure 127**).
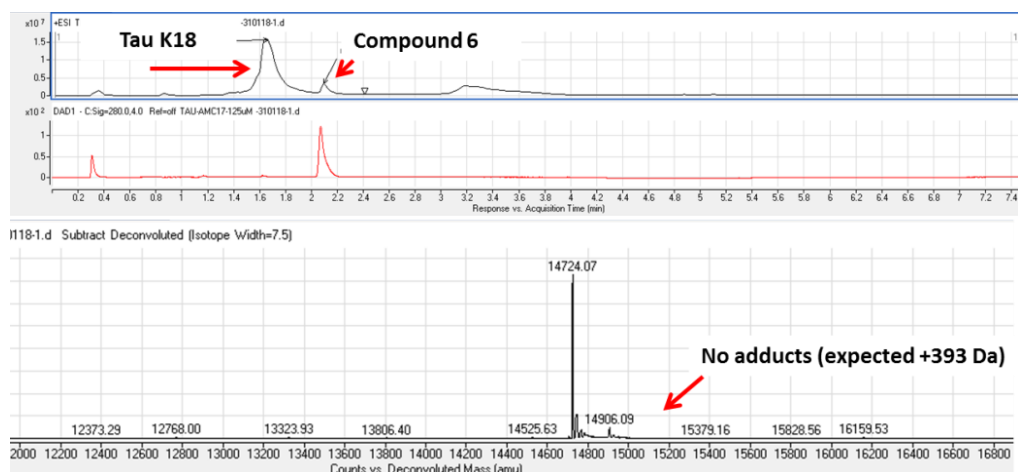
*Figure 126. HPLC-MS data for 50 µM 6xHis-tau K18(C291S,C322S) with 1 mM compound **6**. Incubation time: 4 h.*
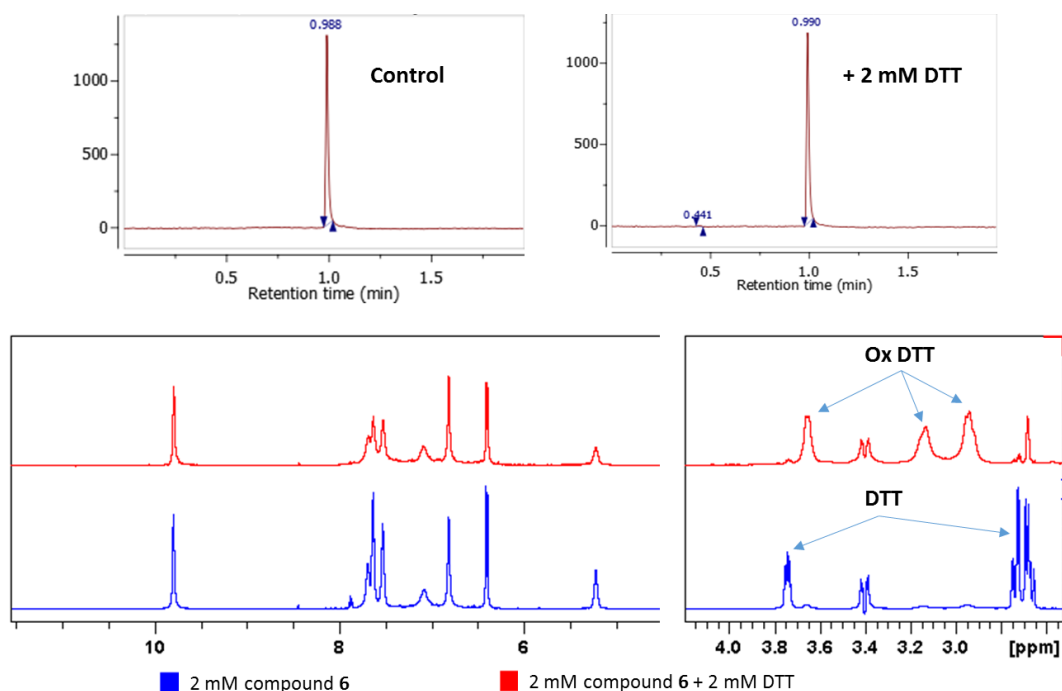
*Figure 127. HPLC-MS and ¹H NMR data of compound **6** in the presence and absence of 2 mM DTT. Incubation time: 72 h.*

To further characterize binding interactions, simple one-step organic synthesis approach was used to produce more of the parent compound **6** and several near neighbours (NNs) (**Figure 128**) (**§7.5**):

- Resynthesized compound **6** (*S*-enantiomer) **(6A)**

- *R*-enantiomer of compound **6**

- Ethyl ester of compound **6**

- N-methyl indole (**6B**)

- 5-methoxy (**6C**)

- 5-hydroxy (**6D**)

- 5-fluoro (**6E**)

- 7-methyl (**6F**)

- 5-chloro (**6G**)

- 5-cyano (**6H**)
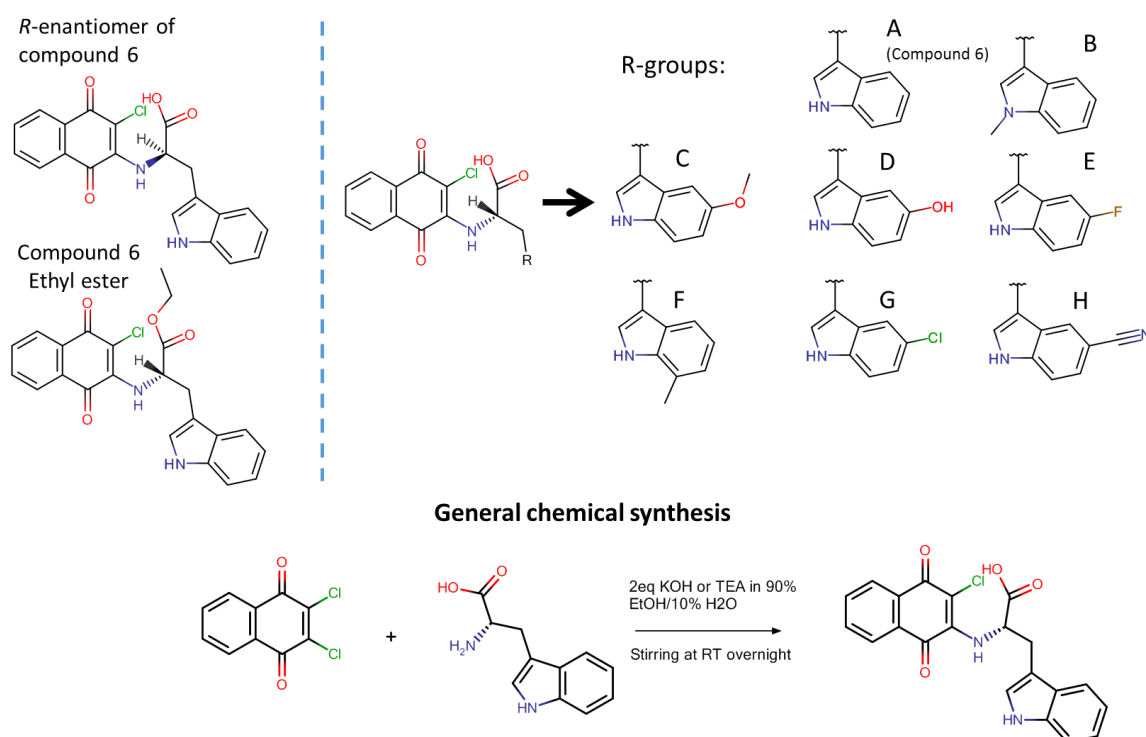
Figure 128. *The chemical structures of compound 6, its derivatives and a representation of chemical synthesis scheme.*

The R enantiomer and most of NNs of original S-enantiomer (compound **6A**) were determined to be soluble up to 4 mM in aqueous solutions and produced similar CSPs for ^{15}N-labelled 6xHis-tau K18(C291S,C322S) (**Figure 129**). This suggested that stereochemistry may not be important for the observed IDP:ligand interactions. The esterification of carboxylic acid decreased the solubility to below 50 μM and made evaluation of the interactions by NMR not possible. 2D ^{1}H-^{13}C HSQC NMR has identified CSPs in the presence of compound **6D** for aromatic carbons of Y310, F346 and Cε1 for His residues (**Figure 130**).
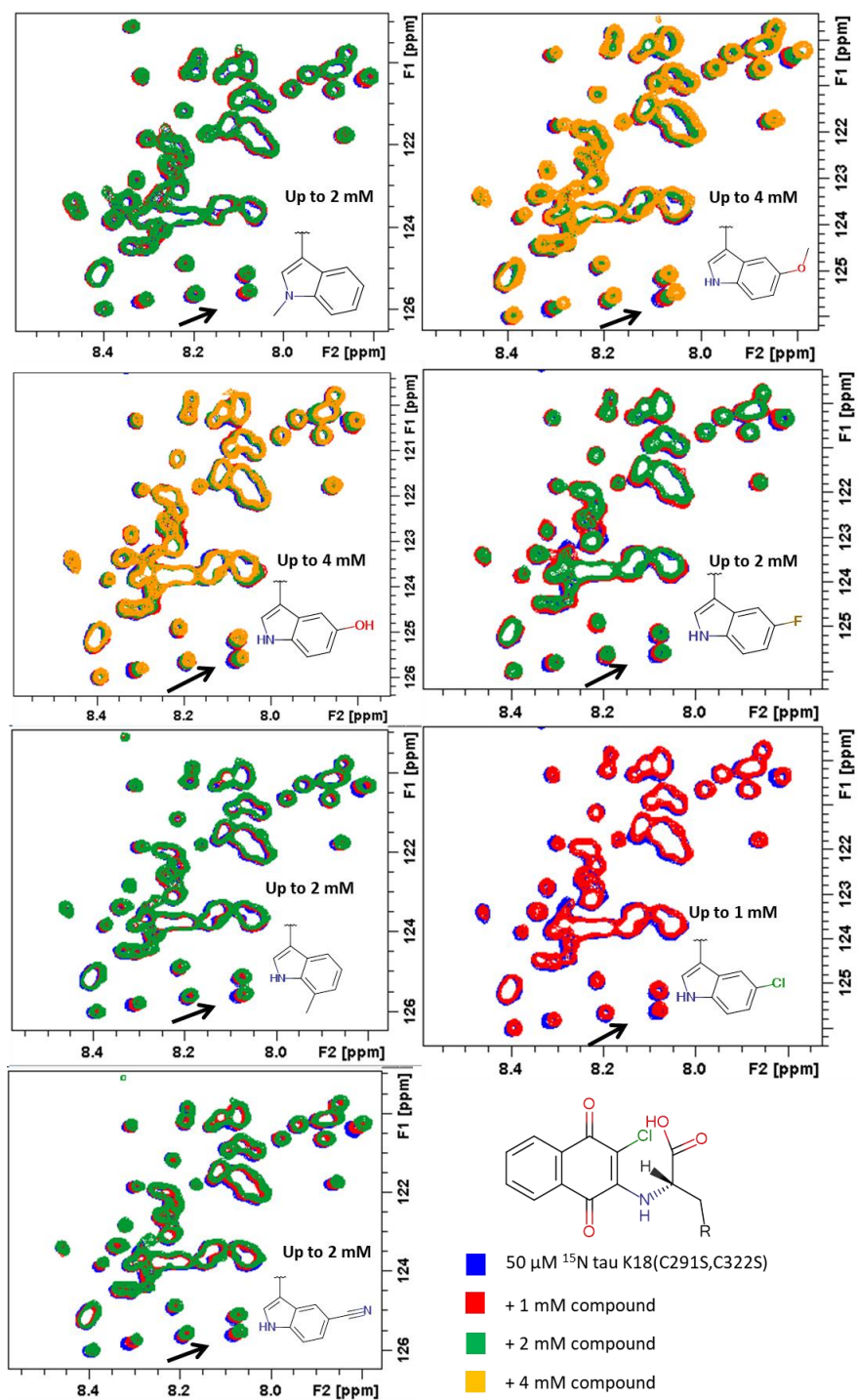
Figure 129. *2D ^1H-^{15}N SF-HMQC spectra in the presence of different concentrations of NNs with ^{15}N-labelled 50 µM 6xHis-tau K18(C291S,C322S). The black arrows indicate the observed vector of CSPs.*

The observed Y310 CSPs in 2D ^1H-^{15}N SF-HMQC agreed with ^1H-^{13}C HSQC data, whereas F346 cross peak was difficult to interpret due to spectral overlap. The CSPs were also noted to be similar for compound **6D** at higher salt concentrations, suggesting non-charge-based interactions (**Figure 131**).
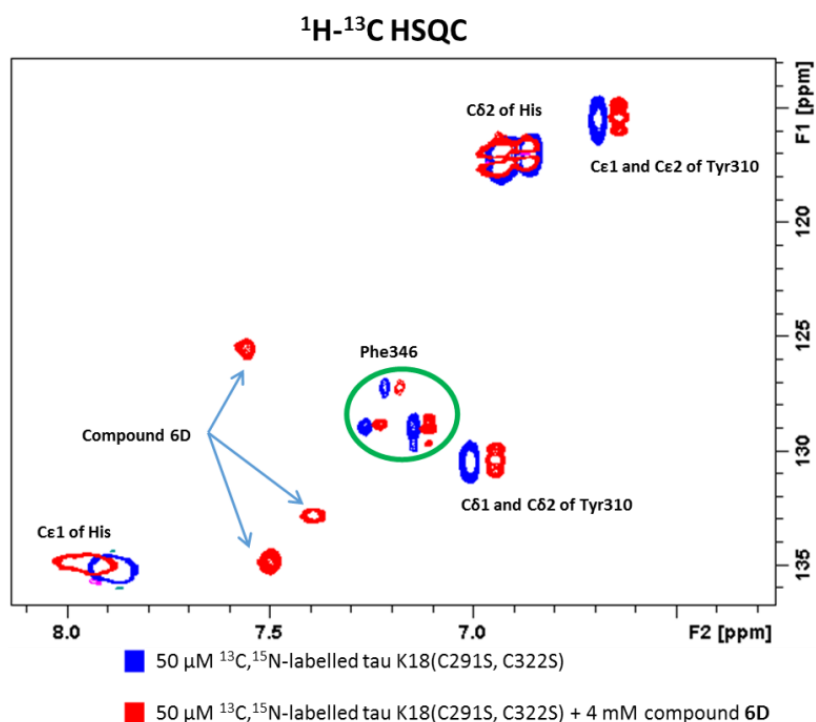
Figure 130. *2D ¹H-¹³C HSQC spectra of ¹³C,¹⁵N-labelled 6xHis-tau K18(C291S,C322S) in the presence and absence of compound **6D**.*
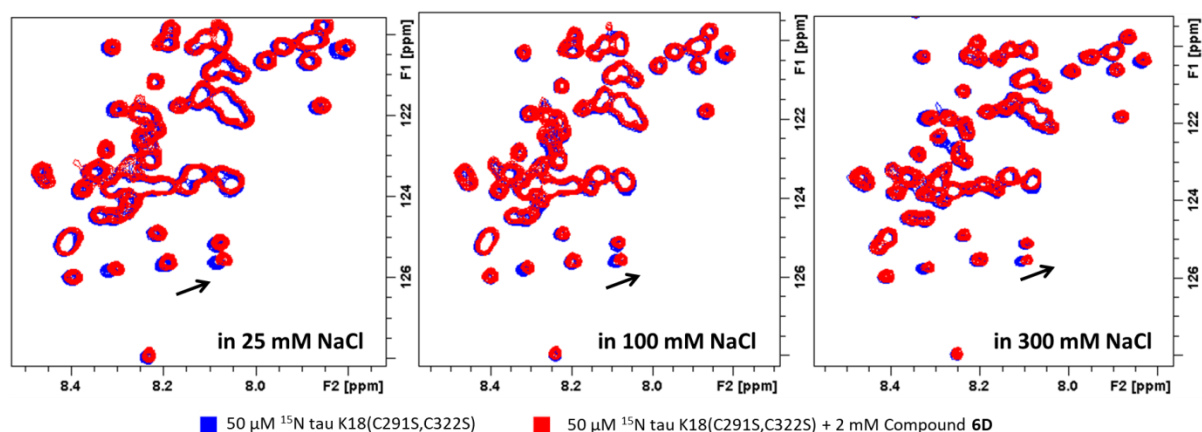
Figure 131. *2D ¹H-¹⁵N SF-HMQC spectra of ¹⁵N 6xHis-tau K18(C291S,C322S) in the presence of 2 mM compound **6D** at different concentrations of NaCl. The black arrows denote the vector of observed CSPs.*

It is important to note that some of the NNs of compound **6A** caused 6xHis-tau K18(C291S,C322S) to precipitate out of solution at concentrations above 1 mM, as determined by ¹H NMR spectra (**Figure 129**, **Figure 132**). Visual inspection of the samples indicated dark red particles in the solution. The precipitation phenomenon was especially noticeable for fluorinated and chlorinated NNs (**6E** and **6G**). Halogen substitutions are known to provide additional protein-ligand interactions and increase the lipophilicity of the parent molecule [261]. It is possible that the halogenated ligands interacted with the protein at higher concentrations in a manner that caused the protein to precipitate out of solution. The

decrease of the compound-specific resonance peaks indicated that compounds have also precipitated out of solution together with the protein.
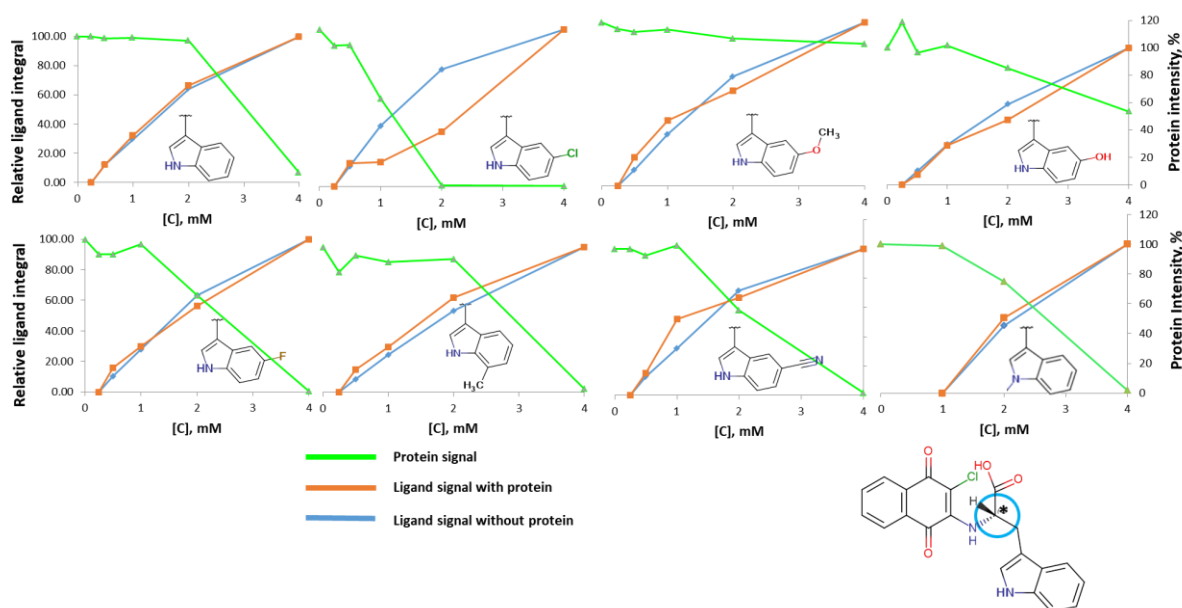
*Figure 132. A comparison of 6xHis-tau K18(C291S,C322S) solubility profile in the presence of compound **6** and its NNs. The well-resolved C-H proton (**circled in blue**) and protein methyl envelope (0.76-1 ppm) were used to track ligand and 6xHis-tau K18(C291S,C322S) signals, respectively.*

After this observation, it was hypothesized that compound **6A** and its NNs may affect the diffusion rate of 6xHis-tau K18(C291S,C322S) before the observed precipitation phenomenon occurred. Protein-observed NMR DOSY indicated that the diffusion rate of 6xHis-tau K18(C291S,C322S) was not significantly affected by the presence of compound **6A** or some of its NNs (**Figure 133**). Some diffusion rates for the protein in the presence of several NNs of compound **6A** could not be determined above certain concentration due to precipitation of the protein. It is possible that any conformational changes that are happening may be too small to have a large effect on the diffusion rate of 6xHis-tau K18(C291S,C322S).
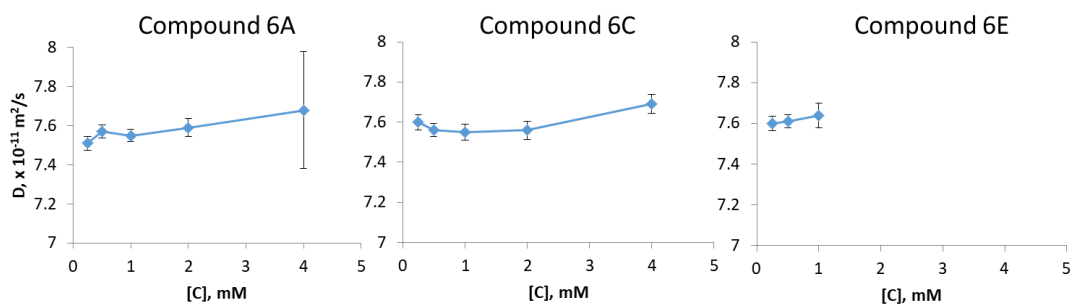
*Figure 133. The diffusion rate changes for 6xHis-tau K18(C291S,C322S) in the presence of different concentrations of compound **6A, 6C** and **6E**.*

It has been previously reported that denatured globular or folded proteins follow a power-law dependency which allows to estimate radius of gyration R_G [262]. In turn, this relation allowed estimating the diffusion rates of the denatured globular protein using Einstein-Stokes relation, which agreed with the experimentally determined diffusion values for the same proteins. This relation has also shown that the shortening of sequence by 20 amino acids would change the R_G by 1.5-2 Å or diffusion rate by 10%. If it is assumed that 20 amino acids become structured for an IDP upon LMW ligand binding, the expected diffusion rate change would potentially be lower than 10%. Therefore, it can be expected than any increased local structural integrity of an IDP upon ligand binding may not be evident in NMR DOSY, if power-law dependency is assumed. However, it has also been suggested that IDPs may have specific physics in solution in comparison to globular proteins and power-law cannot be reliably applied to predict IDP behaviour due to their intrinsic flexibility and a more in-depth investigation is necessary.

As mentioned previously, the usage of polysorbate 20 (Tween-20) above its critical micelle concentration (CMC) did not change the observed behavior of the compound **6A** in CPMG-filtered data, suggesting that the ligand is not forming aggregates in the solution at higher concentrations (**Figure 120**). Lack of compound aggregation was also implied by ligand-observed NMR DOSY data which has shown that theoretical (SEGWE model) and experimental diffusion rates of compound **6A** and its NNs were comparable (**Table 14**) [256]. SEGWE model suggests that any aggregate-forming compounds may exhibit lower diffusion rates as they become a part of a larger, slower moving structure. The observed differences are possibly due to theoretical diffusion model using values for samples in pure D_2O rather than a specific buffer solution.

Table 14. A comparison of diffusion coefficients between theoretical (SEGWE) and NMR DOSY fitted values for 1 mM compound 6A and its NNs. Δ = 10 ms, δ = 8 ms.

	Diffusion coefficient $D \times 10^{-10}$, m^2/s		
Compound	Theoretical	Experimental	Difference, %
6A	3.9	4.48 +/- 0.09	15
6B	3.9	4.38 +/- 0.12	12
6D	3.9	4.42 +/- 0.1	13
6E	3.9	4.49 +/- 0.08	15

In order to further investigate binding interactions of compound **6A** with tau K18(C291S,C322S), X-filtered ^{13}C HSQC-NOESY experiments were performed. These experiments provide data that can be used to unambiguously determine close contacts arising from a significant bound state of the ligand to the protein. The data can then be used to make an NMR-guided interaction models (NGMs) when crystallography approach is difficult or impossible [92].

The experimental results did not indicate any NOEs from compound **6C** to ^{15}N,^{13}C-labelled 6xHis-tau K18(C291S,C322S), probably due to the affinity of the ligand being too low (**Figure 134**). Several naphthoquinone cross peaks were observed at 7.54 and 7.64 ppm. However, the cross peaks did not connect to any other peaks of the protein and were determined to come from the ligand itself due to its high concentration in solution (4 mM). At this concentration, the sample contained large enough presence of naturally abundant ^{13}C (~1% of all carbon atoms) from the ligand itself to exhibit the signal.
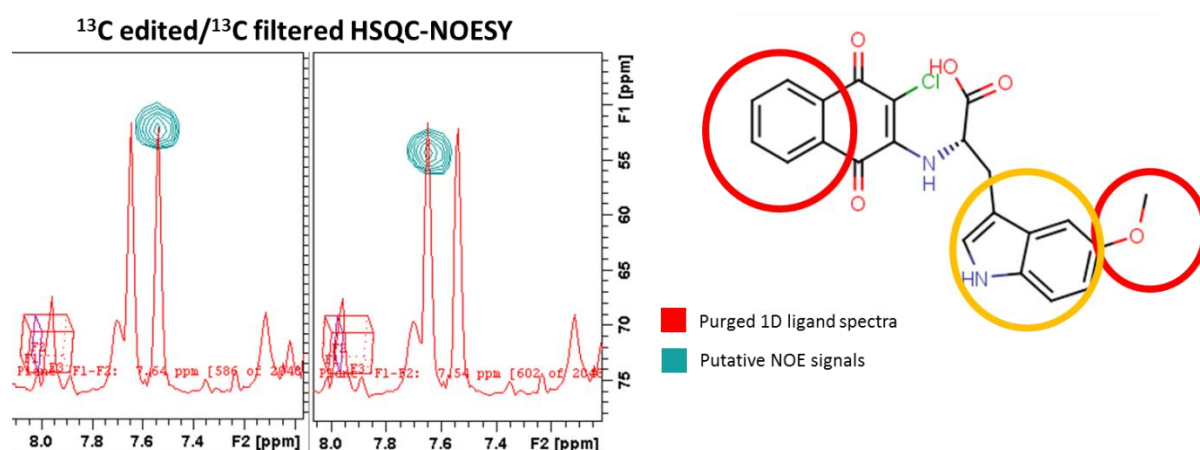
Figure 134. *^{13}C HSQC-NOESY ^{15}N,^{13}C 6xHis-tau K18(C291S,C322S) spectra in the presence of compound **6C**. (**Yellow** – ambiguous NOEs, **Red** – no NOEs observed)*

In addition to this, the presence of 6xHis tag prevented objective observations of any indole NOEs due to the signal "bleed-through" effect, where, because of the unusual one bond J coupling present in the imidazole ring of histidine, the signal is incompletely filtered by the compromise delay used in the pulse sequence (**Figure 135**). While J_{CH} value of 140 Hz is used to filter out non-interacting aliphatic and aromatic ^{13}C signals, imidazoles require J_{CH} value of around 200 Hz. Removal of 6xHis tag can prevent this artefact effect and allow unambiguous evaluation of signals from naphthoquinone or indole ring.
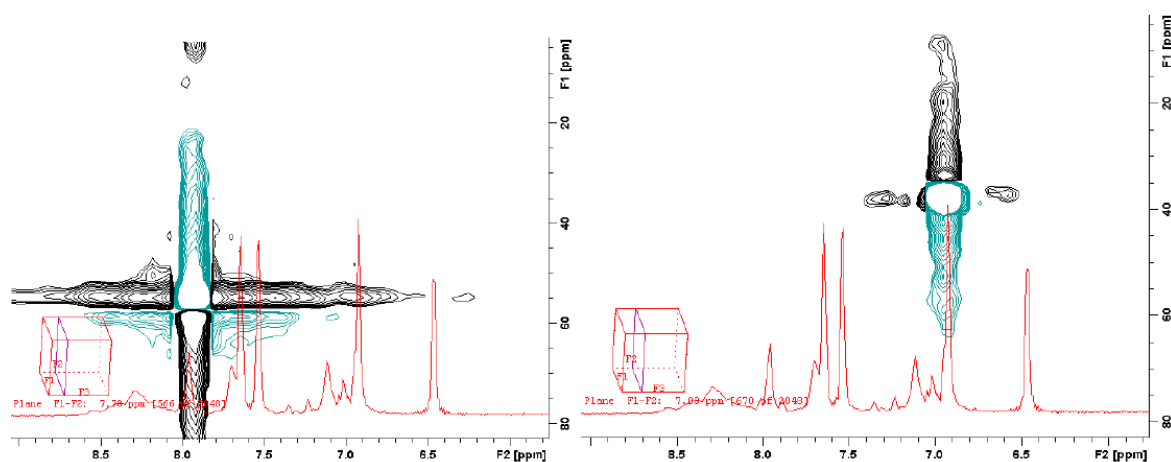
Figure 135. "Bleed-through" effect during X-filtered NOESY experiments with $^{15}N, ^{13}C$ 6xHis-tau K18(C291S,C322S).

Following this, ^{19}F NMR was used to probe the IDP:ligand interactions using previously expressed ^{19}F labelled 6xHis-tau K18(C291S,C322S, 3-fluoro-Y310) (**§5.2**). It has been determined that fluorine atom has an inherently increased sensitivity to its local environment [263]. The protein-observed ^{19}F NMR data have indicated only small CSPs in the presence of compound **6D** with a non-saturating dose response profile (**Figure 136**). The results also suggested that compound **6D** may not be interacting with the side chain of Y310 directly as the observed CSPs are relatively small.
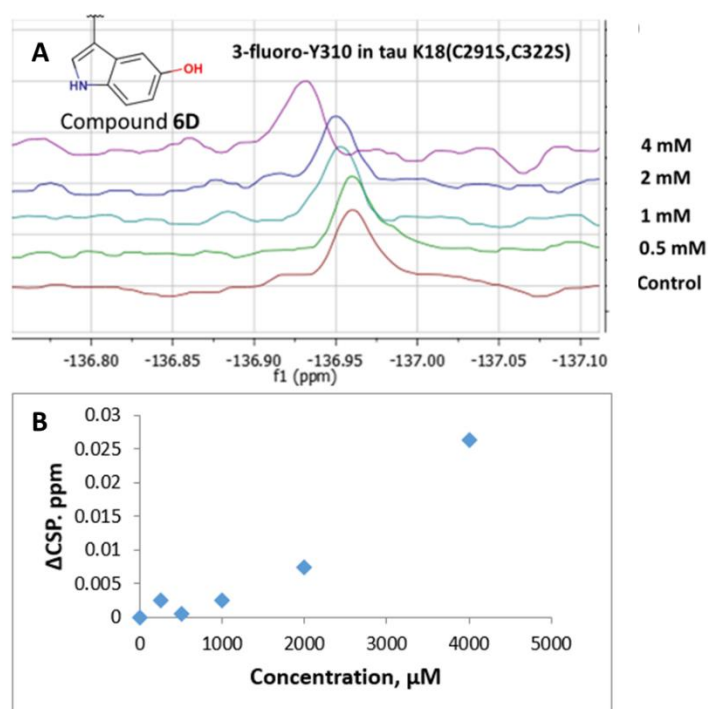
*Figure 136. ^{19}F protein-observed NMR spectra and observed CSPs for 6xHis-tau K18(C291S,C322S, 3-fluoro-Y310) in the presence of compound **6D**. (A) Spectral representation for ^{19}F-labelled 6xHis-tau K18(C291S,C322S, 3-fluoro-Y310) in the presence of different concentrations of compound **6D**. (B) Graphical representation of ΔCSPs for 6xHis-tau K18(C291S,C322S, 3-fluoro-Y310) at different concentrations of compound **6D**.*

Following this, additional ligand-observed NMR approaches were performed to better characterize the behaviour of the ligands in the absence and presence of 6xHis-tau K18(C291S,C322S) protein. ^{19}F CPMG-filtered ligand-observed experiments showed similar ^{19}F T$_2$ relaxation values for compound **6E** in the absence and presence of 6xHis-tau K18(C291S,C322S) (**Figure 137, Table 15**). This suggested that compound **6E** did not interact with the target protein in a detectable manner using ^{19}F CPMG-filtered experiments. In addition to this, the broad ^{19}F peak width of ~1 ppm suggests internal compound dynamics or possible aggregation, possibly due to chemical exchange events.
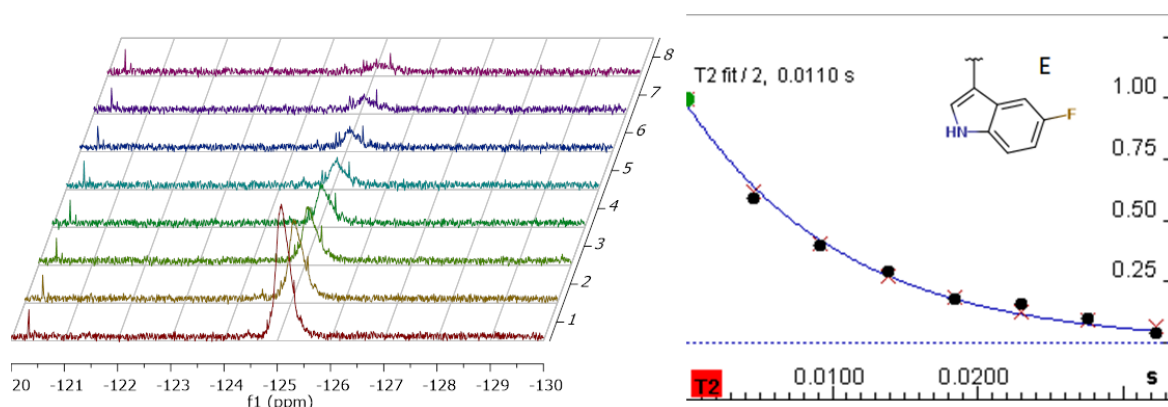
Figure 137. ^{19}F CPMG-filtered spectra of 1 mM compound *6E* at various τ delays (*Left*). Data fitting of ^{19}F T$_2$ relaxation data for compound *6E* (*Right*).

Table 15. ^{19}F T$_2$ values for compound *6E* in the presence and absence of 50 μM 6xHis-tau K18(C291S, C322S).

Ligand concentration, mM	^{19}F T$_2$ value without protein, ms	^{19}F T$_2$ value with protein, ms
0.1	9.3 +/- 0.1	9.4 +/- 0.3
0.3	9.9 +/- 0.2	9.1 +/- 0.1
1	10.9 +/- 0.1	11.2 +/- 0.1

Temperature gradient experiments for compound **6D** have shown possible chemical exchange which may explain the increased T$_2$ relaxation rate observed earlier (**Figure 119, Figure 138**). The chemical exchange may be occurring between stacked and non-stacked ring conformations of the ligand which causes changing chemical environment in the NMR time scale. This can be observed as some peaks coalesce, split and sharpen again with changing temperature (**Figure 138**). Similar behaviour was previously observed for DNA oligos which would dimerize or melt at different temperatures which perturbed the chemical environment of specific protons in DNA bases [264]. Multiple potential conformations of compound **6D** were generated using MarvinSketch (**Figure 139**).
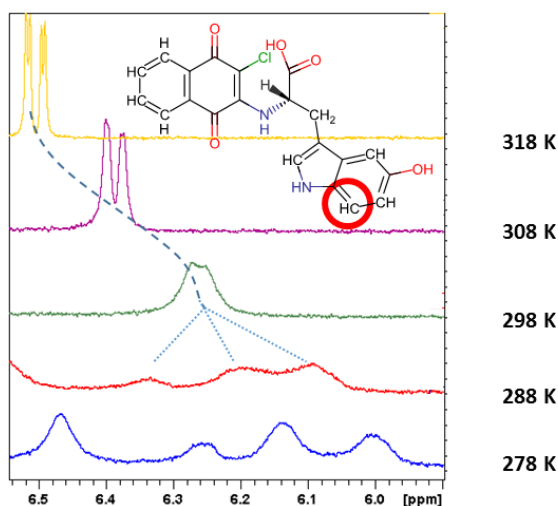
Figure 138. ^{1}H NMR spectral changes for marked proton (red circle) of compound *6D* at different temperatures.
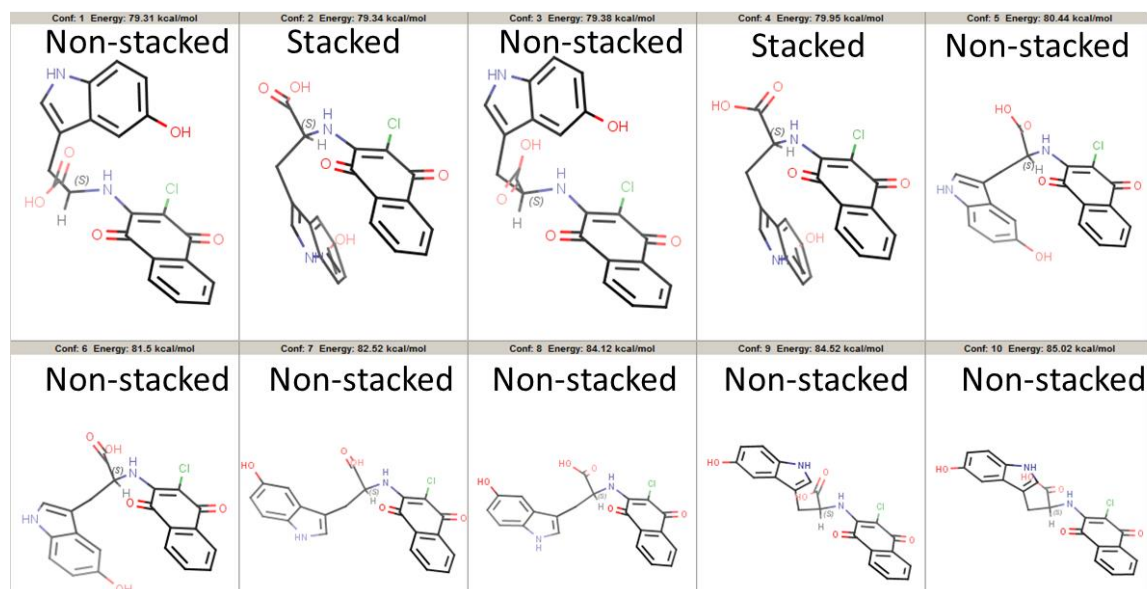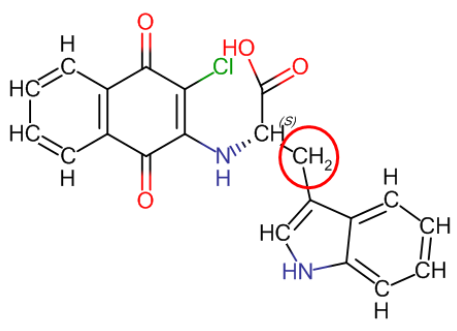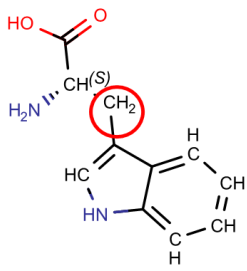
Figure 139. Stacked and non-stacked conformations of compound *6D* (MarvinSketch).

The results showed that compound **6D** exists as a conformational ensemble where the structure interchanges between stacked and non-stacked conformations. Additional molecular dynamic simulations also confirmed this observation (J. Davidson, personal communication).

As a follow-up investigation, T_2 relaxation values for easily distinguishable tryptophan methylene group of multiple NNs were determined using ^{1}H CPMG-filtered NMR experiments at different concentrations with and without protein (**Table 16**). The T_2 values were similar for ligands at 100 and 1000 μM which are indicative of intramolecular rather than intermolecular interactions. Interestingly, T_2 values have decreased in the presence of 6xHis-tau K18(C291S,C322S) which is consistent with IDP:ligand interactions. This indication confirmed

previous observations by wLOGSY and 2D SF-HMQC (**Figure 117, Figure 119**). The T_2 values for the methylene group of L-Trp were determined to be one order of magnitude higher than for compound **6.** This again indicated that the increased T_2 relaxation times for compound **6** and its NNs were caused by intramolecular rather than intermolecular interactions.

Table 16. *T_2 relaxation values for tryptophan methylene group at different concentrations of compound **6** and its NNs in the presence or absence of 6xHis-tau K18(C291S,C322S). L-Trp T_2 values were measured without any protein present.*

Compound	T2 value for tryptophan methylene group, ms			
	Without protein		With protein	
	100 µM	1000 µM	100 µM	1000 µM
6A	50 +/- 10	39 +/- 12	9 +/- 2	18 +/- 6
6B	41 +/- 6	33 +/- 8	13 +/- 4	14 +/- 2
6D	36 +/- 6	26 +/- 5	4 +/- 1	13 +/- 2
6E	24 +/- 3	22 +/- 3	8 +/- 2	12 +/- 2
6F	27 +/- 5	15 +/- 1	10 +/- 2	8 +/- 2
6G	34 +/- 6	34 +/- 5	13 +/- 2	18 +/- 5
6H	53 +/- 11	36 +/- 7	11 +/- 6	26 +/- 4

Compound	T2 value for tryptophan methylene group, ms	
	Without protein	
	100 µM	1000 µM
L-Trp	557+/- 84	559 +/- 63

After learning about possible artefacts and specific signs of interactions using SPR and NMR techniques, SPR fragment screening campaign was initiated in order to identify potential fragments that interact with 6xHis-tau K18(C291S,C322S) in a specific manner.

5.8 Fragment screening via SPR

Due to unavailability of biotinylated tau K18(C291S,C322S) construct, the screening campaign was performed using amine coupling immobilization approach (**§7.4.4, §7.4.4.2**).

5.8.1 Clean Screen (CS)

After immobilization and CS assay, 'sticky' fragments to 6xHis-tau K18(C291S,C322S) were removed in order to prevent false positive hits exhibiting non-specific, residual binding behaviour to empty or derivatized CM5 S Series chip surface, as described earlier (**Figure 97, Table 7**) (**§4.7.2**).

Table 17. The number of 'sticky' compounds identified for empty surface and tau K18(C291S,C322S).

Surface derivatization	Sticky compounds	Surface-specific sticky compounds
Empty	2	1
6xHis-Tau K18(C291S, C322S)	8	7

Two compounds were identified to exhibit slow dissociation rate from empty CM5 S Series chip surface. Of these, one was also interacting non-specifically to 6xHis-tau K18(C291S, C322S). Seven additional compounds were identified to have a residual binding profile specifically to 6xHis-tau K18 (C291S, C322S).
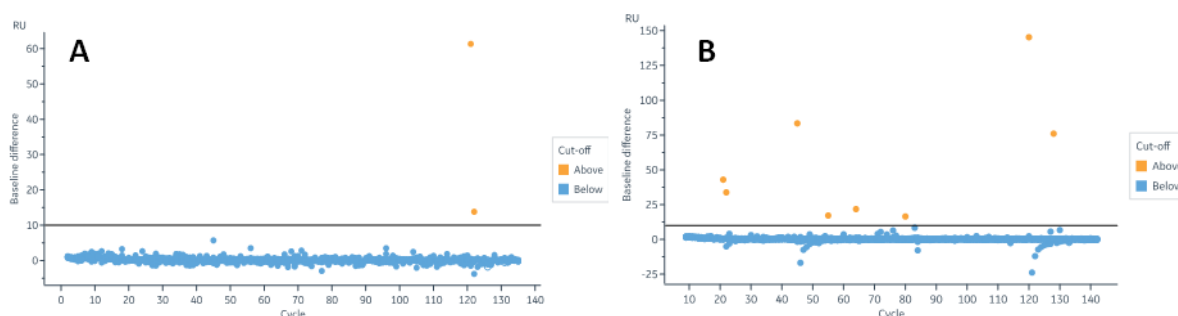
*Figure 140. Representation of N_i - N_{i-1} plots for different derivatizations of the chip surface. **A**. Empty Surface (0 RU). **B**. 6xHis-tau K18(C291S, C322S) (~2800 RU).*

5.8.2 Binding Level Screen (BLS)

After removing the 'sticky' compounds, the BLS assay was used to identify and prioritize hits without artefactual binding behavior markers using the same reasoning and procedure as for PTP1B constructs (**§4.7.3**) (**Figure 141**).
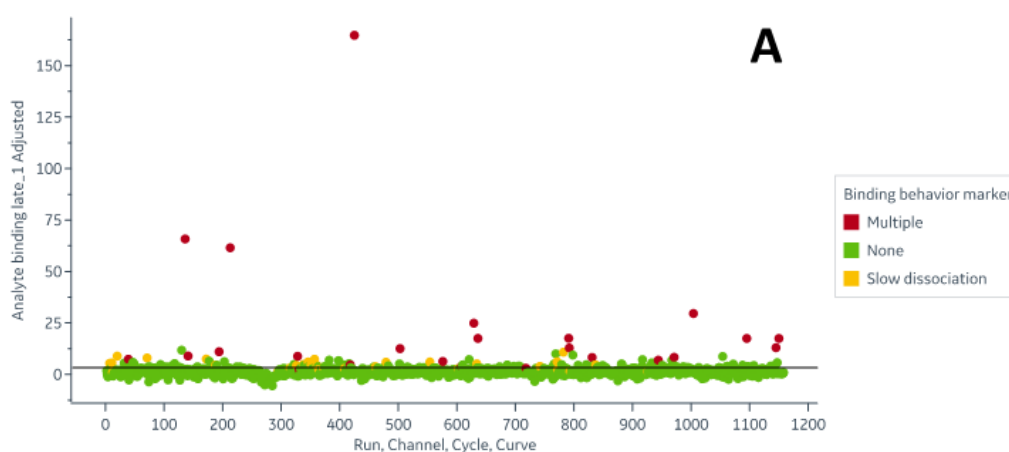
Figure 141. Identification of well-behaved fragments using BLS approach. The samples exhibiting artefactual binding behaviour markers were flagged with specific colours. Immobilization level of 6xHis-tau K18(C291S, C322S): ~3000 RU

The BLS step automatically prioritized 107 compounds without undesirable binding behaviour (**Table 18**). The selected number represented a standard 10% of the library to be used in the

subsequent AS step. After comparative data analysis with PTP1B data (**§4.7.3**) it was noted that out of 107 compounds, 93 were unique to 6xHis-tau K18(C291S, C322S) (**Table 18**). This indicated that 6xHis-tau K18(C291S,C322S) may have a different class of binders when compared to PTP1B constructs.

Table 18. Prioritization of 6xHis-tau K18(C291S, C322S) hits in comparison to PTP1B hits from automated BLS data evaluation.

Protein	BLS hits	Unique hits
6xHis-Tau K18(C291S, C322S)	107	93 out of 107 when compared to PTP1B constructs

Similarly to PTP1B constructs, the observed response levels for 6xHis-tau K18(C291S,C322S) were consistently lower than theoretically calculated R_{max} of 30 RU for 150 Da fragment with 1:1 interaction model. It is possible that low responses were due to only a fraction of immobilized 6xHis-tau K18(C291S,C322S) remaining functional on the surface. 6xHis-tau K18(C291S,C322S) has 20 Lys residues and it cannot be reliably determined which Lys residues participate in amine coupling to the surface of the CM5 S Series chip. Therefore, the generated surface could have exhibited a preferred conformation of 6xHis-tau K18(C291S,C322S), preventing any specific interactions with certain fragments. In addition to this, the carboxymethyl dextran matrix could have also altered the conformation of the immobilized 6xHis-tau K18(C291S,C322S) due to its intrinsic disorder. It was shown previously that depending on the pH of the solution, some proteins may prefer specific conformations once immobilized on the surface in comparison to when they are free in solution [257]. This observation may be especially relevant for IDPs.

5.8.3 Affinity Screen (AS)

After BLS, fragment screening was followed by AS step which identified multiple unique hits against 6xHis-tau K18(C291S,C322S) (**Table 19**). Comparative analysis also showed that only 3 of the 45 hits taken to AS were also identified for PTP1B constructs previously.

Table 19. The identified hits for 6xHis-tau K18(C291S,C322S) construct after AS step.

Protein	Hits	Number of unique hits	% of total library
6xHis-Tau K18(C291S, C322S)	45	42 (out of which 3 found for PTP1B)	4.2%

The majority of potential hits were determined to contain carboxylic acid or phenol moieties (35 out of 45) (**Figure 142, Figure 143, Figure 144**). The hydroxyl groups of phenols have theoretical pKa values close to the pH of the solution (pH = 7.4) which suggests that some of

the compounds could exist as phenolates in the solution and carry negative charge together with carboxylates. Additional fragments with other motifs were also identified (10 out of 45) (**Figure 145**). Interestingly, several chemical groups that may carry negative charge were also noted, such as tetrazole that is a carboxylate bioisostere. 6xHis-tau K18(C291S,C322S) exists as a positively charged molecule at pH 7.4 due to its high theoretical pI value of ~9.8. This may cause non-specific charge-charge interactions. However, if the interactions are non-specific to positively charged 6xHis-tau K18(C291S,C322S), it is expected that the responses would be higher than the currently observed ones and the dose response is linear. Additional experiments in the presence of higher salt concentrations (>150 mM) may characterize potential charge-charge interactions. However, it is also possible that current salt concentration of 150 mM is already blocking the majority of non-specific interactions and the observed interactions are representing only a specific binding mode.
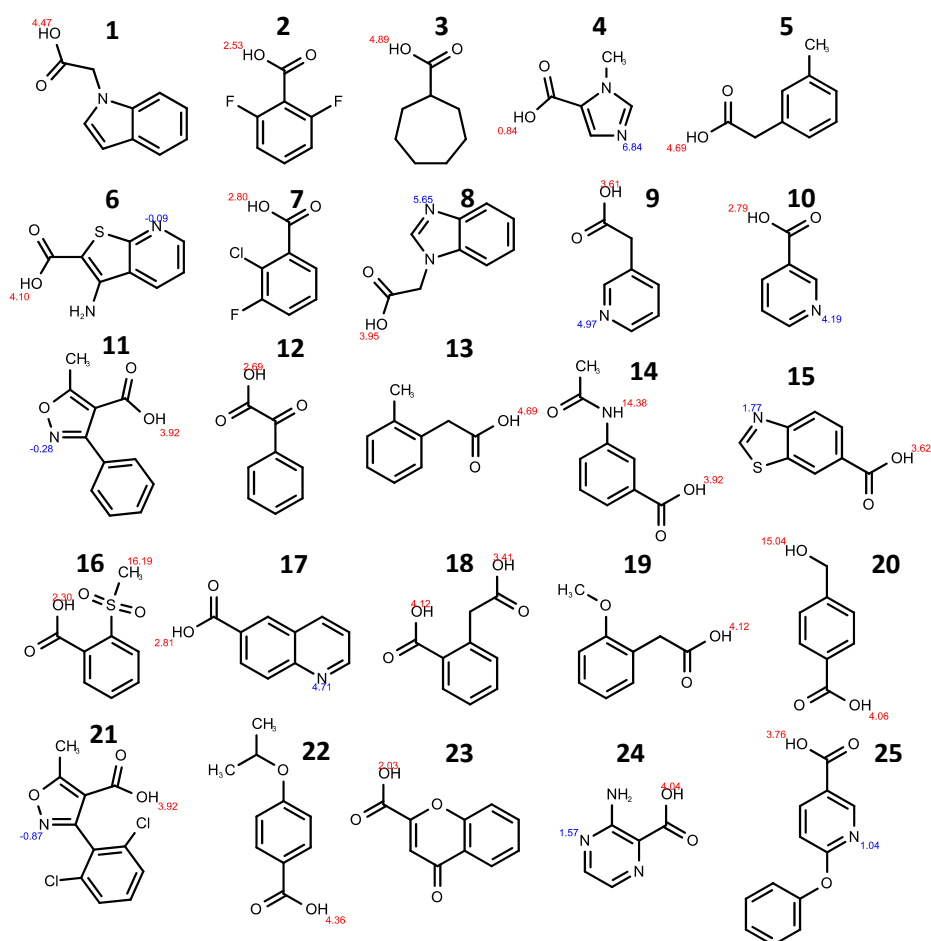
*Figure 142. The identified hits for 6x-His-tau K18(C291S, C322S) from AS that contain carboxylic acid groups. The acidic(**red**)/basic (**blue**) pKa theoretical values were calculated with Marvin Sketch software.*
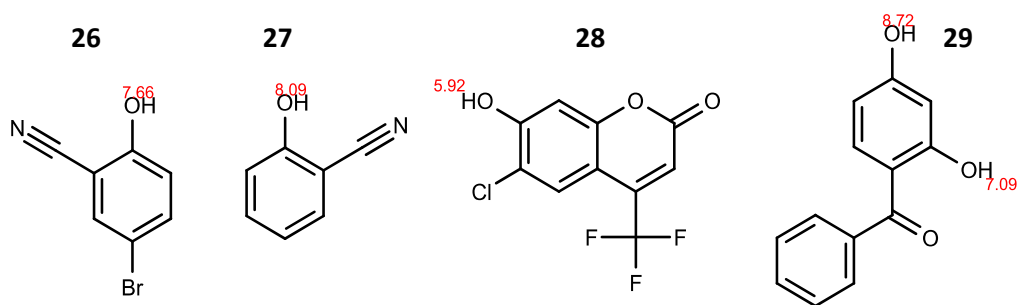
Figure 143. *The identified hits for 6xHis-tau K18(C291S,C322S) from AS that contain phenol groups. The acidic (**red**) pKa theoretical values were calculated with Marvin Sketch software.*
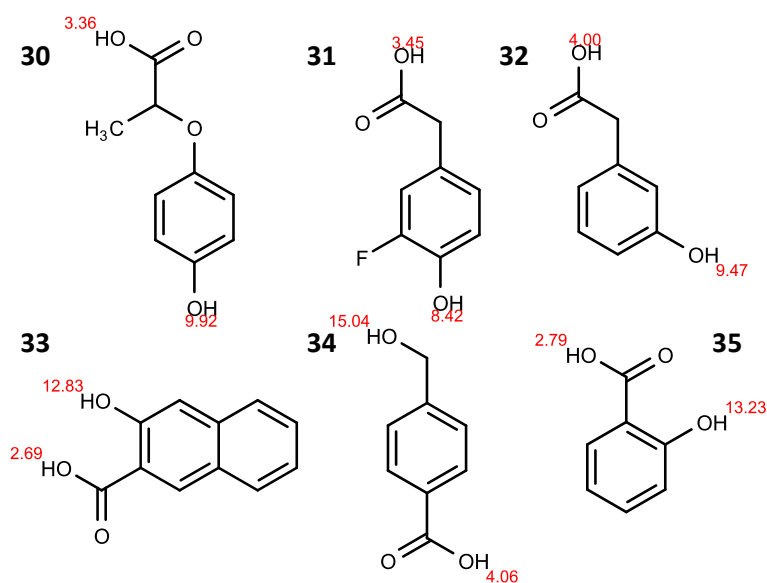
Figure 144. *The identified hits for 6xHis-tau K18(C291S,C322S) from AS that contain both carboxylic acid and phenol groups. The acidic (**red**) pKa theoretical values were calculated with Marvin Sketch software.*
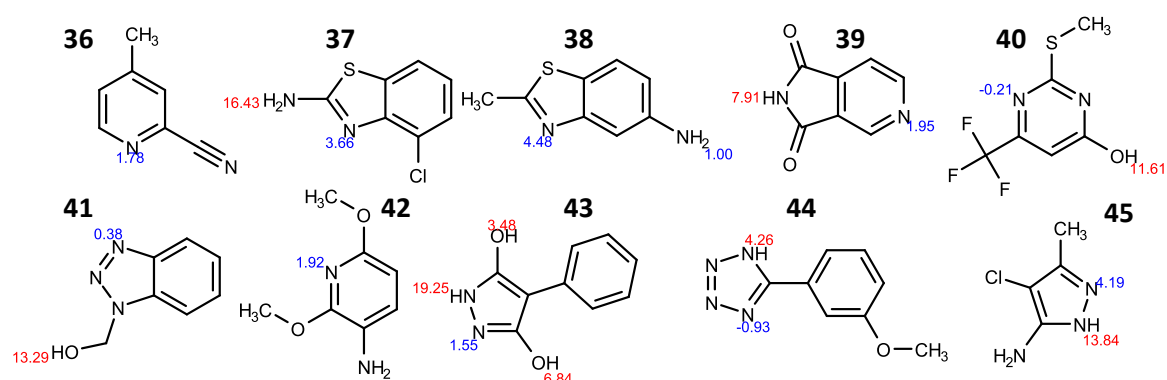
Figure 145. *The identified hits for 6xHis-tau K18(C291S,C322S) from AS that do not contain carboxylic acid or phenol moieties. The acidic(**red**)/basic (**blue**) pKa theoretical values were calculated with Marvin Sketch software.*

After evaluation of steady-state plots, 19 out of 45 hits for 6xHis-tau K18(C291S,C322S) were prioritized for confirmation using NMR approach (**§5.8.4**). Similarly to PTP1B, the evaluation was primarily based on indicative saturating dose response behavior (**§4.7.4**).

It was also noted that potentially non-specific interactions were observable when the concentration of fragment was higher than 125 to 250 μM; this effect was compound-dependent (**Figure 146**).
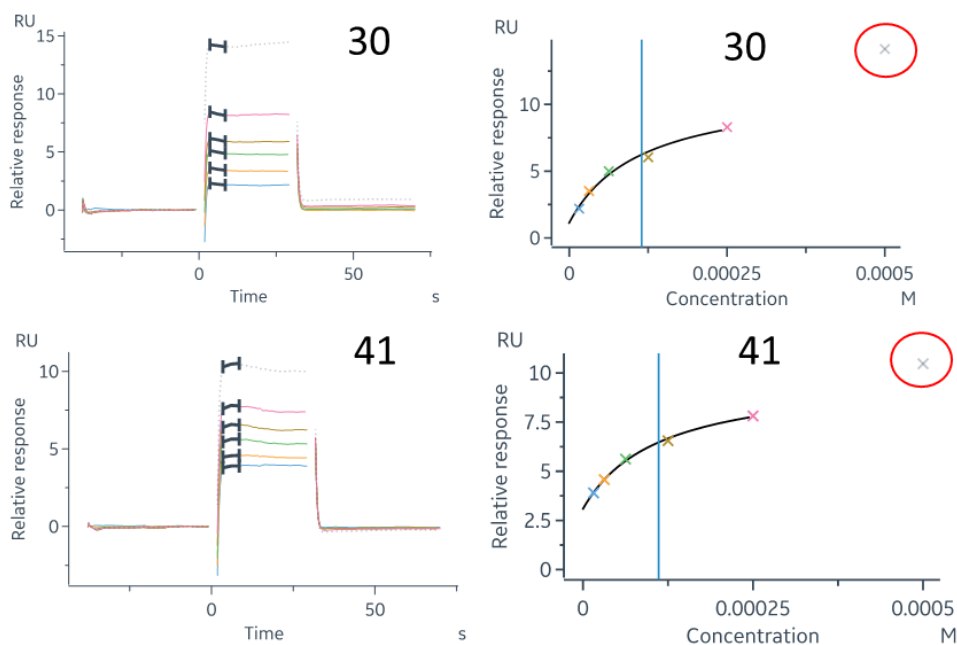
Figure 146. The representative SPR steady-state fitted plots for 6xHis-tau K18(C291S,C322S). Red circle indicates excluded data point. The coloured vertical line represents the calculated K_D values according to the steady-state model. Immobilization level: ~3600 RU. The bolded marks in kinetic sensorgrams denote the report points that were used to produce the steady-state data plots.

Higher than 1:1 stoichiometry is typically observed if the molecules bind non-specifically to the protein due to aggregate formation, charge-charge interactions or other secondary binding effects [202]. In case of 6xHis-tau K18(C291S,C322S), this phenomenon was not observed up to 125 to 250 μM which suggests that the observed interactions below such concentrations were potentially specific. Similar behavior was also observed for PTP1B constructs (**§4.7.4**).

After prioritization of the hits, the ligands were further verified using NMR.

5.8.4   Hit confirmation via NMR

The 1D and 2D NMR techniques were used as an orthogonal approach to confirm the identified hits for 6xHis-tau K18(C291S,C322S) from SPR assay and characterize their binding interactions (**§7.4.2**).

The solubility studies concluded that all of the prioritized 19 fragments were soluble at 500 μM in the same aqueous buffer solution as used for SPR screening in order to rule out any

possible artefacts caused by potential aggregation or precipitation of the compounds (**Figure 147**). DSS was used as a reference compound for approximating the concentration of the ligands in solution and some deviations were observed. However, the deviations up to 20% can be attributed to experimental errors, such as pipetting. Higher deviations would indicate that there is a problem with compound solubility or original stock concentrations of the ligands.
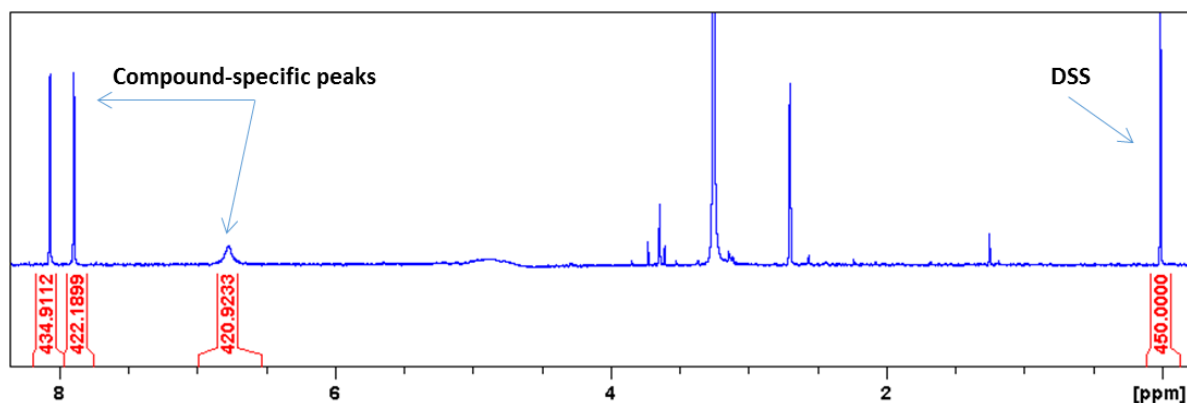
Figure 147. *Solubility assessment of a 500 μM fragment by integration using 1D ^{1}H NMR and DSS as a reference.*

The ligand-observed NMR data indicated no binding events of the fragments to 6xHis-tau K18(C291S,C322S). None of the compounds were found to exhibit increased T_2 relaxation as determined by 1D ^{1}H CPMG-filtered experiments (**Figure 148**).
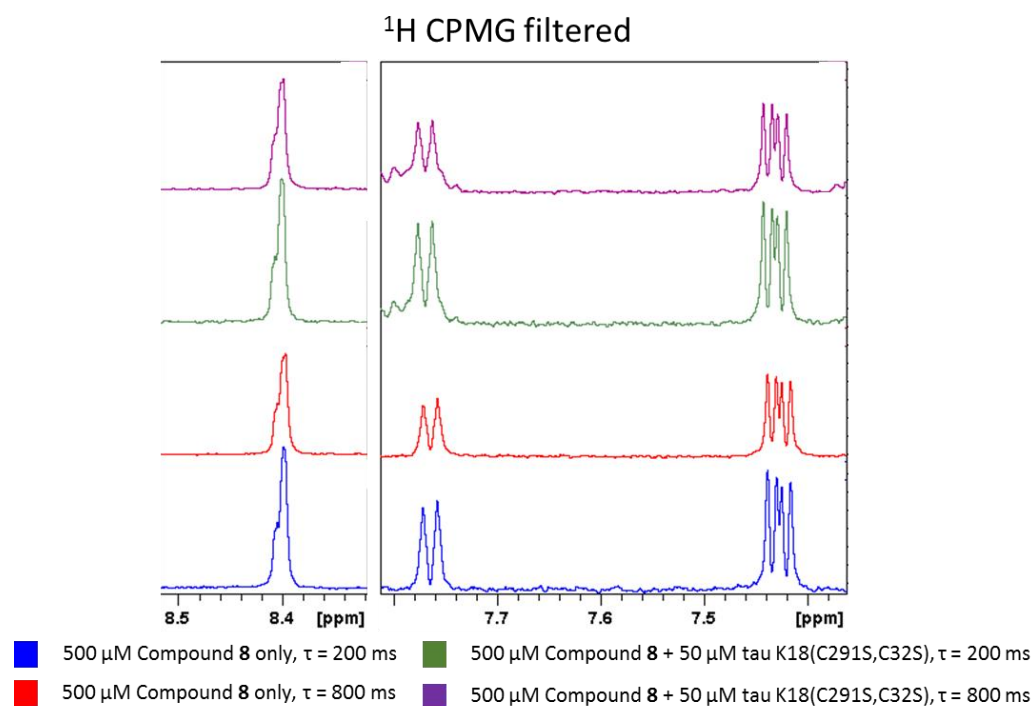
Figure 148. *An example of ^{1}H ligand-observed CPMG spectra of fragment **A** in the absence and presence of 6xHis-tau K18(C291S,C322S).*

NMR DOSY data indicated less than 10% change in diffusion rate in the presence of fragments suggesting no significant conformational changes for 6xHis-tau K18(C291S,C322S) (**Table 20**).

Table 20. The calculated diffusion coefficients for 6xHis-tau K18(C291S,C322S) in the presence of 500 μM fragments.

Compound	Diffusion coefficient D x 10^{-11}, m^2/s	Difference, %
Control	8.74 +/- 0.31	0
1	8.1 +/- 0.03	-7.32
2	8.31 +/- 0.08	-4.92
3	8.1 +/- 0.07	-7.32
4	8.21 +/- 0.07	-6.06
5	8.17 +/- 0.05	-6.52
6	8.34 +/- 0.06	-4.58
7	8.13 +/- 0.04	-6.98
8	8.39 +/- 0.05	-4.00
9	8.04 +/- 0.08	-8.01
10	8.35 +/- 0.07	-4.46
11	8.32 +/- 0.08	-4.81
12	8.17 +/ 0.04	-6.52
13	8.02 +/ 0.04	-8.24
14	8.25 +/- 0.05	-5.61
15	8.18 +/- 0.04	-6.41
16	8.05 +/- 0.05	-7.89
17	8.09 +/- 0.05	-7.44
18	8.12 +/- 0.06	-7.09
19	8.12 +/- 0.03	-7.09

Protein-observed 2D SF-HMQC indicated similar CSPs for all of the fragments (**Figure 149, Left**). The CSPs were possibly caused by slight pH changes or possible charge-charge interactions rather than compound-specific interactions. The CSPs were observed for compounds that have a negatively charged moiety, such as carboxylate or phenolate at pH 7.4. This observation was confirmed by similar CSPs in the presence of 1 mM acetic acid (**Figure 149, Right**). This agrees with previous assumption that the fragments with negatively charged moieties may interact non-specifically with the positively charged protein in SPR assay.

^{1}H-^{15}N SF-HMQC

50 µM ^{15}N Tau K18(C291S, C322S)

50 µM ^{15}N Tau K18(C291S, C322S) + 500 µM prioritized fragment A

50 µM ^{15}N Tau K18(C291S, C322S)

50 µM ^{15}N Tau K18(C291S, C322S) + 1 mM acetic acid
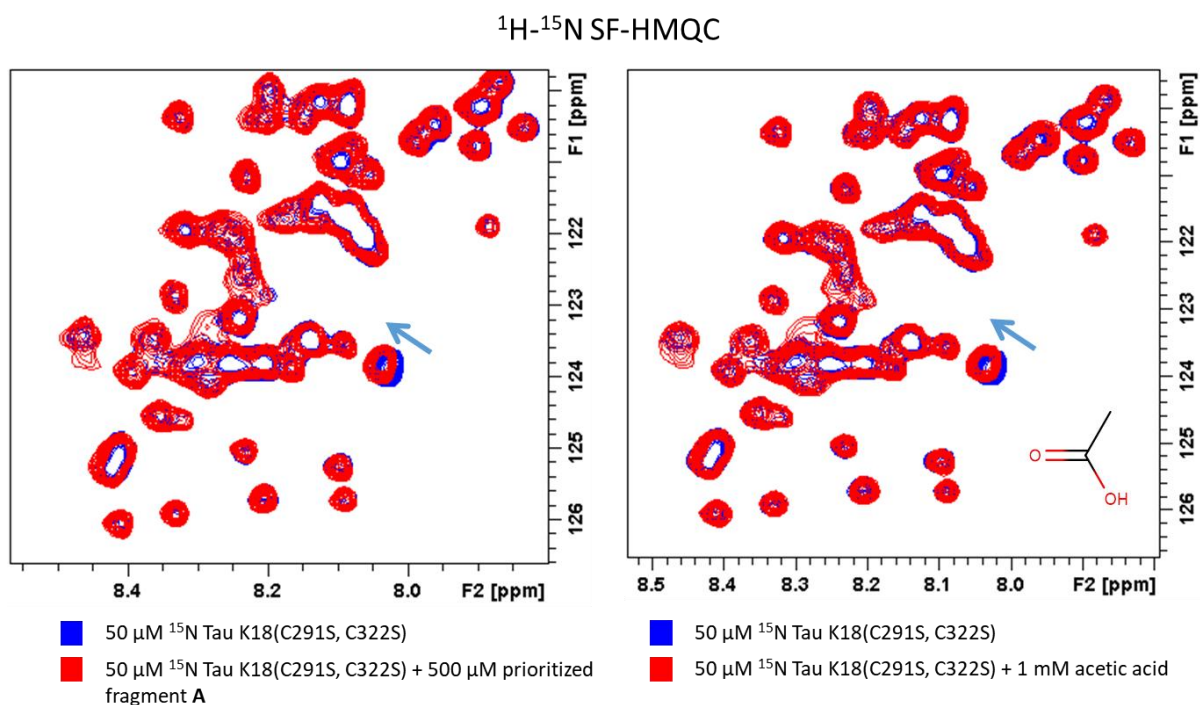
***Figure 149.** 2D ^{1}H-^{15}N SF-HMQC spectra of ^{15}N-labelled 6xHis-tau K18(C291S,C322S) in the presence of fragment A **(Left)** and acetic acid **(Right)**. The blue arrows indicate the vectors of observed CSPs.*

In general, the SPR and NMR data suggest that fragment screening for certain IDPs may benefit from higher salt concentrations in order to prevent artefactual binding due to charge-based interactions. This would be applicable to IDPs that have their pI values significantly higher or lower than the pH of screening buffer solution. In addition to this, oriented immobilization approach, such as biotin:streptavidin, would increase the probability that the immobilized protein is functional, in comparison to a potentially random immobilization using standard amine coupling.

After the prioritized hits from SPR were shown not to interact with tau K18(C291S,C322S) with NMR, a covalent fragment screening campaign was launched with native tau K18(C291,C322). This approach was used to further probe and learn about small IDP:ligand interactions with NMR and HPLC-MS, when there is a 100% occupancy of the binding site.

5.9 Covalent fragment screening

The fluorinated covalent fragment library was used to target native Cys residues in 6xHis-tau K18(C291,C322), in collaboration with Gyorgy Keseru (Hungarian Academy of Sciences, HU). Cys residues are considered to be involved in toxic filament formation in tauopathies, as discussed previously (**§5.1**). By specifically targeting those regions with covalent ligands, it may be possible to prevent the formation of toxic oligomers and filaments. In general,

covalent binders have not been as widely used in the field of drug discovery as non-covalent ligands due to their potential non-specificity and off-target interactions. However, there are now multiple successful cases of targeted covalent inhibitors that have been developed to specifically interact with targeted proteins [197]. The selectivity of covalent inhibitors can be achieved by targeting key nucleophilic groups in the protein which are responsible for its activity, such as deprotonated side chain of Lys or thiol of Cys residue. It is also important to mention that the reactivity of covalent binder may also depend on its conformational positioning amongst other properties. Energetically non-favourable positioning of electrophilic group would make the compound less reactive towards the nucleophile. This can be used to an advantage as it decreases the non-specific reactivity to off-targets. In addition to identifying covalent binders to 6xHis-tau K18(C291,C322), the full occupancy of the binding site would provide additional information about which NMR methods could detect this change.

As a first step, HPLC-MS was used to quickly evaluate the stoichiometry and reactivity of the fluorinated electrophilic fragments towards 6xHis-tau K18(C291,C322) (**Figure 150**) (**§7.4.1, §7.4.1.2**). It has been shown that vinyl sulfone (BTF-29), maleimide (BTF-19, BTF-20) and bromoacetone (BTF-25) warheads reacted with Cys residues of native tau K18(C291,C322) construct to completion by forming 2x adducts. In contrast, other warheads such as epoxide, isothiocyanate, chloroacetamide, acrylamide and alkynes were determined to be less reactive because mixed species of unreacted 6xHis-tau K18(C291,C322), dimerized tau K18(C291,C322) with 1x adduct and tau K18(C291,C322) with 2x adducts were detected (**Figure 151**).
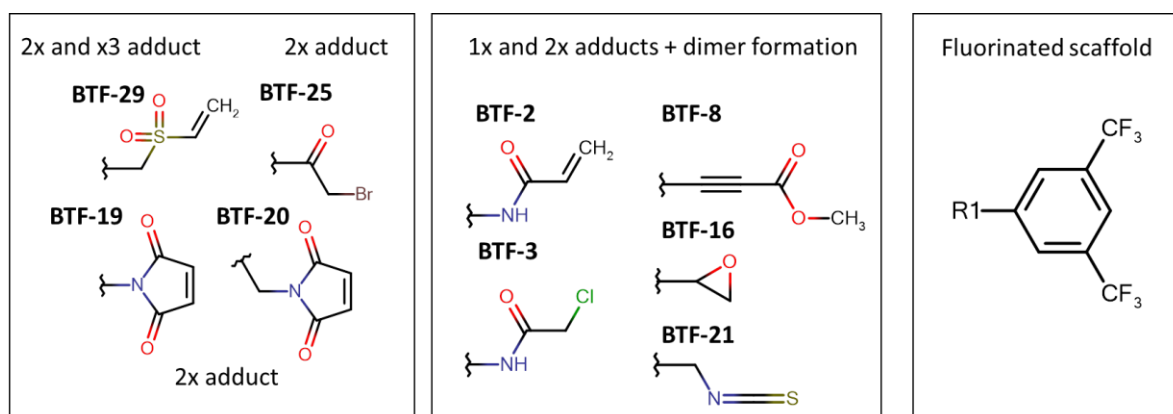
Figure 150. *The chemical structures of identified hits for 6xHis-tau K18(C291,C322).*
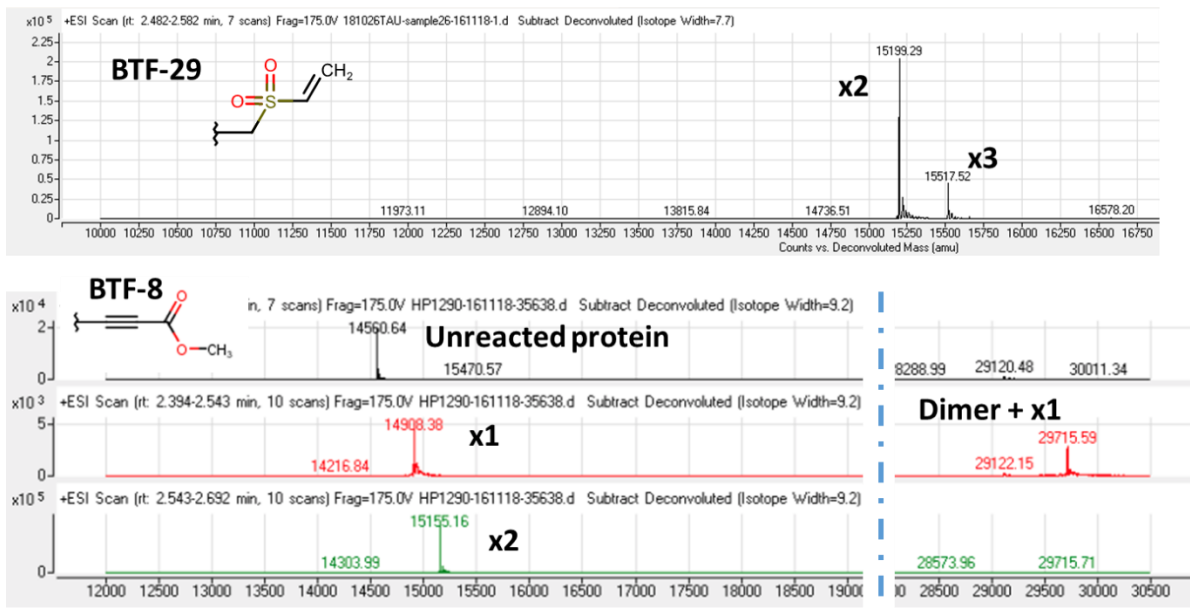
Figure 151. The HPLC-MS spectra of 6xHis-tau K18(C291,C322) in the presence and absence of BTF-29 and BTF-8.

It was also noted that intramolecular disulphide bridges had formed in some mixtures during incubation period. HPLC-MS data indicated a decrease of MW by 2 Da which represented a loss of two protons (**Figure 152**). It is possible that the time-dependent formation of S-S bond may have influenced the reactivity of some molecules with Cys residues in the protein.
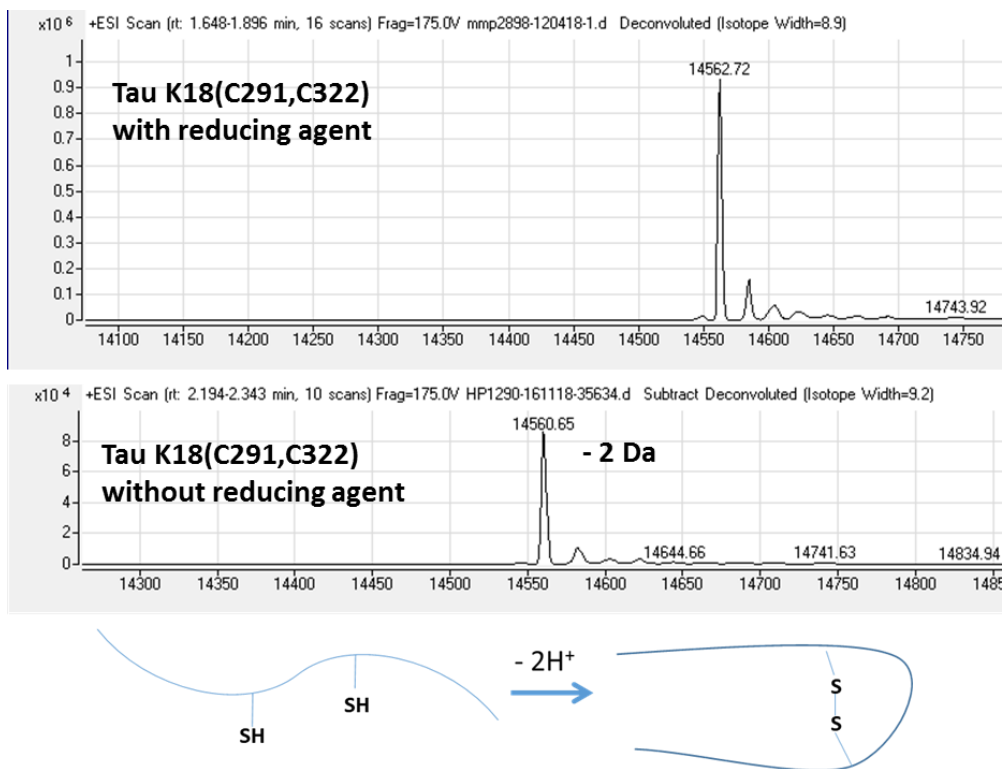
*Figure 152. The HPLC-MS spectra of 6xHis-tau K18(C291,C322) in the presence (**Left**) and absence (**Right**) of reducing agent.*

The reactivity of the identified hits by HPLC-MS was also confirmed by ^{19}F ligand-observed NMR (**§7.4.2**). The observed spectral changes were consistent with reactivity with the appearance of new peaks and the decreased intensity of initial peaks (**Figure 153**).
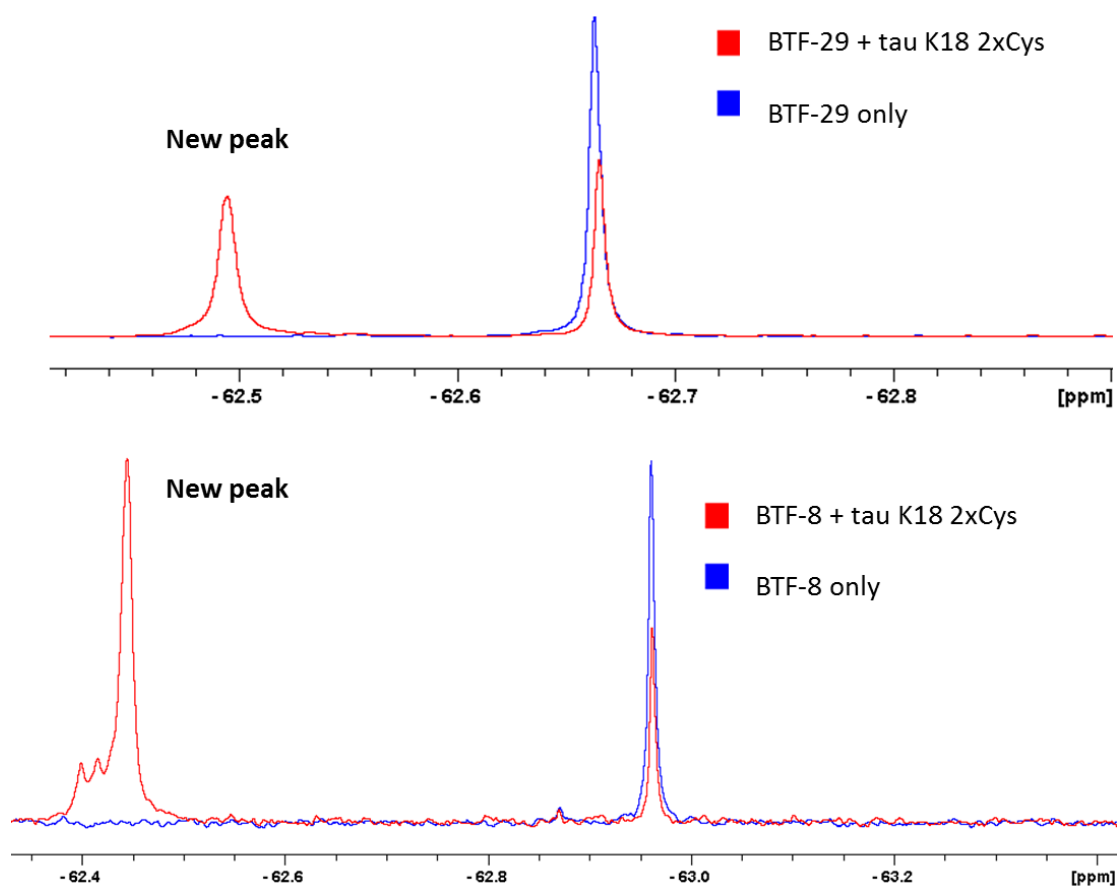
Figure 153. ^{19}F ligand-observed NMR spectra of BTF-29 and BTF-8 before and after reacting with 6xHis-tau K18(C291,C322) during overnight incubation.

While HPLC-MS and ^{19}F NMR have shown the reactivity and stoichiometry of electrophilic molecules with 6xHis-tau K18(C291,C322), 2D [1]H-^{15}N SF-HMQC spectra indicated the affected amino acid residues. No additional MS/MS studies have been performed to confirm that the fragments bind directly to Cys residues. However, the CSPs were observed for amino acids that are close to Cys residues in native tau K18(C291,C322) sequence. Moreover, the CSPs were not of the same magnitude for the different binders. The fully reacted BTF-29 caused larger CSPs than partially reacted BTF-3 which can be observed as two distinct populations of 6xHis-tau K18(C291,C322) (**Figure 150, Figure 154**). It is also possible that larger CSPs were observed due to different chemical structures of the ligands.

2D ¹H-¹⁵N SF-HMQC

C291,C322

● Tau K18(C291,C322) only
● Tau K18(C291,C322) + BTF-29
● Tau K18(C291,C322) + BTF-3

```
       246          256          266          276          286          296
GHHHHHHQTA PVPMPDLKNV KSKIGSTENL KHQPGGGKVQ IINKKLDLSN VQSKCGSKDN

       306          316          326          336          346          356
IKHVPGGGSV QIVYKPVDLS KVTSKCGSLG NIHHKPGGGQ VEVKSEKLDF KDRVQSKIGS

       366   372
LDNITHVPGG GNKKIE
```
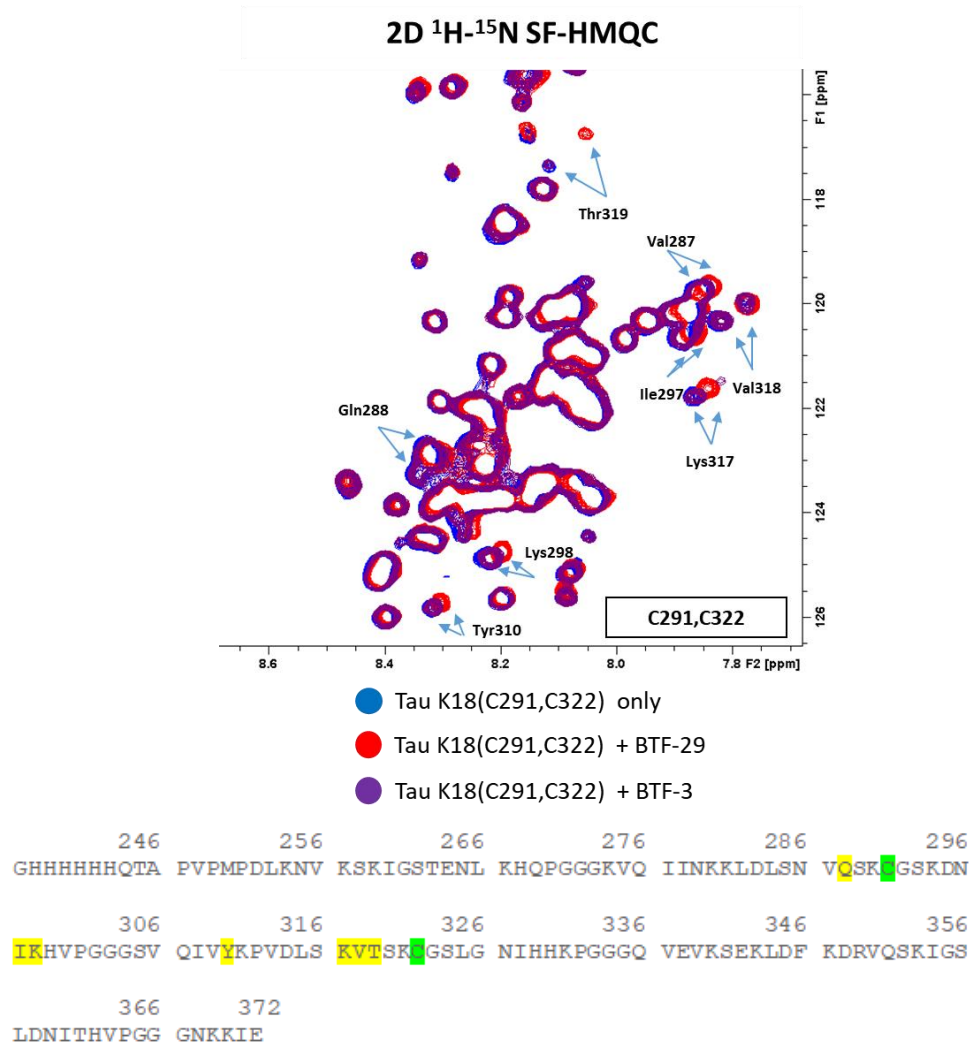
*Figure 154. A comparison of 2D SF-HMQC spectra of 6xHis-tau K18(C291,C322) in the presence and absence of BTF-29 or BTF-3. The sequence marks Cys residues (**green**) and affected protein backbone amides (**yellow**).*

Due to close to 100% occupancy of the binding site by covalent fragments in 6xHis-tau K18(C291,C322) protein, it was hypothesized that diffusion rate may get perturbed. However, the diffusion rate of covalently modified 6xHis-tau K18(C291,C322) was determined to change by less than 10% by NMR DOSY, indicating that no large structural changes have occurred (**Figure 155**). The standard deviation of N = 2 for 6xHis-tau K18(C291,C322) was calculated to be 0.13. The experimental diffusion coefficients for the covalently modified samples did not differentiate more than 3 standard deviations from the control sample.

Compound	Diffusion coefficient D x 10^{-11}, m²/s	Difference, %
Control (n = 2)	7.74 +/- 0.04	0.00
BTF-2	7.94 +/- 0.04	2.58
BTF-3	8.03 +/- 0.04	3.75
BTF-8	7.74 +/- 0.06	0.00
BTF-16	8.08 +/- 0.04	4.39
BTF-19	7.97 +/- 0.03	2.97
BTF-21	7.71 +/- 0.04	-0.39
BTF-29	7.86 +/- 0.04	1.55
BTF-20	7.78 +/- 0.03	0.52
BTF-25	7.47 +/- 0.07	-3.49
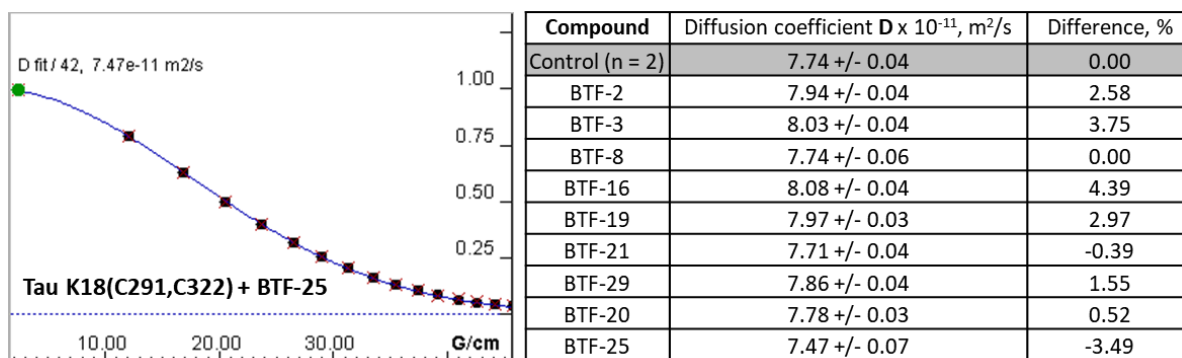
Figure 155. A comparison of 6xHis-tau K18(C291,C322) diffusion rate values in the presence of reactive compounds.

It can be hypothesized that the fragments in general may not cause large enough conformational changes in the protein structure to affect the diffusion rate, even with 100% binding site occupancy. The conformation of a protein may be affected more significantly if larger molecules, such as peptides, interact with the protein. The larger ligands may provide additional interactions for the protein to cause significant conformational and structural changes. The identified reactive warheads, such as BTF-29, can be potentially attached to a known peptide binder of 6xHis-tau K18(C291,C322) which may increase specific interactions and decrease the general reactivity of the warheads with Cys. Several research groups have already identified multiple putative tau K18 binders, such as rhodanines and targeted peptides [227], [265], [266]. These ligands are thought to bind close to Cys residues in the PHF6 region and may serve as a part of a molecule that could be used to selectively bind to native tau K18(C291,C322) and inhibit the formation of fibrils.

5.10 Conclusions

Multiple tau K18 protein constructs were produced in order to investigate tau K18:ligand interactions using NMR, SPR and MST approaches.

2D protein-observed NMR data have shown that only compound **6** (Cl-NQTrp) out of 10 literature-reported ligands interacted with monomeric tau K18(C291S,C322S). Such results suggested that other ligands may be interacting with oligomeric or aggregated forms of tau K18 in solution. Similar CSPs were also observed with NNs of compound **6.** Interestingly, some NNs affected the solubility of tau K18 suggesting some NN-specific interactions. The tau K18:compound **6** interactions were also confirmed using 1D ligand-observed wLOGSY and ¹H CPMG-filtered data, in addition to protein-observed ¹⁹F NMR techniques. The observed increased T_2 relaxation rates for Cl-NQTrp were attributed to intramolecular interactions,

154

suggesting that good characterization of the ligand is necessary in order to correctly interpret the data. The follow-up X-filtered NMR experiment failed to identify specific conformational interactions potentially due to weak IDP:ligand complex.

SPR and MST techniques allowed using lower concentrations of the literature-reported ligands in order to observe binding events. The used techniques confirmed additional ligands binding to tau K18, including compound **1**, **2**, **5** and **6**. However, the possibility of artefacts had to be considered due to them not being detected by NMR previously, with the exception of compound **6**.

Due to experimental design and inherent sensitivity, SPR approach seems to be prone to detecting charge-charge interactions, as shown with confirmatory NMR experiments after SPR-based fragment screening campaign which identified multiple hits to tau K18. Some charged ligands can appear as false positives and higher salt concentrations may be needed to counter this phenomenon. This means that a good knowledge of IDP system is required to objectively establish optimal experimental assays. In addition to this, IDPs may be immobilized in a non-functional or non-native conformation if non-oriented immobilization approach is used, such as amine coupling. Therefore, oriented immobilization approaches, such as biotin-streptavidin system, may provide a more representative protein conformation for screening purposes while providing similar surface stability over time.

MST method was identified to produce ambiguous data which requires to be confirmed by more robust and label-free methods, such as NMR. MST may be particularly prone to artefacts due to usage of dyes and possible conformational changes of the IDP when interacting with a ligand.

Covalent binders for wild-type tau K18 (C291,C322) provided additional insight about what behaviour to expect when an IDP has a fully bound LMW ligand. The diffusion rate was not significantly perturbed with covalent fragment which concluded that NMR DOSY may not be reliably used in fragment screening campaigns, as fragments may not cause large enough conformational changes. Nevertheless, the identified covalent warheads may be potentially developed into tau K18-specific binders when attached to larger molecules.

6 Final conclusions and future challenges

The IDPs and IDRs have emerged as potential therapeutic targets that are involved in a variety of important cell signalling cascades that also are present in multiple diseases. In addition to this, the discovery of IDPs has started changing the paradigm of '*structure->function*' that existed for decades prior. However, the potential of novel therapeutic targets also introduced multiple challenges which made the currently established drug discovery approaches prone to failure, including lack of enzymatic activity and structure for IDPs, and limited knowledge about the IDP partner proteins in occurring PPIs.

Nevertheless, the presence of several successful examples at targeting IDPs suggested that finding IDP-modulating ligands is possible. It was also noted that the majority of previously literature-reported IDP ligands were developed by following the phenotypic effect without much structural information obtained from biophysical methods. This investigation of four therapeutically relevant IDP systems contributed to the general knowledge about IDPs, their behaviour in solution, stability and potential to be characterized using a variety of biophysical techniques.

6.1 Protein production

It was identified that some IDPs were be prone to proteolytic degradation during expression in procaryotic expression systems due to their disordered state, as shown for PTP1B. The problem was solved by introducing N-terminal and C-terminal tags which helped to stabilize and purify intact forms of IDPs. In contrast, other IDPs, such as tau K18 and HASPA, were found to be resistant to *in vivo* degradation during expression, possibly due to their specific properties, including elevated general charge and specific pseudo-structural elements. As some IDPs exhibited not satisfactory expression levels, the yields were increased by using optimized expression vectors with T7 promoter, where possible. The pET vectors were shown to be a preferred expression system over commercially available pJ411 expression system.

The expression of short peptides in fusion proteins has also been proven to be difficult due to *in vivo* degradation, as shown for cMyc. The proteolysis problem was solved by expressing the peptides in *E. coli* as insoluble inclusion bodies. The efficient chemical cleavage was potentially prevented by lack of sample purity. A robust purification assay should be developed in order to complete the peptide production protocol.

Additional labeling of proteins during protein production may allow additional characterization of IDPs or ligands, as shown with ^{19}F and biotin labelling for tau K18 constructs.

6.2 NMR

NMR spectroscopy was used as the main biophysical approach to characterize IDP:ligand interactions due to its ability to directly observe the behaviour of the protein and ligand in solution. It was shown that if the data is consistent with binding in 2D NMR, the binding was also detected in some of 1D ligand-observed NMR experiments, including wLOGSY and ^{1}H CPMG-filtered. The knowledge of the physicochemical processes in solution is essential, including chemical exchange, ionic strength of the solution, T_2 relaxation, intramolecular interactions and potential aggregate formation, in order to avoid artefacts.

It was noted that NMR DOSY technique did not detect any binding events with any of the confirmed binders as the diffusion coefficient was not affected. It is possible that the conformational changes were not significant enough to modulate the diffusion rate. One of potential drawbacks of NMR DOSY is that the technique is not able to detect the general changes in hydrodynamic radius for the whole molecule, only for molecule-specific peaks. However, the technique may still be a useful orthogonal method in assessing potential aggregation of the compounds in solution, together with ^{1}H CPMG-filtered experiments.

6.3 SPR

SPR technique was used to further characterize the identified binders to the IDPs and assess the capabilities of SPR to be used in fragment screening campaigns.

It was noticed that the IDPs may be required to be immobilized on the surface in a specific conformation in order to obtain representative results. Since only specific conformations of IDPs may be functionally relevant, oriented immobilization approach should be a preferred option for the highest chance of IDPs to have enough conformational freedom and stability to represent *in vivo* activity on the surface. The oriented and stable immobilization can be achieved by, for example, introducing a biotin tag to the protein and using a streptavidin-derivatized surface.

Due to its inherent sensitivity, SPR was also noted to be prone to detecting charge-charge interactions. Such observations have shown the importance of orthogonal assays in order to

identify such artefacts and avoid false positives. Such possibility should be considered before starting any experiments by assessing such parameters as pI of the protein and ionic strength of the buffer solution. Moreover, any conformational changes on the surface may also produce ambiguous results.

6.4   MST

The use of MST for detecting IDP:ligand interactions was noted to produce ambiguous results due to the ligands potentially interacting with the dye, as the fluorescence of the dye was affected even without the protein present, as shown for tau K18. However, a larger change of the MST data with protein than without the protein may indicate the presence of potentially specific IDP:ligand interactions.

Due to their unrestricted conformational freedom, the IDPs may assume specific conformation in the presence of a ligand that directly and specifically affect the fluorescence of the dye. It would be of interest to compare the IDP:ligand interactions using different labelling approaches to learn more about this phenomenon, including covalent and random dye labelling. It is possible that the preferred His-tag labelling approach may not be the most representative as the IDPs may not undergo significant enough changes.

6.5   Fragment-based ligand discovery and fragment libraries

While FBLD covers a larger chemical space and increases a chance of identifying a specific ligand, it has been determined that the IDP:fragment interactions may have to be either target-specific with reasonable affinity or risk to become false-positives due to potential non-specific interactions, such as charge-charge or aggregation. Moreover, larger fragments should be considered to be included in the libraries for an increased interaction potential between IDP and the fragment. As an alternative approach, screening at very high fragment concentrations could also be an option in order to identify fragments with very low affinities using encapsulation method with NMR approach.

Multiple fragment screening attempts targeting HASPA, PTP1B and tau K18 showed that it is important to be aware of advantages and disadvantages of each biophysical technique, and any potential artefacts.

In addition to this, covalent fragment library was used to identify specific binders to tau K18. Covalent binders may offer another approach to contribute to the development of the IDP-specific binders by attaching the identified hits to larger molecules to increase specificity.

6.6   General applicability of biophysical methods for detecting IDP:ligand interactions

Different biophysical approaches offer different kind of information which can help identifying and validating protein's interactions with LMW ligands. Due to IDPs being specific kind of proteins, special approaches are also needed in order to produce reliable data and avoid ambiguity.

One of the main NMR advantages over other biophysical methods is that a protein does not have to be dye-labelled or immobilized on the surface, potentially altering the therapeutically relevant conformation of the protein. Such restriction of freedom may be especially relevant for IDPs that potentially prefer a specific, free-in-solution conformation. The 1D and 2D NMR methods can then be used to observe the behaviour of ligands and IDPs which can help identifying potential artefacts early and track any structural changes to the IDP upon interacting with a LMW molecule. While 1D STD technique may not work due to lack of folded domains in ID regions, wLOGSY, CPMG-filtered and 2D HSQC techniques can potentially be used in identification of binders, if the affinities are high enough. If short peptides are being investigated, 2D ^{1}H-^{1}H TOCSY and NOESY can also provide important information about potential interactions, as shown in the case of cMyc. In some cases, site-specific ^{19}F labelling of the protein or ligand can also help identifying and characterizing the binding events.

When it comes to the limitations of NMR technique, due to it being a relatively insensitive, low expression yields of the protein of interest may become a problem. If that appears to be the case, isotopic labelling can also become a relatively expensive step once it is necessary to do more advanced NMR experiments. In addition to this, NMR DOSY may require larger conformational changes to be present for them to be detected, as shown in the case of PTP1B. Moreover, some NMR methods require long experimental times which may hinder the ligand screening process further.

As an alternative approach, SPR technique also has a potential to be used in IDP drug screening campaigns. Due to its inherent sensitivity, SPR approach requires only ng amounts of protein for its assays, similarly to MST, whereas NMR can use up to mg quantities of protein

per sample. Moreover, SPR technique can determine the stoichiometry and kinetics of the interactions in a relatively fast manner. This information can also provide some insights about potential artefacts, similarly as NMR. In addition to this, the throughput of screening is faster using SPR than NMR, when using a multi-injection instrumentation.

One of the main disadvantages of SPR method is that the protein of interest has to be immobilized on the surface. The immobilization may cause the IDPs to lose their therapeutically relevant conformation and provide ambiguous data. However, it may be possible to minimize such conformational restrictions of IDPs by using specific, IDP-tailored immobilization approaches, which can include immobilizing the protein via flexible linker. The same reasoning can potentially be applied for the ligands by immobilizing them onto the surface rather than the protein. However, such approach may become a rather expensive approach and would require special machinery. SPR technique is also prone to charge-charge artefacts if the ligands or proteins have a highly charged structure. Specific assay development would be required to mitigate such interactions, including increasing salt concentration or chemically modifying the dextran surface.

MST can be used as an orthogonal technique to validate IDP:ligand interactions. Its fast data acquisition time and small material consumption may help in cases when sample amount is a problem. While MST may be able to detect the binding of the ligand to IDP, the need to label the protein with a fluorescent dye may prevent accurate evaluation of interactions for IDPs. Such inaccuracy may arise due to IDP's potential sensitivity to any surface modifications. Moreover, the ligands may interact with the dye itself rather than the protein. Therefore, additional control experiments are necessary.

6.7   Future challenges and opportunities

Despite investigating four different IDP systems, further research is still necessary to challenge and push the current understanding of IDP:ligand interactions, including the ways to detect that.

As the results of the project suggest, FBLD may not be the right approach for identifying ligands to IDPs due to low affinities. If the protein binding partners of IDPs/IDRs are unknown, it may be of interest to screen chemically more complex molecules using NMR as a primary approach. This may potentially increase the chance of identifying ligands with high-enough

affinity. The hits can then be further validated by SPR and MST to get a better understanding of the system.

If the NMR approach is not possible, an SPR approach may be considered as alternative approach after immobilizing the IDP via a long, flexible linker. This would potentially allow the IDPs to maintain their native conformation. If any hits are identified, next step should involve confirmatory NMR methods to identify the binding site and any artefacts that may have been not detected.

If there are known binding partners of IDPs of interest, it can be beneficial to perform a high throughput fluorescence-based assays first, indicating the disruption of targeted PPIs. NMR, MST and SPR may also indicate such disruptions of PPIs, although with a lower throughput. However, careful experimental planning would be necessary for such approaches to avoid artefacts. Any identified ligands can then be validated by NMR for further structural analysis and development.

It is evident that even if a fragment/ligand is identified to interact in a specific manner with an IDP, it can still be a challenge to develop the hit further due to lack of structural information. However, molecular modelling and advanced NMR-guided models may provide some structural guidance to understand the occurring interactions better. Fortunately, high affinity ligands may not be required with the discovery of proteolysis targeting chimeras (PROTACs) that change the concept of drugs from occupancy-based to event-based model. The IDPs could be degraded in a desirable manner even with low affinity but specifically interacting molecules. Nevertheless, the exact inhibitory effects have to be studied as IDPs are involved in multiple signalling cascades which may cause non-desirable side effects once the function of an IDP is suppressed.

7   Materials and Methods

7.1   Materials

All reagents and solvents were obtained from commercial vendors unless stated otherwise.

7.1.1   HASPA

HASPA coding sequence in pET-28b expression vector was provided by Michael Plevin (University of York, UK).

The fragment library contained 559 fragments from an in-house Maybridge library and 51 3D fragments from Hanna Klein (University of York). The fragments were split into 93 fragment mixtures with 6 compounds each at 33.33 mM in 100% d_6-DMSO.

7.1.2   cMyc

Soluble fusion protein KRas-EK-cMyc(363-381) and KRas-EK-cMyc(402-412) coding plasmids in pJ411 expression vector were ordered from DNA 2.0 (US). Inclusion body-forming ketosteroid isomerase (KSI) construct with cMyc coding sequences were ordered from GenScript (US) in pET-31b(+) vector.

7.1.3   PTP1B

6xHis-TEV-PTP1B(1-301) and 6xHis-TEV-PTP1B(1-393) coding plasmids were kindly provided by Nicholas Tonks (ColdSpring Harbour, USA). Other codon-optimized constructs were obtained from either DNA 2.0 (USA) or GenScript (USA) with 5'-NdeI and 3'-XhoI restriction sites. All reagents and solvents were obtained from commercial suppliers unless indicated otherwise.

The fragment library of 1000 compounds was provided by Sci-Life Lab (Stockholm, SE) as 50 mM stocks in DMSO. 80 additional fragments were added to the library from collaborators in FragNet consortium (ESR 1, 2 and 7, personal communication).

The fluorinated CF_3 fragment library of 1664 compounds was provided by Novartis (CH).

The expression plasmids that contain various tags (MBP, Trx, SUMO, GST) were obtained from an in-house library.

7.1.4   Tau K18

The initial batch of ^{15}N labelled 6xHis-tau K18(C291S,C322S) and several literature-reported tau K18 ligands were kindly provided by Servier (FRA). The construct coding for 6xHis-tau K18(C291S,C322S) was codon-optimized and obtained from DNA 2.0 (USA) in pJ411 vector.

The construct coding for 6xHis-TEV-AviTag-tau K18(C291S,C322S) was obtained in T7-driven expression vector pET-24a from GenScript (USA). All inserts were designed with 5'-NdeI and 3'-XhoI restriction sites to enable sub-cloning approach, if required. Such restriction sites were chosen because the majority of in-house vectors had their cloning site designed with NdeI and XhoI restriction sites. 6xHis tag was used as an affinity tag for purification and biophysical methods. AviTag was used for *in vitro* biotinylation and biophysical experiments.

The electrophilic fragment library was provided as dry powder by Gyorgy Keseru (Hungarian Academy of Sciences, HU) and prepared as 100 mM stocks in d_6-DMSO.

The fragment library for SPR screening was the same as for PTP1B fragment screening campaign (**§7.1.3**).

The pET-24a vector with NdeI/XhoI cloning site was obtained from an in-house library.

7.2   Molecular biology

7.2.1   Heat shock transformation

The required *E. coli* strain cells were removed from -80°C and thawed on ice for 5 min. This was followed by addition of 50-500 ng of target plasmid DNA. The sample was mixed gently by flicking the tube. The mixture was then incubated on ice for 30 min. Following this, the sample was incubated in 42°C water for 45 sec and put on ice for 2 min. Then, 300 µL of SOC outgrowth medium (NEB, UK) was added and the cells were incubated at 37°C for 1 hour. After this, 20-50 µL and the rest of the mixture was then plated on agar plates with appropriate antibiotics and incubated at 37°C overnight.

7.2.2   Cloning

All digestion enzymes and cleavage buffers were obtained from NEB (UK). Undigested plasmids were prepared as reference samples for control purposes. Plasmids containing either insert or empty vector were double-digested for 1 hour at 37°C in the following mixtures:

	Insert reference, µL	Insert digested, µL
DNA (~0.5 - 1 µg)	4	4
10 x Cleavage buffer	4	4
Digestion enzymes	0	0.5 of each enzyme
Distilled H$_2$O	32	31
Total Volume	**40**	**40**

	Vector reference, μL	Vector digested, μL
DNA (~0.1 - 0.5 μg)	2	2
Cleavage buffer	4	4
Digestion enzymes	0	0.5 of each enzyme
Distilled H$_2$O	34	33
Total Volume	**40**	**40**

Enzyme mixtures	Cleavage buffer
NdeI and XhoI	CutSmart® Buffer (NEB)
BamHI and XhoI	NEBuffer® 3.1 (NEB)

The digested plasmids were separated on 1% agarose gel in TAE buffer solution (40 mM Tris pH 8.5, 20 mM acetate, 1 mM EDTA) for 1 hour at 100 V and extracted using QIAquick Gel Extraction kit (Qiagen). DNA fragments were excised from the agarose gel, weighed and mixed with Buffer QG with a ratio of 3:1 (1:1 - 100 μL of Buffer QG to 100 mg of gel). The mixture was incubated at 50°C for 10 min, vortexed and 1 volume of propan-2-ol was added. After mixing, the mixture was loaded on QIAquick spin column and centrifuged at 10000 g for 1 min. Then, the spin column was washed by adding 750 μL of Buffer PE to the column which was spun down twice for 1 min at 10000 g. Finally, the DNA was eluted by adding 50 μL of Buffer EB (10 mM Tris-HCl, pH 8.5), incubating for 1 min and spinning down for 1 min at 10000 g.

The ligation was performed at RT using T4 ligase (NEB, UK) in total volume of 20 μL in the following mixtures:

	Vector, μL	Insert, μL	10x Ligase Buffer, μL	T4 ligase, μL	Distilled H$_2$O, μL
Ligation	5	12	2	1	0
Control	5	0	2	1	12

1x Ligase buffer: 50 mM Tris-HCl pH 7.5, 10 mM MgCl$_2$,1 mM ATP, 10 mM DTT.

The ligation mix was used to transform NEB 5-alpha competent *E. coli* cells (NEB, UK) using heat shock protocol (**§7.2.1**). After overnight incubation, 2 to 5 discrete colonies were picked and used to inoculate 5 mL of LB media with appropriate antibiotic (see **§8.2** for plasmid antibiotic resistances) and the cells were grown overnight at 37°C at 200 RPM. The plasmid was purified using QIAprep Spin MiniPrep Kit (Qiagen). The ligated sequence was confirmed by Sanger sequencing by GeneWiz (UK).

7.2.3 Site-directed mutagenesis

Single-site mutations were achieved using site-directed mutagenesis with QuikChange Lightning SDM Kit (Agilent). Primers were designed using Agilent website tool (https://www.agilent.com/store/primerDesignProgram.jsp) and acquired from Sigma-Aldrich.

7.2.3.1 Tau K18

The primers used to introduce Cys mutations to 6xHis-tau K18(C291S,C322S) are as follows:

- Primer A: Ser to Cys (S291C)

- Primer B: Ser to Cys (S322C)

- Primer C: Ser to Cys (S291C) with S322C mutation already present

A

Forward Primer 5'-GCAATGTGCAGTCTAAG**T**GCGGTTCCAAAGACAA-3'

Reverse Primer 5'-TTGTCTTTGGAACCGC**A**CTTAGACTGCACATTGC-3'

B

Forward Primer 5'-GGTTACCAGCAAG**T**GCGGTAGCCTGGG-3'

Reverse Primer 5'-CCCAGGCTACCGC**A**CTTGCTGGTAACC-3'

C

Forward Primer 5'-CCTGAGCAATGTGCAGTCTAAG**T**GCGGTTCCAAAGACAATATC-3'

Reverse Primer 5'-GATATTGTCTTTGGAACCGC**A**CTTAGACTGCACATTGCTCAGG-3'

Targeted nucleotides are marked in red. The nucleotide sequence for primer C is the same as for primer A but with extensions (cyan) that made the mutagenesis approach possible (see **§5.2** for primer design discussion).

7.3 Protein production

All plasmids were transformed into designated *E. coli* cell strains using standard heat shock protocol (**§7.2.1**)

7.3.1 Expression in LB or TB media

One transformed colony with plasmid of interest was used to inoculate 100 mL of LB or TB media with appropriate antibiotics for overnight hour growth at 37°C with continuous

agitation at 200 RPM. On the subsequent day, 1 mL of overnight culture was used to inoculate 1 L of LB or TB media with appropriate antibiotics. The cells were grown at 37°C with continuous agitation at 200 RPM until OD_{600} of 0.8, and induced with appropriate amount IPTG for timed expression at specific temperature at 200 RPM agitation. After expression, the cell pellets were collected by centrifugation at 4500 RPM for 20 min at 4°C using JLA 8.1000 rotor and stored at -80°C. Multiple non-labelled cMyc, PTP1B and tau K18 constructs were expressed in LB or TB media (**Table 21, Table 22, Table 23**). See **§8.1**, **§8.2** for construct sequences.

Table 21. The expression conditions for various cMyc constructs.

Construct	Cell line	Vector	Conditions
KRas-EK-cMyc(363-381)	*E. coli* BL21(DE3) pLysS	pJ411	1 mM IPTG for overnight expression at 18°C in LB media
KRas-EK-cMyc(402-412)	*E. coli* BL21(DE3) pLysS	pJ411	1 mM IPTG for overnight expression at 18°C in LB media
KSI-cMyc(363-381)	*E. coli* BL21(DE3) pLysS	pET-31b	1 mM IPTG for 3 hour expression at 37°C in LB media
KSI-cMyc(402-412)	*E. coli* BL21(DE3) pLysS	pET-31b	1 mM IPTG for 3 hour expression at 37°C in LB media

Table 22. The expression conditions for various PTP1B constructs.

Construct	Cell line	Vector	Expression conditions
6xHis-TEV-PTP1B(1-301)	*E. coli* BL21(DE3) RIL Codon Plus	pET-28b	1 mM IPTG for overnight expression at 18°C in LB media
PTP1B(1-393)-6xHis	*E. coli* BL21(DE3) pLysS	pJ411	1 mM IPTG for overnight expression at 18°C in TB media
MBP-TEV-PTP1B(1-393)-6xHis	*E. coli* BL21(DE3) pLysS	pJ411	1 mM IPTG for overnight expression at 18°C in TB media
Trx-TEV-PTP1B(1-393)-6xHis	*E. coli* BL21(DE3) pLysS	pJ411	1 mM IPTG for overnight expression at 18°C in TB media
SUMO-TEV-PTP1B(1-393)-6xHis	*E. coli* BL21(DE3) pLysS	pJ411	1 mM IPTG for overnight expression at 18°C in TB media
GST-3C-PTP1B(1-393)-6xHis	*E. coli* BL21(DE3) pLysS	pJ411	1 mM IPTG for overnight expression at 18°C in TB media
GST-3C-PTP1B(1-393)-6xHis	*E. coli* BL21(DE3)	pGex-6p-1	1 mM IPTG for overnight expression at 18°C in TB media
GST-3C-PTP1B(1-393)-6xHis	*E. coli* BL21(DE3)	pET-24a	1 mM IPTG for overnight expression at 18°C in TB media

Table 23. The expression conditions for various tau K18 constructs.

Construct	Cell line	Vector	Conditions
6xHis-tau K18(C291S,C322S)	*E. coli* BL21(DE3)	pET-24a	0.4 mM IPTG for 4 hour expression at 37°C in LB media
6xHis-tau K18(C291, C322S)	*E. coli* BL21(DE3)	pET-24a	0.4 mM IPTG for 4 hour expression at 37°C in LB media
6xHis-tau K18(C291S,C322)	*E. coli* BL21(DE3)	pET-24a	0.4 mM IPTG for 4 hour expression at 37°C in LB media
6xHis-tau K18 (C291,C322)	*E. coli* BL21(DE3)	pET-24a	0.4 mM IPTG for 4 hour expression at 37°C in LB media
AviTag-tau K18(C291S, C322S)	*E. coli* BL21(DE3)	pET-24a	0.4 mM IPTG for 4 hour expression at 37°C in LB media

7.3.2   Expression in M9 media

7.3.2.1   M9 media preparation

The following components were dissolved in 954 mL of dH$_2$O, pH adjusted to 7.0 with NaOH/HCl and autoclaved:

Components	MW	Final concentration
K$_2$HPO$_4$ (anhydrous)	174.2 g/mol	60.85 mM (10.6 g/L)
KH$_2$PO$_4$ (anhydrous)	136.08 g/mol	36.05 mM (4.9 g/L)
NaCl	58.44 g/mol	8.55 mM (0.5 g/L)
NH$_4$Cl	53.48 g/mol (54.48 g/mol for ^{15}N)	18.35 mM (1 g/L)

The mixture was then supplemented with the following sterile-filtered component mixture to bring the total volume to 1000 mL:

Supplement components	Volume added, mL	Final concentration
25% w/v D-Glucose (or ^{13}C-D-Glucose) in dH$_2$O	40	1% w/v (0.2% for ^{13}C-labelled)
1 M MgSO$_4$	2	2 mM
0.1 M CaCl$_2$	1	0.1 mM
1000x Trace metal mix	1	1x
1000x Vitamin mix	1	1x
Antibiotics	1	100 µg/mL Amp and/or 50 µg/mL Kan and/or 34 µg/mL Cam (see **§8.2** for plasmid antibiotic resistances)

1x Trace metal mix:

Compound	1000x concentration, g/L
FeCl$_2$ x 4H$_2$O	10
CaCl$_2$ x 2H$_2$O	0.368
H$_3$BO$_3$	0.128
CoCl$_2$ x 6H$_2$O	0.036

CuCl$_2$ x 2H$_2$O	0.008
ZnCl$_2$	0.68
Na$_2$MoO$_4$ x 2H$_2$O	1.21
MnCl$_2$ x 4H$_2$O	0.08

1x Vitamin mix:

Vitamin	1000x concentration, g/L
Choline chloride	1
Folic acid	1
Pantothenic acid	1
Nicotinamide	1
Pyridoxal HCl	1
Thiamine HCl	5
Riboflavin	0.1
Biotin	1

7.3.2.2 Expression protocol

A discrete colony transformed with plasmid of interest was used to inoculate 15 mL of LB media with appropriate antibiotics and allowed to grow for 6 hour at 37°C with continuous agitation at 200 RPM. Then, an appropriate volume of cultured cells was centrifuged at 4000 x g for 15 min in order to give starting OD$_{600}$ of 0.05 when resuspended in 50 mL of fresh M9 media with supplements by using formula below:

$$V(mL) = \frac{0.05}{OD_{600,measured}} * V(final\ volume, mL)$$

This was done in order to increase the isotopic labelling efficiency by preventing the carry-over of non-labelled material from LB growth media. Following this, the cells were incubated overnight at 37°C with continuous agitation at 200 RPM. On the subsequent day, an appropriate volume of cells using previously mentioned formula was centrifuged at 4000 g for 15 min and transferred to 1 L flasks to give starting OD$_{600}$ of 0.05. The cells were then grown at 37°C with continuous agitation at 200 RPM until OD$_{600}$ of 0.8 was reached and induced with appropriate amount IPTG for timed expression. The cells were collected using the same protocol as mentioned in non-labelled media expression section (**§7.3.1**).

HASPA and tau K18 labelled proteins were expressed in M9 media under certain conditions (**Table 24, Table 25**).

Table 24. *The expression conditions for HASPA construct.*

Construct	Cell line	Vector	Labelling	Conditions
HASPA-6xHis	*E. coli* BL21(DE3)	pET-28b	^{15}N	1 mM IPTG for 6 hours at 30°C in M9 media

Table 25. *The expression conditions for various tau K18 constructs.*

Construct	Cell line	Vector	Labelling	Conditions
6xHis-tau K18(C291S,C322S)	*E. coli* BL21(DE3)	pET-24a	^{15}N ^{15}N,^{13}C	0.4 mM IPTG for 4 hour expression at 37°C in M9 media
6xHis-tau K18 (C291,C322)	*E. coli* BL21(DE3)	pET-24a	^{15}N	0.4 mM IPTG for 4 hour expression at 37°C in M9 media

^{19}F labelling of tau K18

In order to express 3-fluoro-Y310 labelled tau K18(C291S,C322S), the cell culture was prepared as described for M9 media. Then, 1 g of glyphosate was added to 1 L of culture together with 60 mg of L-Trp, L-Phe and 3-fluoro-DL-Tyr before induction at OD$_{600}$ of 0.8 [267]. The glyphosate inhibits aromatic amino acid synthesis pathway and the cells use the manually provided aromatic amino acids. The culture was further incubated for 1 hour at 37°C, 200 RPM and expression was induced with 0.4 mM IPTG for expression for 4 hours at 37°C.

7.3.3 Protein purification

All purification steps were performed using AKTA Explorer (GE Healthcare) equipment.

7.3.3.1 HASPA

The cells were thawed on ice and resuspended in ice cold Buffer A (1:3 ratio w/v of cell paste to Buffer A) (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM Imidazole) together with appropriate amount of EDTA-free Roche cOmplete protease inhibitor tablets (Roche) which was followed by two cycles of sonication on wet ice (3 seconds on, 7 seconds off). The cell lysate was clarified by centrifugation at 40000 g with JA 25.50 rotor for 40 min at 4°C. The clarified supernatant was loaded onto a 5 mL HisTrap HP column, equilibrated with Buffer A, and HASPA was eluted using Buffer B gradient to 100% (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 500 mM imidazole) over 20 CV. The QC was carried out using SDS-PAGE on 15% agarose gel. The protein-containing fractions were concentrated to ~3 mL using Amicon 3K MWCO concentrator. The concentrated sample was then loaded 6 times as 500 µL injections onto Superdex 10/300 GL 75pg that was pre-equilibrated with Buffer C (20 mM HEPES pH 6.5, 50

mM NaCl) at 0.5 mL/min. The separation was done at 0.5 mL/min and pure fractions were concentrated using Amicon 3K MWCO concentrator, snapfrozen in liquid N_2 and stored at -80°C. Concentration was determined using with OPA assay (**§7.3.8**).

7.3.3.2   cMyc

Soluble cMyc fusion constructs

The cell pellets were thawed on ice and resuspended in cold Buffer A (1:3 ratio w/v of cell paste to lysis buffer) (50 mM HEPES pH 8.0, 500 mM NaCl, 5% glycerol, 5 mM $MgCl_2$). Additionally, EDTA-Free Protease inhibitor Cocktail Tablets (Roche) and 250 µL of DNAse1 (5 mg/mL in PBS, Sigma-Aldrich) were added. The mixture was stirred for 30 min at 4°C and passed through French press homogenizer (Stansted) twice. The clarification of sample was achieved by centrifugation for 1 hour at 40000 g using JA-25.50 rotor at 4°C. The supernatant was loaded on 5 mL HisTrap FF column at 5 mL/min which was pre-equilibrated with Buffer A using AKTA Explorer system (GE Healthcare). The protein was eluted using 3-step gradient (5, 20 and 100%) to Buffer B (25 mM HEPES pH 8.0, 500 mM NaCl, 5% glycerol, 5 mM $MgCl_2$, 500 mM Imidazole). Pure fractions were concentrated using Amicon Ultra-15 3.5K MWCO concentrator (Merck Millipore) to 10 mL and further purified at 2.6 mL/min on pre-equilibrated with Buffer C (25 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol, 5 mM $MgCl_2$) HiLoad Superdex 26/60 75 pg column. All samples were evaluated using SDS-PAGE after each step. Concentration of the protein was determined using A_{280} with ε = 13410 1/M*cm (**§7.3.6**).

The pure fusion protein samples were desalted to EK cleavage buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM $CaCl_2$, 0.5 mM TCEP, 5 mM $MgCl_2$) using HiPrep Desalt 26/10 column at 6 mL/min and a recommended amount of EK (NEB, UK) (1 unit per 25 µg of protein) was added to the samples for overnight cleavage at RT. The cleavage efficiency was checked using HPLC-MS assay (**§7.4.1**)

Insoluble cMyc constructs with KSI

The cell pellets were thawed on ice and resuspended in cold Buffer A (1:3 ratio w/v of cell paste to lysis buffer) (10 mM Sodium phosphate pH 7.4, 137 mM NaCl, 2.7 mM KCl, 1 mM EDTA). Additionally, EDTA-Free Protease inhibitor Cocktail Tablets (Roche) and 250 µL of DNAse1 (5 mg/mL, Sigma-Aldrich) were added. The mixture was stirred for 30 min at 4°C. The sample was passed through French press homogenizer (Stansted) twice. The inclusion bodies

were collected by centrifugation for 15 minutes at 10000 g using JA-25.50 rotor at 4°C. The pellets were further washed with:

- Buffer B: 10 mM Sodium phosphate pH 7.4, 137 mM NaCl, 2.7 mM KCl
- Buffer C: 10 mM Sodium phosphate pH 7.4, 1 M NaCl, 2.7 mM KCl
- Buffer D: 10 mM Sodium phosphate pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.2% Tween-20

The pellets were solubilized in 5 mL of Buffer E (PBS buffer solution pH 7.4, 6 M Gdn-HCl). The concentration was determined using A_{280} with ε = 2980 1/M*cm for cMyc(363-381) and ε = 4470 1/M*cm for cMyc(402-412) constructs.

The solubilized cMyc pellets in Buffer E were transferred to a round-bottom flask with a magnetic stirrer in fumehood. Then, 98% formic acid was added to the solution to amount for 70% of the total volume. This was followed by addition of 100 molar excess over methionine residues of 5 M CNBr in acetonitrile (Sigma). The sample was stirred overnight at RT in dark. The cleavage efficiency was checked using HPLC-MS assay (**§7.4.1**)

7.3.3.3   PTP1B
PTP1B(1-301)

The cells were thawed on ice and mixed in ice cold Lysis Buffer (1:3 ratio w/v of cell paste to Lysis Buffer) (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM Imidazole, 0.5 mM MTG). An appropriate amount of Roche cOmplete protease inhibitor tablets (Roche) (1 tablet per 50 mL of lysate) were also added together with 500 μL of 2 mg/mL DNAseI (Sigma). The mixture was left to stir at 4°C for 30 min. The cells were then homogenized using Stansted French press system. Following this, the cell lysate was clarified by centrifugation at 40000 g with JA 25.50 rotor for 40 min at 4°C. The clarified supernatant was loaded onto a 5 mL HisTrap FF column, equilibrated with Buffer A (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.5 mM MTG) and PTP1B(1-301) was eluted using three-step gradient to Buffer B (5, 20, 100%) (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 500 mM imidazole, 0.5 mM MTG). After QC by SDS-PAGE, fractions with acceptable purity were desalted to Buffer C (50 mM Tris-HCl pH 7.5, 25 mM NaCl, 0.5 mM MTG) using HiPrep Desalt 26/10 column at 6 mL/min and loaded on pre-equilibrated with Buffer C anion exchange column MonoQ 10/100 GL at 4 mL/min. PTP1B(1-301) was eluted at 2 mL/min using gradient over 20 CV from 0 to 50% of Buffer D (50 mM Tris-HCl pH 7.5, 1 M

NaCl, 0.5 mM MTG). Pure fractions were pooled, concentrated with Amicon 10K MWCO concentrator to 15 mL and loaded 3 times in total as 5 mL loads on HiLoad 16/60 Superdex 75pg column at 1 mL/min, pre-equilibrated with Buffer E (50 mM HEPES pH 6.8, 150 mM NaCl, 0.5 mM TCEP). Pure fractions were pooled, concentrated with Amicon 10K MWCO concentrator before assessing protein concentration using A_{280} with ε = 46940 1/M*cm. Samples were aliquoted, snap frozen in liquid N_2 and stored at -80°C. The intact mass was checked using HPLC-MS assay (**§7.4.1**)

PTP1B(1-393)

The cells were thawed on ice and mixed in ice cold Lysis Buffer (1:3 ratio of cell paste to Lysis Buffer) (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM Imidazole, 0.1% Triton X-100, 0.5 mM DTT). An appropriate amount of Roche cOmplete protease inhibitor tablets (Roche) (1 tablet per 50 mL of lysate) were also added together with 500 µL of 2 mg/mL DNAseI (Sigma). The mixture was left to stir at 4°C for 30 min. The cells were then homogenized using Stansted French press system. Following this, the cell lysate was clarified by centrifugation at 40000 g with JA 25.50 rotor for 40 min at 4°C. The clarified supernatant was loaded onto a 5 mL HisTrap FF column twice, equilibrated with Buffer A (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.5 mM DTT) and PTP1B(1-393) was eluted using three-step gradient to Buffer B (5, 20, 100%) (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM imidazole, 0.5 mM DTT). After QC by SDS-PAGE, fractions with acceptable purity were desalted to Buffer C (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) using HiPrep Desalt 26/10 column at 6 mL/min and 150 µL of PreScission protease (GE Healthcare) was added to the mixture for overnight incubation at 4°C. After QC which confirmed cleavage of GST tag, the sample was loaded onto pre-equilibrated with Buffer C 5 mL GSTrap HP column at 1 mL/min. Cleaved PTP1B(1-393) was found in flow-through fractions whereas contaminants were found in the 100% elution fractions with Buffer D (25 mM Tris HCl pH 8.0, 150 mM NaCl, 10 mM reduced glutathione, 1 mM DTT). Following this, PTP1B(1-393) was desalted to Buffer E (25 mM Tris-HCl pH 7.5, 25 mM NaCl, 1 mM DTT) using HiPrep Desalt 26/10 at 6 mL/min and loaded on MonoQ 10/100 GL column at 4 mL/min. The protein was eluted using gradient from 0 to 30% over 12 CV with Buffer F (25 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM DTT). Pure fractions were concentrated using Amicon 10K MWCO concentrator to 10 mL and further purified with pre-equilibrated with Buffer F (25 mM HEPES pH 6.8, 150 mM NaCl, 2 mM DTT) HiLoad 26/60 75pg column at 2

mL/min. After QC, the protein concentration was determined using A_{280} with ε = 53400 1/M*cm. Samples were aliquoted, snap frozen in liquid N_2 and stored at -80°C. The intact mass was checked using HPLC-MS assay (**§7.4.1**)

7.3.3.4 Tau K18

Non-labelled, isotopically and ^{19}F-labelled tau K18

Cells were thawed on ice and resuspended in ice-cold Buffer A (1:3 ratio w/v of cell paste to Buffer A) (50 mM NaPi pH 7.5, 500 mM NaCl, 10 mM imidazole. The mixture was supplemented with cOmplete Protease Inhibitor Cocktail EDTA-free tablets (1 tablet per 50 mL of lysate), 500 uL of 2 mg/mL DNAse1 in PBS (Sigma-Aldrich) and stirred for 30 min at 4°C. Then, the mixture was homogenized using Stansted French press system and incubated at 90°C for 15 min under mild stirring. Following this, the lysate was cooled down in ice bath and further clarified by centrifugation at 4°C using JA 25.50 rotor at 40000 g for 1 hour. The supernatant was then loaded onto 5 mL HisTrap FF column (GE Healthcare) pre-equilibrated with Buffer B (25 mM NaPi pH 7.5, 500 mM NaCl, 10 mM imidazole) at 5 mL/min. The protein was eluted using 3-step gradient (5, 20 and 100%) of Buffer C (25 mM NaPi pH 7.5, 500 mM NaCl, 500 mM imidazole). After QC with SDS-PAGE, protein-containing fractions were pooled and desalted to Buffer D (25 mM NaPi pH 7.0, 25 mM NaCl, 1 mM EDTA) using HiPrep Desalt 26/10 column at 6 mL/min. Following this, tau K18 was loaded on 5 mL HiTrap SP FF column at 4 mL/min for cation exchange step. The protein was eluted with gradient over 10 CV from 0 to 100% of Buffer D (25 mM NaPi pH 7.0, 1 M NaCl, 1 mM EDTA). Protein-containing fractions were pooled, concentrated using Amicon 3k MWCO concentrator to 5 mL and further purified via size-exclusion chromatography using HiPrep Superdex 16/60 75pg column at 1 mL/min, pre-equilibrated with Buffer E (25 mM NaPi pH 6.6, 150 mM NaCl, 100 µM EDTA). The pure samples were pooled, concentrated with Amicon MWCO 3K concentrator, snap frozen in liquid N_2 and stored at -80°C. Concentration was determined using Pierce BCA assay kit (ThermoScientific) (**§7.3.7**). The intact mass was checked using HPLC-MS assay (**§7.4.1**)

During purification of Cys-containing constructs all buffer solutions included 1 mM MTG. Final size exclusion buffer solution contained 2 mM DTT.

Avi-tagged tau K18

The cell lysate was prepared in the same manner as non-labelled tau K18 protein. The lysate was then loaded onto 1 mL HisTrap FF column (GE Healthcare) equilibrated with 25 mM NaPi pH 7.5, 500 mM NaCl, 10 mM imidazole. The protein was eluted using an imidazole gradient (10 - 500 mM). Protein-containing fractions were pooled and desalted to 25 mM NaPi pH 7.5, 500 mM NaCl, 10 mM imidazole using HiPrep 26/10 column at 6 mL/min. Then, 2.5 mg of TEV protease (in-house) was added to the solution to cleave the 6xHis tag at 4°C overnight. After QC, the sample was further purified by loading the cleavage mixture on new 1 mL HisTrap FF column equilibrated with 25 mM NaPi pH 7.5, 500 mM NaCl, 10 mM imidazole. After QC with SDS-PAGE, flow through and 5% elution fractions were pooled together, concentrated with Amicon MWCO 3K to 5 mL and further purified via size-exclusion chromatography using HiPrep Superdex 16/60 75pg column at 1 mL/min, equilibrated with 25 mM NaPi pH 6.6, 150 mM NaCl. The pure samples were pooled, desalted to 50 mM Bicine pH 8.3 using HiPrep 26/10 column and the protein concentration was adjusted to 40 μM (Pierce BCA assay, (**§7.3.7**)). This was followed by the addition of 10 mM ATP (from 100 mM stock dissolved in 50 mM Bicine pH 8.3), 100 μM D-biotin (from 50 mM stock in DMSO), 10 mM MgCl$_2$, and 5 μg of BirA (in-house) per 10 nmol of AviTag-tau K18(C291S, C322S). The mixture was incubated at room temperature and biotinylation process was followed by HPLC-MS (**§7.4.1**). After biotinylation, the sample was put over HiPrep Superdex 16/60 75 pg as 5 mL load at 1 mL/min, pre-equilibrated with 25 mM NaPi pH 6.6, 150 mM NaCl to remove BirA. Pure fractions were concentrated with Amicon MWCO 3K to 4.66 mg/mL, as determined by Pierce BCA assay (ThermoScientific), snap frozen in liquid N$_2$ and stored at -80°C. The intact mass was checked using HPLC-MS assay (**§7.4.1**)

7.3.4 Initial expression evaluation

Expression levels of proteins were evaluated using Maxwell 16 Polyhistidine Protein Purification Kit (Promega). Small scale expression of 10 mL is performed in LB media as described previously (**§7.3.1**). After expression, the sample is prepared according to the standard kit protocol. The method uses magnetic particles that bind His-tagged protein in the cell lysate which is later eluted with 500 mM imidazole. QC of the eluted sample via SDS-PAGE and HPLC-MS helps to evaluate if protein of interest is expressed intact and in high yields.

Transformed colonies are then stored as glycerol stocks in Pro-Lab Diagnostics™ Microbank™ tubes (Thermofisher) at -80°C. The intact mass was checked using HPLC-MS assay (**§7.4.1**)

7.3.5 SDS-PAGE

Bolt Bis-Tris Plus (4-12%) gels (Invitrogen) were used to run SDS-PAGE in Bolt MES SDS Running buffer (Invitrogen) for 20 min at 200 V. The samples for SDS-PAGE were prepared by incubating protein samples with Bolt LDS Sample Buffer (Invitrogen) at 95°C for 5 min.

7.3.6 Protein concentration determination using A_{280}

Protein concentration was determined by measuring absorbance of target protein at A_{280} and calculating the concentration using Beer-Lambert law:

$$A = \varepsilon * C * l$$

where A = absorbance, ε = extinction coefficient ($M^{-1}cm^{-1}$), C = concentration (M), l = path length (cm). Extinction coefficient ε is estimated using amino acid sequence with ProtParam tool (https://web.expasy.org/protparam/).

7.3.7 Protein concentration determination using BCA assay

If protein of interest has a low extinction coefficient, concentration was determined using Pierce BCA assay kit (Thermofisher). This method measured the concentration of a target protein sample by using BSA standard curve in linear range from 25 to 2000 µg/mL. The standard curve samples were made by diluting 2 mg/mL BSA stock solution to 1.5, 1.0, 0.75, 0.5, 0.25, 0.125 and 0.025 mg/mL using storage buffer of protein of interest. Storage buffer was also used as a blank. The target protein sample was diluted in storage buffer to 1:1, 1:5, 1:10 and 1:20. 25 µL of BSA and target protein samples were put into separate wells of clear round-bottom 96 well plate (Pierce). The samples were mixed with 200 µL of Working Solution which was prepared by mixing standard solutions from the kit as follows:

- 50 parts of Reagent A (sodium carbonate, sodium bicarbonate, bicinchoninic acid and sodium tartrate in 0.1 M sodium hydroxide)
- 1 part of Reagent B (4% cupric sulphate)

The plate was incubated at 37°C for 30 min. Following this, the plate was left to cool to room temperature for 5 min and the absorbance was measured at 562 nm. The data was processed with Microsoft Excel software using linear regression.

7.3.8 Protein concentration determination using OPA assay

For HASPA, the concentration was determined using *ortho*-phthalaldehyde (OPA) assay due to unusual amino acid composition (M. Plevin, personal communication). The concentration of a target protein sample is measured by using BSA standard curve in linear range from 50 to 500 µg/mL. BSA stock solution of 10 mg/mL was prepared in target protein buffer. The BSA stock solution was then diluted to 50, 100, 200, 300, 400 and 500 µg/mL. The stock HASPA sample was diluted in protein buffer to 3 dilutions (1:10, 1:100 and 1:1000). 20 µL of diluted BSA and HASPA samples were put into separate wells of black flat-bottom 96 well plate (Greiner). The samples were mixed with 200 µL of Phthaldialdehyde Reagent Solution Complete (Sigma) which was regenerated with 2 µL of 2-mercaptoethanol per mL of Reagent Solution. The Reagent Solution has pH of 10 with Brij-35 detergent which denatures proteins of interest to expose all residues to the solvent. This resulted in 6 standard points of BSA, 3 points of diluted protein samples and 1 well of target protein buffer to serve as a blank. After incubation of 2 minutes, the readings were performed on microplate reader (CLARIOstar) at 340 nm excitation and 455 nm emission (with 30 nm band width). The data was processed using MARS software using linear regression.

7.3.9 Peptide mass fingerprint

Peptide mass fingerprinting was performed as a commercial service by Cambridge Centre for Proteomics (Cambridge, UK).

7.3.10 Enzyme activity assay

Enzymatic activities of 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis were checked using EnzChek Phosphatase Assay Kit (Thermofisher) at [E] = 5 nM, 100 µM of substrate DiFMUP in 50 mM HEPES pH 6.9, 150 mM NaCl, 1 mM EDTA, 2 mM DTT and 1% DMSO. The reaction was followed using Corning black flat-bottomed 96 well plates in SynergyNeo plate reader with excitation at 358 nm and emission at 455 nm.

7.4 Biophysical methods

7.4.1 HPLC-MS/HRMS-ESI

Exact protein molecular mass was determined using Agilent HPLC 1290 series coupled to a 6230 TOF (positive ESI mode) with Agilent MassHunter software. The column (BioResolve RP mAB polyphenyl, 150 x 2.1 mm, 2.7 µm, Waters) was run under gradient condition at pH = 4

(5 to 95% A to B, A = 0.08% formic acid 0.01% TFA, B = MeCN+ 0.08% formic acid, 0.01% TFA) over 5 min at a rate of 0.8mL/min at 55°C).

LMW organic compound studies were performed on Agilent 1200 SL series instrument linked to an Agilent MSD 6140 single quadrupole with an ESI-APCI multimode source or Agilent 1290 Infinity II series instrument connected to an Agilent TOF 6230 with an ESI-jet stream source. Column: Thermo Accucore 2.6 μm, C18, 50 mm x 2.1 mm at 55 °C. Injection volume: 1 μL. Elution: 95% A / 5% B to 5% A / 95% B using HPLC-grade solvents.

- Solvent A: 10 mM aqueous ammonium formate + 0.04% (v/v) formic acid
- Solvent B: Acetonitrile + 5.3% (v/v) Solvent A + 0.04% (v/v) formic acid.

Time (min)	LC-MS-V-B1		Flow (mL/min)
	Solvent A (%)	Solvent B (%)	
0	95	5	1.1
0.12	95	5	1.3
1.30	5	95	1.3
1.35	5	95	1.6
1.85	5	95	1.6
1.90	5	95	1.3
1.95	95	5	1.3

7.4.1.1   Covalent binders to PTP1B
Samples of 20 μM 6xHis-PTP1B(1-301) and PTP1B(1-393)-6xHis containing 500 μM compound **V1** in 25 mM HEPES pH 6.9, 150 mM NaCl, 250 μM TCS401, 1 mM TCEP (or 1 mM DTT, or no reducing agent) were incubated for 18 hours at RT before MW evaluation using HPLC-MS.

7.4.1.2   Covalent binders to tau K18
Samples of 50 μM native 6xHis-tau K18(C291,C322) containing 250 μM compound **V1** in 50 mM NaPi pH 6.6, 25 mM NaCl, 50 μM DSS, 4% d_6-DMSO, 5% D_2O were incubated for 18 hours at RT before MW evaluation using HPLC-MS and NMR.

7.4.2   NMR
Data was processed using TopSpin 4.0.2 (Bruker) and MNova 11.0 (MestreLab Research). Sample purity, integrity and solubility of LMW organic compounds were assessed in d_6-DMSO or aqueous solutions using Bruker NMR spectrometers. Protein and protein:ligand studies were performed in aqueous solutions.

Protein, ligand QC and protein:ligand interaction experiments were performed with Bruker NMR equipment using one or more of the following pulse sequences:

- 1D ^{1}H: 1D sequence using 30 degree flip angle (for 1D ^{1}H NMR spectra acquired in d$_6$-DMSO and for automatic pulse width determination in aqueous samples)
    - PULPROG = zg30 (standard Bruker sequence)
- 1D ^{1}H: 1D sequence with water suppression using excitation sculpting (for 1D ^{1}H NMR spectra acquired in aqueous samples)
    - PULPROG = zgesgp (standard Bruker sequence)
- 2D ^{1}H-^{15}N SF-HMQC: 2D SOFAST-Heteronuclear Multiple Quantum Coherence with ^{1}H-^{15}N correlation
    - PULPROG = sfhmqcf3gpph (standard Bruker sequence)
- 1D ^{1}H wLOGSY: Water-Ligand Observed via Gradient SpectroscopY
    - PULPROG = ephogsygpno.2 (standard Bruker sequence) or bdLOGSYesgpdf.rh (in-house)
- 1D ^{1}H CPMG-filtered: T_2 measurement with CPMG and water suppression
    - PULPROG = cpmg_esgp2d (standard Bruker sequence) or bdcpmgesdf.rh (in-house)
- 1D ^{1}H STD: Saturation Transfer Difference Spectroscopy
    - PULPROG = bdSTDesgpdf.rh (in-house)
- 1D ^{19}F: ^{19}F spectra with proton decoupling
    - PULPROG = zgig30 (standard Bruker sequence)
- 1D ^{19}F broadband CPMG: ^{19}F T_2 filter with broadband CPMG sequence
    - PULPROG = bbCPMG_bb19F_dec_phase_opt.awp (in-house) or hk_19Fcpmg_screen_adia (in-house)
- 1D ^{19}F CPMG: ^{19}F T_2 filter with CPMG sequence
    - PULPROG = bd_zgigcpmg2d (in-house)
- 2D w1- ^{13}C,^{15}N-filtered,edited NOESY using WATERGATE
    - PULPROG = noesygpphwgx1 (standard Bruker sequence)
- 3D w2- ^{13}C,^{15}N-filtered,edited ^{1}H-^{13}C-HSQC-NOESY using WATERGATE
    - PULPROG = hsqcgpnowgx33d (standard Bruker sequence)
- 2D ^{1}H-^{13}C HSQC: Heteronuclear Single Quantum Coherence with ^{1}H-^{13}C correlation

- o PULPROG = Chsqc.rh (in-house)
- Pseudo 2D Diffusion Ordered Spectroscopy (DOSY)
    - o PULPROG = bd-ledbpgppr2s (in-house)
- 2D ^{1}H-^{1}H-Total Correlation Spectroscopy (TOCSY)
    - o PULPROG = dipsi2esgpph (standard Bruker sequence)
- 2D ^{1}H-^{1}H-Nuclear Overhauser Effect Spectroscopy (NOESY)
    - o PULPROG = noesyesgpph (standard Bruker sequence)

For in-house pulse sequences, see **§8.4.**

7.4.2.1   HASPA

HASPA samples were desalted to 50 mM NaPi pH 6.5, 50 mM NaCl using 3.5K MWCO D-tube dialyzer (Millipore) overnight at 4°C. All NMR experiments were performed using Bruker Avance Neo 700 MHz spectrometer with QCI cryoprobe at University of York at 298 K.

A total of 609 fragments were screened against ^{15}N-labelled HASPA in 5 mm NMR tubes, out of which 559 (93 mixtures of 6 fragments each at 33.33 mM in 100% d_6-DMSO) were a subset of in-house Maybridge fragment library and 51 3D fragments were kindly provided by Hanna Klein (ESR2, University of York). The signal of DSS at 0.0 ppm was used as a reference.

For NMR screening, 1D ^{1}H experiments were used for QC purposes. 1D ^{1}H CPMG-filtered experiments were used to evaluate resonance signal decays of the ligands at different delays. 2D ^{1}H-^{15}N SF-HMQC experiment was used to acquire 2D data and quickly evaluate the system for any specific CSPs. wLOGSY was used as an additional control experiment for putative hits against HASPA. All pulse sequences were used from standard Bruker pulse sequence catalogue for ligand- and protein-observed NMR screening. The following experimental parameters were used for NMR experiments:

Experiment	NS	SW, ppm	FID	Extra parameters	PULPROG
1D ^{1}H	64	15.5	32768	-	zgesgp
1D ^{1}H CPMG-filtered	128	16	42350	τ = variable	cpmg_esgp2d
1D ^{1}H wLOGSY	128	15.5	32768	-	ephogsygpno.2
2D ^{1}H-^{15}N SF-HMQC	4	F1 = 35 F2 = 16.2	F1 = 512 F2 = 2048	NUS = 50%	sfhmqcf3gpph

Initial fragment reference experiments were acquired for the following samples:

- 500 µM fragment mixture, 50 µM DSS, 50 mM NaPi pH 6.5, 50 mM NaCl, 2% d_6-DMSO, 5% D_2O
  - 1D ^{1}H NMR
  - 1D ^{1}H CPMG
- 25 µM ^{15}N-labelled HASPA, 50 µM DSS, 50 mM NaPi pH 6.5, 50 mM NaCl, 2% d_6-DMSO, 5% D_2O
  - 2D ^{1}H-^{15}N SF-HMQC
    - Up to 4 mM ethanolamine or acetic acid
    - Up to 5% d_6-DMSO

This was followed by fragment screening experiments at a single fragment concentration:

- 500 µM fragment mixture, 25 µM ^{15}N-labelled HASPA, 50 µM DSS, 50 mM NaPi pH 6.5, 50 mM NaCl, 2% d_6-DMSO, 5% D_2O
  - 1D ^{1}H NMR
  - 1D ^{1}H CPMG-filtered
  - 2D ^{1}H-^{15}N SF-HMQC

After data evaluation and hit prioritization, the potential hits were titrated up to 4 mM in the presence of HASPA and CSPs were evaluated using data from the following experiments:

- 25 µM ^{15}N-labelled HASPA, 50 µM DSS, 50 mM NaPi pH 6.5, 50 mM NaCl, 2% d_6-DMSO, 5% D_2O
  - 1D ^{1}H CPMG
  - 1D ^{1}H wLOGSY
  - 2D ^{1}H-^{15}N SF-HMQC

7.4.2.2   cMyc

The NMR experiments were performed in 5 mm NMR tubes with commercially synthesized peptides (Severn Biotech, UK) in 100 mM NaPi pH 6.6, 25 mM NaCl, 50 µM DSS, 1% d_6-DMSO, 5% D_2O using Bruker Avance HD III 600 MHz spectrometer at 298 K. Data was processed and evaluated using TopSpin 4.0.1 (Bruker).

The parameters used for NMR experiments were as follows:

Experiment	NS	SW, ppm	FID	Extra parameters	PULPROG
1D ^{1}H	128	16	32768	-	zgesgp
1D ^{1}H CPMG-filtered	64	16	32768	τ = variable	bdcpmgesdf.rh
1D ^{1}H wLOGSY	128	16	32768	-	bdLOGSYesgpdf.rh
1D ^{1}H STD	2048	16	32768	On-resonance = 0.7 ppm, off-resonance = -27.6 ppm	bdSTDesgpdf.rh
Pseudo 2D DOSY	32	F1 = 10, F2 = 20	F1 = 16 F2 = 2048	Δ = 20 ms, δ = 8 ms	bd-ledbpgppr2s
2D ^{1}H,^{1}H TOCSY	24	F1 = 10 F2 = 10	F1 = 512 F2 = 2048	Mixing time d9 = 60 and 90 ms	dipsi2esgpph
2D ^{1}H-^{1}H NOESY	48	F1 = 10 F2 = 10	F1 = 512 F2 = 2048	Mixing time d8 = 200 ms	noesyesgpph

Variable gradient strength (x-axis) for DOSY experiment was calculated following the formula:

$$G_i = \sqrt{\frac{i}{n}} * G_{max}$$

where n = gradient steps (16), i = index of gradient value, G_{max} = maximum gradient value (G/cm), G_i applied gradient strength (G/cm)

7.4.2.3   PTP1B

The screening samples for ^{19}F CPMG filtered experiments were prepared in 3 mm NMR tubes that contained 10 µM of protein 6xHis-PTP1B(1-301) or PTP1B(1-393)-6xHis in 25 mM HEPES pH 6.9, 150 mM NaCl, 250 µM TCS401, 1 mM TCEP, 100 µM DSS, 10% D_2O, 4% d_6-DMSO. TCS401 was used as an active site inhibitor to reduce potential hit rate [196]. Each cocktail mix contained 32 CF_3-fluorinated compounds in a total of 52 cocktails.

The ^{19}F ligand-observed CPMG-filtered NMR experiments were performed using Bruker AV-500 and AV-600 MHz spectrometers with QCI cryoprobe at 296 K. The data was evaluated by comparing 320 ms relaxation profile of the compounds in the presence of different proteins with reference spectra.

Experimental parameters used for primary ^{19}F CPMG-filtered NMR screening were as follows:

Experiment	NS	SW, ppm	FID	Extra parameters	PULPROG
1D ^{1}H	64	16	32768	-	zgesgp
^{19}F CPMG-filtered broadband	256	100	65536	τ = 20 and 320 ms	bbCPMG_bb19F_dec_phase_opt.awp

The confirmatory follow-up ^{19}F CPMG-filtered singleton experiments were performed in the same buffer solution as primary screen using Bruker 600 MHz at 296K. Additional stocks of putative hits in powder form were obtained from LifeChemicals (DE).

The protein- and ligand-observed ^{1}H CPMG-filtered and NMR DOSY experiments were performed in 5 mm NMR tubes in 25 mM HEPES pH 6.9, 150 mM NaCl, 250 µM TCS401, 1 mM TCEP (or 1 mM DTT), 100 µM DSS, 10% D_2O, 4% d_6-DMSO using Bruker 600 MHz at 298K.

The parameters used for ^{19}F CPMG-filtered and other NMR experiments were as follows:

Experiment	NS	SW, ppm	FID	Extra parameters	PULPROG
1D ^{1}H	64	16	32768	-	zgesgp
^{19}F CPMG-filtered broadband	256	180	131072	τ = 20 and 320 ms	hk_19Fcpmg_screen_adia
1D ^{1}H CPMG-filtered	64	16	32768	τ = 200 and 800 ms	bdcpmgesdf.rh
Pseudo 2D DOSY	32	F1 = 10, F2 = 20	F1 = 16, F2 = 2048	Δ = 50 ms, δ = 8 ms	bd-ledbpgppr2s

7.4.2.4   Tau K18

All NMR experiments were performed in 5 mm NMR tubes using Bruker 400, 600, 700 or 800 MHz spectrometers at 298 K. The parameters used for NMR experiments were as follows:

Experiment	NS	SW, ppm	FID	Extra parameters	PULPROG
1D ^{1}H	64	16	32768	-	zgesgp
1D ^{1}H CPMG-filtered	256	16	32768	τ = variable	bdcpmgesdf.rh
1D ^{1}H STD	128	16	32768	On-resonance = 0.7 ppm, off-resonance = -27.6 ppm	bdSTDesgpdf.rh
1D ^{1}H wLOGSY	32	16	32768	-	bdLOGSYesgpdf.rh
1D ^{19}F NMR	256	237.2	262144	-	zgig30
1D ^{19}F NMR CPMG-filtered	64	F1, F2 = 10	F1 = 8, F2 = 16384	τ = variable	bd_zgigcpmg2d
^{19}F CPMG-filtered broadband	128	F1 = 10, F2 = 180.72	F1 = 4, F2 = 131072	τ = variable	hk_19Fcpmg_screen_adia
Pseudo 2D DOSY	32	F1 = 10, F2 = 20	F1 = 16, F2 = 2048	Δ and δ = variable	bd-ledbpgppr2s
2D ^{1}H-^{15}N SF-HMQC	32	F1 = 31.8, F2 = 16	F1 = 128, F2 = 2048	NUS = 50%	sfhmqcf3gpph
2D ^{1}H-^{13}C HSQC	16	F1 = 31.8, F2 = 16	F1 = 128, F2 = 2048	NUS = 50%	Chsqc.rh
w2- ^{13}C,^{15}N-filtered/edited 3D ^{1}H-^{13}C-HSQC-NOESY	16	F1 = 80, F2 = 13.7, F3 = 13.7	F1 = 48, F2 = 128, F3 = 4096	J_{CH} = 140 Hz	hsqcgpnowgx33d

Compound reference and solubility NMR spectra were obtained in 50 mM NaPi pH 6.6, 25 mM NaCl, 100 μM DSS, 5% D_2O. All NMR experiments with tau K18 constructs were performed in 50 mM NaPi pH 6.6, 25 mM NaCl, 100 μM DSS, 5% D_2O, 4% d_6-DMSO, unless stated otherwise.

Theoretical diffusion coefficient was calculated using Stokes–Einstein Gierer-Wirtz Estimation (SEGWE) method [268].

^{1}H and ^{19}F T_2 relaxation values were calculated by fitting relaxation data to the following formula using Dynamics Center 2.5.4 (Bruker):

$$I = I_0 * \exp(-\frac{t}{T_2})$$

where I = relative integral of signal, I_0 = relative amplitude of signal, t = applied delay time in ms, T_2 = relaxation value in ms

7.4.3   MST

The experiments were performed on Monolith NT Automated (NanoTemper) machine with the following parameters:

- MST power = 40 %
- Excitation power = 15 %
- Temperature = 25°C
- Acquisition mode = Pico-Red

The data was processed using MO.Affinity Analysis v2.2.4 (NanoTemper). The change in fluorescence was used to fit the data to either 1:1 K_D model or Hills equation.

The affinity of RED-Tris-NTA dye to 6xHis-tag was determined by preparing 12 samples which contained 5 nM of the dye and His-tagged protein from 0.01 to 200 nM in appropriate buffer solution. The protein concentrations were prepared as two-fold dilutions. After preparation, the samples were incubated for 30 min, loaded into premium-coated capillaries and measured as described above.

For protein:ligand interaction measurements, 6xHis-tagged protein was first labelled with 5 nM RED-Tris dye. The protein concentration used for labelling was 10 times the value of the K_D determined for the RED-Tris-NTA dye:protein interaction, as described above. By having a

large excess of protein in solution, all of the dye was bound to the protein which enabled accurate measurements. Following this, an appropriate number of the labelled protein samples with the ligand were prepared by mixing 1 μL of ligand in d_6-DMSO with 24 μL of labelled protein. The samples were incubated for 30 min, loaded into premium-coated capillaries and measured as described above.

For non-labelled approach, the concentration of fluorescent ligand was kept constant in the presence of different concentrations of the protein. After sample preparation, the samples were loaded into premium-coated capillaries, incubated for 30 min and measured as described above.

7.4.3.1 PTP1B

The affinity of RED-Tris dye to PTP1B(1-393)-6xHis was determined as described previously in assay buffer 25 mM HEPES pH 7.4, 150 mM NaCl, 1 mM DTT, 0.05% Tween 20 (**§7.4.3**).

In order to measure interactions between 6xHis-tau K18(C291S,C322S) and ligands, 5 nM of RED-Tris dye was incubated with 50 nM of protein for 30 min in assay buffer. Then, the labelled protein samples with 12 different concentrations of ligand were prepared, incubated for 30 min and measured. All final samples contained 50 nM of labelled protein in 25 mM HEPES pH 7.4, 150 mM NaCl, 1 mM DTT, 0.05% Tween and 4% DMSO. Measurements were performed with MST power of 40% and excitation power of 20%.

7.4.3.2 Tau K18

The affinity of RED-Tris dye to 6xHis-tau K18(C291S, C322S) was determined as described previously in assay buffer 50 mM HEPES pH 7.4, 150 mM NaCl, 0.05% Tween 20 (**§7.4.3**).

In order to measure interactions between 6xHis-tau K18(C291S,C322S) and ligands, 5 nM of RED-Tris dye was incubated with 200 nM of protein for 30 min in assay buffer. Then, the labelled protein samples with 12 different concentrations of ligand were prepared, incubated for 30 min and measured. All final samples contained 200 nM of labelled protein in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.05% Tween and 4% DMSO. Measurements were performed with MST power of 40% and excitation power of 20%.

For non-labelled approach, 24 samples were prepared from 1000 to 5.9 nM of 6xHis-tau K18(C291S,C322S) as 1.25x dilutions in the presence of 100 nM of compound **1.**

Experimental data was analysed using Biacore Insight Evaluation Software (GE Healthcare) and SimFit (https://www.simfit.org.uk/). Experiments were performed on Biacore T200 and Biacore 8K+.

Immobilization assays were performed as follows:

- For amine coupling, the surface of CM5 S Series chip was activated using 1:1 mixture of 400 mM EDC+ 100 mM NHS for 420 s at 10 µL/min at 25°C. This was followed by injection of 25 µg/mL target protein in its immobilization buffer for appropriate time at 10 µL/min. The surface was then deactivated with 1 M ethanolamine for 420 s at 10 µL/min. The reference flow cell was activated and deactivated using the same protocol but without protein. Eight startup cycles were used for stabilizing the surface after the immobilization

- For biotin-SA coupling, the surface of SA S Series chip was first conditioned with three consecutive 1 min injections of 1 M NaCl + 50 mM NaOH at 10 µL/min at 25°C. This was followed by injection of 100 nM of biotinylated protein in appropriate buffer solution at 5 µL/min until immobilization level of 1000 RU. Eight startup cycles were used for stabilizing the surface after the immobilization

- For His tag coupling, the surface of Ni-NTA S Series chip was loaded with 0.5 mM NiCl$_2$ for 1 min at 10 µL/min at 25°C. Then, 200 nM target protein was injected over the surface at 10 µL/min until immobilization level of 1000 RU in appropriate buffer solution. After kinetic and affinity studies, the surface was regenerated by injection of 350 mM EDTA for 1 minute at 10 µL/min. Eight startup cycles were used for stabilizing the surface after the immobilization

Affinity and kinetic measurements were performed as follows:

- Single-cycle kinetics experiments were performed at 5 different concentrations with appropriate association and dissociation times at 30 µL/min at 25°C in appropriate running buffer. Blank injections were also used referencing. Solvent correction was performed with 8 point samples at appropriate DMSO concentrations. The flow system was washed with 50% DMSO after each cycle. Additional regeneration and washing procedures were used as necessary.

- Multi-cycle kinetics experiments were performed using 8 different concentrations with appropriate association and dissociation times at 30 µL/min at 25°C in appropriate running buffer. Blank injections were also used for referencing. Solvent correction was performed with 8 point samples at appropriate DMSO concentrations. The flow system was washed with 50% DMSO after each cycle. Additional regeneration and washing procedures were used as necessary.

For fragment screening campaigns, the immobilization levels were aimed to produce R_{max} values between 20 and 40 RU for 150 Da fragment, assuming 100% ligand activity and 1:1 binding behaviour. The immobilization levels were calculated using the following formula:

$$R_{max} = \frac{Immobilization\ level\ of\ ligand, RU}{MW\ of\ ligand, Da} * 150\ Da * 1(valency)$$

The workflow of fragment screening included Clean Screen (CS), Binding Level Screen (BLS) and Affinity Screen (AS). These assays were used as follows:

- **CS**: after target immobilization, the injections of 500 µM fragments were performed over reference and protein-immobilized surfaces with 10 s contact time and 0 s of dissociation time at 15°C at 30 µL/min in Running Buffer (**RB**). The cut-off was set to 10 RU for evaluating N vs N_{i-1} baseline changes. The flow system was washed with 50% DMSO after each cycle.

- **BLS**: after target immobilization, the injections of 500 µM fragments were carried out over reference and the immobilized protein surface with 30 s contact time and 15 s of dissociation time at 15°C at 30 µL/min in **RB**. Solvent correction was performed with 8 point samples at appropriate DMSO concentrations. The flow system was washed with 50% DMSO after each cycle. Default binding behaviour settings (slope, slow dissociation, R>R_{max}) were used for data evaluation. A standard 10% of hits with highest response and without binding behavior markers were automatically selected by the evaluation software.

- **AS**: after target immobilization, the injections at concentrations from 15.625 to 500 µM (6 in total) as two-fold dilutions were carried out over the reference and immobilized protein surface with 30 s contact time and 30 s of dissociation time at 15°C in **RB**. Two zero concentration injections were used for blank subtraction. Solvent correction was performed with 8 point samples at appropriate DMSO concentrations.

The flow system was washed with 50% DMSO after each cycle. The dose response curves were fitted using 1:1 binding model with free R_{max}.

7.4.4.1 PTP1B

Fragment screening

SPR screening campaign was performed using Biacore 8K+ with the following PTP1B constructs:

- 6xHis-PTP1B(1-301)
- PTP1B(1-393)-6xHis

Amine coupling was used to immobilize 6xHis-PTP1B(1-301) or (1-393)-6xHis on CM5 S Series chip to achieve R_{max} of 20-40 RU (**§7.4.4**). The immobilization buffer used was 10 mM sodium acetate pH 5.5, 1 mM DTT. Positive control assay with suramin was developed using multi-cycle kinetics approach (**§7.4.4**). The buffer solution used for the experiments contained 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1 mM DTT, 1% DMSO. This was followed by Clean Screen (CS), Binding Level Screen (BLS) and Affinity Screen (AS) assays. Running buffer of 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% DMSO was used for all fragment screening assays. Solvent correction was performed from 0.5 to 1.8%

Protein:ligand interactions

For IDP:ligand interactions, PTP1B(1-393)-6xHis was immobilized on Ni-NTA S Series chip to 1000 RU using Biacore T200, as described (**§7.4.4**). The running buffer of 25 mM HEPES pH 7.4, 150 mM NaCl, 0.05% Tween-20, 1% DMSO was used for multicycle kinetics assays. Solvent correction was performed from 0.5 to 1.8%

7.4.4.2 Tau K18

Fragment screening campaign

Amine coupling was used to immobilize 6xHis-tau K18(C291S,C322S) in 10 mM sodium borate pH 8.5 to achieve R_{max} of 20-40 RU. This was followed by CS, BLS and AS steps. Running buffer of 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% DMSO was used for all fragment screening assays. Solvent correction was performed from 0.5 to 1.8%

Protein:ligand interactions

Biotinylated AviTag-tau K18(C291S,C322S) was immobilized on SA S Series chip to 1000-2000 RU in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.05% Tween-20, as described previously (**§7.4.4**).

Single-cycle and multi-cycle kinetics experiments were performed in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.05% Tween-20 and 4% DMSO with solvent correction from 3.5 to 4.8%.

7.5   Organic synthesis

The following tau K18 compounds were synthesized and purified using flash chromatography:

- (S)- and (R)-Cl-NQTrp enantiomers

- (S)-Cl-NQTrp derivatives:

    - 5-fluoro, 5-chloro, 5-hydroxy, 5-methoxy, 5-cyano, 7-methyl and N-methyl

    - Ethyl ester

For synthesis protocols and QC data, see **§8.3**

8 Appendix

8.1 Protein construct sequences

8.1.1 HASPA

```
        10         20         30         40         50         60
GAYSTKDSAK EPQKRADNID TTTRSDEKDG IHVQESAGPV QENFGDAQEK NEDGHNVGDG

        70         80
ANGNEDGNDD QPKEHAAGNL EHHHHHH
```

Number of amino acids: 87

Molecular weight: 9445.66

Theoretical pI: 4.91

Figure 156. The amino acid sequence for HASPA with C-terminal 6xHis tag in pET-28b vector.

8.1.2 cMyc

```
        10         20         30         40         50         60
MHHHHHHENL YFQGSTEYKL VVVGADGVGK SALTIQLIQN HFVDEYDPTI EDSYRKQVVI

        70         80         90        100        110        120
DGETCLLDIL DTAGQEEYSA MRDQYMRTGE GFLCVFAINN TKSFEDIHHY REQIKRVKDS

       130        140        150        160        170        180
EDVPMVLVGN KSDLPSRTVD TKQAQDLARS YGIPFIETSA KTRQGVDDAF YTLVREIRKH

       190        200
KEKDDDDKER QRRNELKRSF FALRDQI
```

Number of amino acids: 207

Molecular weight: 24082.88

Theoretical pI: 5.68

Figure 157. The amino acid sequence for 6xHis-TEV-KRas-EK-cMyc(363-381) in pJ411 expression vector. Cyan - 6xHis tag, green - EK cleavage sequence, yellow – cMyc(363-381) peptide.

```
             10         20         30         40         50         60
    MHHHHHHENL YFQGSTEYKL VVVGADGVGK SALTIQLIQN HFVDEYDPTI EDSYRKQVVI

             70         80         90        100        110        120
    DGETCLLDIL DTAGQEEYSA MRDQYMRTGE GFLCVFAINN TKSFEDIHHY REQIKRVKDS

            130        140        150        160        170        180
    EDVPMVLVGN KSDLPSRTVD TKQAQDLARS YGIPFIETSA KTRQGVDDAF YTLVREIRKH

            190
    KEKDDDDKYI LSVQAEEQK
```

Number of amino acids: 199

Molecular weight: 22927.54

Theoretical pI: 5.23

Figure 158. The amino acid sequence for 6xHis-TEV-KRas-EK-cMyc(402-412) in pJ411 expression vector. Cyan - 6xHis tag, green - EK cleavage sequence, yellow – cMyc(402-412) peptide.

```
             10         20         30         40         50         60
    MHTPEHITAV VQRFVAALNA GDLDGIVALF ADDATVEDPV GSEPRSGTAA IREFYANSLK

             70         80         90        100        110        120
    LPLAVELTQE VRAVANEAAF AFTVSFEYQG RKTVVAPIDH FRFNGAGKVV SIRALFGEKN

            130        140
    IHACQMERQR RNELKRSFFA LRDQI
```

Number of amino acids: 145

Molecular weight: 16048.24

Theoretical pI: 6.42

Figure 159. The amino acid sequence for KSI-cMyc(363-381) in pET-31b expression vector. Yellow – cMyc(363-381) peptide, Red - methionines.

```
             10         20         30         40         50         60
    MHTPEHITAV VQRFVAALNA GDLDGIVALF ADDATVEDPV GSEPRSGTAA IREFYANSLK

             70         80         90        100        110        120
    LPLAVELTQE VRAVANEAAF AFTVSFEYQG RKTVVAPIDH FRFNGAGKVV SIRALFGEKN

            130
    IHACQMYILS VQAEEQK
```

Number of amino acids: 137

Molecular weight: 14892.91

Theoretical pI: 5.30

Figure 160. The amino acid sequence for KSI-cMyc(402-412) in pET-31b expression vector. Yellow – cMyc(402-412) peptide, Red - methionines.

8.1.3 PTP1B

```
              10         20         30         40         50         60
MGSDKIHHHH HHENLYFQGH MASMEMEKEF EQIDKSGSWA AIYQDIRHEA SDFPCRVAKL

              70         80         90        100        110        120
PKNKNRNRYR DVSPFDHSRI KLHQEDNDYI NASLIKMEEA QRSYILTQGP LPNTCGHFWE

             130        140        150        160        170        180
MVWEQKSRGV VMLNRVMEKG SLKCAQYWPQ KEEKEMIFED TNLKLTLISE DIKSYYTVRQ

             190        200        210        220        230        240
LELENLTTQE TREILHFHYT TWPDFGVPES PASFLNFLFK VRESGSLSPE HGPVVVHCSA

             250        260        270        280        290        300
GIGRSGTFCL ADTCLLLMDK RKDPSSVDIK KVLLEMRKFR MGLIQTADQL RFSYLAVIEG

             310        320
AKFIMGDSSV QDQWKELSHE DLEPHN
```

Number of amino acids: 326

Molecular weight: 37996.17

Theoretical pI: 6.10

Figure 161. *6xHis-TEV-PTP1B(1-301) in pET-28b vector. Cyan - 6xHis tag, yellow - TEV cleavage sequence.*

```
              10         20         30         40         50         60
MGSDKIHHHH HHENLYFQGH MASMEMEKEF EQIDKSGSWA AIYQDIRHEA SDFPCRVAKL

              70         80         90        100        110        120
PKNKNRNRYR DVSPFDHSRI KLHQEDNDYI NASLIKMEEA QRSYILTQGP LPNTCGHFWE

             130        140        150        160        170        180
MVWEQKSRGV VMLNRVMEKG SLKCAQYWPQ KEEKEMIFED TNLKLTLISE DIKSYYTVRQ

             190        200        210        220        230        240
LELENLTTQE TREILHFHYT TWPDFGVPES PASFLNFLFK VRESGSLSPE HGPVVVHCSA

             250        260        270        280        290        300
GIGRSGTFCL ADTCLLLMDK RKDPSSVDIK KVLLEMRKFR MGLIQTADQL RFSYLAVIEG

             310        320        330        340        350        360
AKFIMGDSSV QDQWKELSHE DLEPPPEHIP PPPRPPKRIL EPHNGKCREF FPNHQWVKEE

             370        380        390        400        410
TQEDKDCPIK EEKGSPLNAA PYGIESMSQD TEVRSRVVGG SLRGAQAASP AKGEPS
```

Number of amino acids: 416

Molecular weight: 47875.27

Theoretical pI: 6.11

Figure 162. *6xHis-TEV-PTP1B(1-393) construct in pET-28b vector. Cyan - 6xHis tag, yellow - TEV cleavage sequence.*

```
            10         20         30         40         50         60
MEMEKEFEQI DKSGSWAAIY QDIRHEASDF PCRVAKLPKN KNRNRYRDVS PFDHSRIKLH

            70         80         90        100        110        120
QEDNDYINAS LIKMEEAQRS YILTQGPLPN TCGHFWEMVW EQKSRGVVML NRVMEKGSLK

           130        140        150        160        170        180
CAQYWPQKEE KEMIFEDTNL KLTLISEDIK SYYTVRQLEL ENLTTQETRE ILHFHYTTWP

           190        200        210        220        230        240
DFGVPESPAS FLNFLFKVRE SGSLSPEHGP VVVHCSAGIG RSGTFCLADT CLLLMDKRKD

           250        260        270        280        290        300
PSSVDIKKVL LEMRKFRMGL IQTADQLRFS YLAVIEGAKF IMGDSSVQDQ WKELSHEDLE

           310        320        330        340        350        360
PPPEHIPPPP RPPKRILEPH NGKCREFFPN HQWVKEETQE DKDCPIKEEK GSPLNAAPYG

           370        380        390
IESMSQDTEV RSRVVGGSLR GAQAASPAKG EPSHHHHHH

Number of amino acids: 399

Molecular weight: 45965.12

Theoretical pI: 6.15
```

Figure 163. PTP1B(1-393)-6xHis construct in pJ411 vector. Cyan - 6xHis tag.

```
            10         20         30         40         50         60
MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI

            70         80         90        100        110        120
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGE

           130        140        150        160        170        180
NLYFQGSMEM EKEFEQIDKS GSWAAIYQDI RHEASDFPCR VAKLPKNKNR NRYRDVSPFD

           190        200        210        220        230        240
HSRIKLHQED NDYINASLIK MEEAQRSYIL TQGPLPNTCG HFWEMVWEQK SRGVVMLNRV

           250        260        270        280        290        300
MEKGSLKCAQ YWPQKEEKEM IFEDTNLKLT LISEDIKSYY TVRQLELENL TTQETREILH

           310        320        330        340        350        360
FHYTTWPDFG VPESPASFLN FLFKVRESGS LSPEHGPVVV HCSAGIGRSG TFCLADTCLL

           370        380        390        400        410        420
LMDKRKDPSS VDIKKVLLEM RKFRMGLIQT ADQLRFSYLA VIEGAKFIMG DSSVQDQWKE

           430        440        450        460        470        480
LSHEDLEPPP EHIPPPPRPP KRILEPHNGK CREFFPNHQW VKEETQEDKD CPIKEEKGSP

           490        500        510        520
LNAAPYGIES MSQDTEVRSR VVGGSLRGAQ AASPAKGEPS HHHHHH

Number of amino acids: 526

Molecular weight: 60302.09

Theoretical pI: 6.02
```

Figure 164. 6xHis-Thr-SUMO-TEV-PTP1B(1-393)-6xHis construct in pJ411 vector. Cyan - 6xHis tag.

```
           10         20         30         40         50         60
MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY QGKLTVAKLN

           70         80         90        100        110        120
IDQNPGTAPK YGIRGIPTLL LFKNGEVAAT KVGALSKGQL KEFLDANLAG SGSGHMHHHH

          130        140        150        160        170        180
HHSSGENLYF QGSMEMEKEF EQIDKSGSWA AIYQDIRHEA SDFPCRVAKL PKNKNRNRYR

          190        200        210        220        230        240
DVSPFDHSRI KLHQEDNDYI NASLIKMEEA QRSYILTQGP LPNTCGHFWE MVWEQKSRGV

          250        260        270        280        290        300
VMLNRVMEKG SLKCAQYWPQ KEEKEMIFED TNLKLTLISE DIKSYYTVRQ LELENLTTQE

          310        320        330        340        350        360
TREILHFHYT TWPDFGVPES PASFLNFLFK VRESGSLSPE HGPVVVHCSA GIGRSGTFCL

          370        380        390        400        410        420
ADTCLLLMDK RKDPSSVDIK KVLLEMRKFR MGLIQTADQL RFSYLAVIEG AKFIMGDSSV

          430        440        450        460        470        480
QDQWKELSHE DLEPPPEHIP PPPRPPKRIL EPHNGKCREF FPNHQWVKEE TQEDKDCPIK

          490        500        510        520        530
EEKGSPLNAA PYGIESMSQD TEVRSRVVGG SLRGAQAASP AKGEPSHHHH HH
```

Number of amino acids: 532

Molecular weight: 60360.42

Theoretical pI: 6.01

Figure 165. *Trx-TEV-PTP1B(1-393)-6xHis construct in pJ411 vector. Cyan - 6xHis tag, yellow - TEV cleavage sequence.*

```
          10         20         30         40         50         60
MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ VAATGDGPDI

          70         80         90        100        110        120
IFWAHDRFGG YAQSGLLAEI TPDKAFQDKL YPFTWDAVRY NGKLIAYPIA VEALSLIYNK

         130        140        150        160        170        180
DLLPNPPKTW EEIPALDKEL KAKGKSALMF NLQEPYFTWP LIAADGGYAF KYAAGKYDIK

         190        200        210        220        230        240
DVGVDNAGAK AGLTFLVDLI KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK

         250        260        270        280        290        300
VNYGVTVLPT FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG LEAVNKDKPL

         310        320        330        340        350        360
GAVALKSYEE ELAKDPRIAA TMENAQKGEI MPNIPQMSAF WYAVRTAVIN AASGRQTVDE

         370        380        390        400        410        420
ALKDAQTENL YFQGSMEMEK EFEQIDKSGS WAAIYQDIRH EASDFPCRVA KLPKNKNRNR

         430        440        450        460        470        480
YRDVSPFDHS RIKLHQEDND YINASLIKME EAQRSYILTQ GPLPNTCGHF WEMVWEQKSR

         490        500        510        520        530        540
GVVMLNRVME KGSLKCAQYW PQKEEKEMIF EDTNLKLTLI SEDIKSYYTV RQLELENLTT

         550        560        570        580        590        600
QETREILHFH YTTWPDFGVP ESPASFLNFL FKVRESGSLS PEHGPVVVHC SAGIGRSGTF

         610        620        630        640        650        660
CLADTCLLLM DKRKDPSSVD IKKVLLEMRK FRMGLIQTAD QLRFSYLAVI EGAKFIMGDS

         670        680        690        700        710        720
SVQDQWKELS HEDLEPPPEH IPPPPRPPKR ILEPHNGKCR EFFPNHQWVK EETQEDKDCP

         730        740        750        760        770
IKEEKGSPLN AAPYGIESMS QDTEVRSRVV GGSLRGAQAA SPAKGEPSHH HHHH
```

Number of amino acids: 774

Molecular weight: 87124.92

Theoretical pI: 5.71

Figure 166. MBP-TEV-PTP1B(1-393)-6xHis construct in pJ411 vector. Cyan - 6xHis tag, yellow - TEV cleavage sequence.

```
        10         20         30         40         50         60
MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL EFPNLPYYID

        70         80         90        100        110        120
GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL DIRYGVSRIA YSKDFETLKV

       130        140        150        160        170        180
DFLSKLPEML KMFEDRLCHK TYLNGDHVTH PDFMLYDALD VVLYMDPMCL DAFPKLVCFK

       190        200        210        220        230        240
KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LEVLFQGPGS MEMEKEFEQI

       250        260        270        280        290        300
DKSGSWAAIY QDIRHEASDF PCRVAKLPKN KNRNRYRDVS PFDHSRIKLH QEDNDYINAS

       310        320        330        340        350        360
LIKMEEAQRS YILTQGPLPN TCGHFWEMVW EQKSRGVVML NRVMEKGSLK CAQYWPQKEE

       370        380        390        400        410        420
KEMIFEDTNL KLTLISEDIK SYYTVRQLEL ENLTTQETRE ILHFHYTTWP DFGVPESPAS

       430        440        450        460        470        480
FLNFLFKVRE SGSLSPEHGP VVVHCSAGIG RSGTFCLADT CLLLMDKRKD PSSVDIKKVL

       490        500        510        520        530        540
LEMRKFRMGL IQTADQLRFS YLAVIEGAKF IMGDSSVQDQ WKELSHEDLE PPPEHIPPPP

       550        560        570        580        590        600
RPPKRILEPH NGKCREFFPN HQWVKEETQE DKDCPIKEEK GSPLNAAPYG IESMSQDTEV

       610        620
RSRVVGGSLR GAQAASPAKG EPSHHHHHH
```

Number of amino acids: 629

Molecular weight: 72676.12

Theoretical pI: 6.03

Figure 167. *GST-3C-PTP1B(1-393)-6xHis construct in pJ411 vector. Cyan - 6xHis tag, yellow – 3C cleavage sequence.*

```
        10         20         30         40         50         60
MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL EFPNLPYYID

        70         80         90        100        110        120
GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL DIRYGVSRIA YSKDFETLKV

       130        140        150        160        170        180
DFLSKLPEML KMFEDRLCHK TYLNGDHVTH PDFMLYDALD VVLYMDPMCL DAFPKLVCFK

       190        200        210        220        230        240
KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LEVLFQGPLG SMEMEKEFEQ

       250        260        270        280        290        300
IDKSGSWAAI YQDIRHEASD FPCRVAKLPK NKNRNRYRDV SPFDHSRIKL HQEDNDYINA

       310        320        330        340        350        360
SLIKMEEAQR SYILTQGPLP NTCGHFWEMV WEQKSRGVVM LNRVMEKGSL KCAQYWPQKE

       370        380        390        400        410        420
EKEMIFEDTN LKLTLISEDI KSYYTVRQLE LENLTTQETR EILHFHYTTW PDFGVPESPA

       430        440        450        460        470        480
SFLNFLFKVR ESGSLSPEHG PVVVHCSAGI GRSGTFCLAD TCLLLMDKRK DPSSVDIKKV

       490        500        510        520        530        540
LLEMRKFRMG LIQTADQLRF SYLAVIEGAK FIMGDSSVQD QWKELSHEDL EPPPEHIPPP

       550        560        570        580        590        600
PRPPKRILEP HNGKCREFFP NHQWVKEETQ EDKDCPIKEE KGSPLNAAPY GIESMSQDTE

       610        620        630
VRSRVVGGSL RGAQAASPAK GEPSHHHHHH
```

Number of amino acids: 630

Molecular weight: 72789.28

Theoretical pI: 6.03

Figure 168. GST-3C-PTP1B(1-393)-6xHis in pGex-6p-1 vector. Cyan - 6xHis tag, yellow – 3C cleavage sequence.

196

```
            10         20         30         40         50         60
MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL EFPNLPYYID

            70         80         90        100        110        120
GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL DIRYGVSRIA YSKDFETLKV

           130        140        150        160        170        180
DFLSKLPEML KMFEDRLCHK TYLNGDHVTH PDFMLYDALD VVLYMDPMCL DAFPKLVCFK

           190        200        210        220        230        240
KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LEVLFQGPGS MEMEKEFEQI

           250        260        270        280        290        300
DKSGSWAAIY QDIRHEASDF PCRVAKLPKN KNRNRYRDVS PFDHSRIKLH QEDNDYINAS

           310        320        330        340        350        360
LIKMEEAQRS YILTQGPLPN TCGHFWEMVW EQKSRGVVML NRVMEKGSLK CAQYWPQKEE

           370        380        390        400        410        420
KEMIFEDTNL KLTLISEDIK SYYTVRQLEL ENLTTQETRE ILHFHYTTWP DFGVPESPAS

           430        440        450        460        470        480
FLNFLFKVRE SGSLSPEHGP VVVHCSAGIG RSGTFCLADT CLLLMDKRKD PSSVDIKKVL

           490        500        510        520        530        540
LEMRKFRMGL IQTADQLRFS YLAVIEGAKF IMGDSSVQDQ WKELSHEDLE PPPEHIPPPP

           550        560        570        580        590        600
RPPKRILEPH NGKCREFFPN HQWVKEETQE DKDCPIKEEK GSPLNAAPYG IESMSQDTEV

           610        620
RSRVVGGSLR GAQAASPAKG EPSHHHHHH
```

Number of amino acids: 629

Molecular weight: 72676.12

Theoretical pI: 6.03

Figure 169. GST-3C-PTP1B(1-393)-6xHis in pET-24a vector. Cyan - 6xHis tag, yellow – 3C cleavage sequence.

8.1.4   Tau K18

6xHis-Tau K18(C291S/C322S)

MGHHHHHHQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKSGSKDNIKHVPGGGSVQIVYK
PVDLSKVTSKSGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIE

6xHis-Tau K18(C291S)

MGHHHHHHQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKSGSKDNIKHVPGGGSVQIVYK
PVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIE

6xHis-Tau K18(C322S)

MGHHHHHHQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK
PVDLSKVTSKSGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIE

6xHis-Tau K18

MGHHHHHHQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK
PVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIE

6xHis-TEV-AviTag-Tau K18(C291S/C322S)

MHHHHHHENLYFQGSGLNDIFEAQKIEWHEGSQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNV
QSKSGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKSGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPG
GGNKKIE

Figure 170. *The amino acid sequences for tau K18 constructs.* **Cyan** *- 6xHis tag,* **yellow** *- Cys and Ser mutations. N-terminal Met marked in* **grey***.*

8.2 Expression vectors

Xho I(158)
Not I(166)
Eag I(166)
Hind III(173)
Sal I(179)
Sac I(190)
EcoR I(192)
BamH I(198)

Sty I(57)
Bpu1102 I(80)

Nhe I(231)
Nde I(238)

Xba I(276)
Bgl II(342)
SgrA I(383)
Sph I(539)

Dra III(5068)

f1 origin (4844-5299)

Pvu I(4367)
Sgf I(4367)

Sma I(4241)

Kan (3936-4748)

Cla I(4058)
Nru I(4024)

pET-24a(+)
(5310bp)

Mlu I(1064)
Bcl I(1078)

lacI (714-1793)

BstE II(1245)
Apa I(1275)

BssH II(1475)
EcoR V(1514)
Hpa I(1570)

Eco57 I(3713)

AlwN I(3581)

ori (3227)

PshA I(1909)

BssS I(3338)

BspLU11 I(3165)

Sap I(3049)

Bst1107 I(2936)

Tth111 I(2910)

Bgl I(2128)
Fsp I(2146)
Psp5 II(2171)

T7 promoter primer #69348-3

Bgl II   T7 promoter   lac operator   Xba I   rbs

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA

Nde I Nhe I   T7•Tag   pET-24a   BamH I EcoR I   Sac I   Sal I Hind III Eag I   Xho I   His•Tag
                                                                                      Not I

TATACATATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA
      MetAlaSerMetThrGlyGlyGlnGlnMetGlyArgGlySerGluPheGluLeuArgArgGlnAlaCysGlyArgThrArgAlaProProProProProLeu

pET-24d   pET-24b   ...GGTCGGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA
Nco I   ...GlyArgAspProAsnSerSerSerValAspLysLeuAlaAlaAlaLeuGluHisHisHisHisHisHisEnd
...TACCATGGCTAGC...
      MetAlaSer...   pET-24c,d ...GGTCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA
                                       ...GlyArgIleArgIleArgAlaProSerThrSerLeuArgProHisSerSerThrThrThrThrThrThrGlu

Bpu1102 I   T7 terminator

GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG

T7 terminator primer #69337-3

pET-24a-d(+) cloning/expression region

Figure 171. The vector map of pET-24a. (Novagen)

199

Xho I(158)
Not I(166)
Eag I(166)
Hind III(173)
Sal I(179)
Sac I(190)
EcoR I(192)
BamH I(198)
Nhe I(231)
Nde I(238)
Nco I(296)

Bpu1102 I(80)

Xba I(335)
Bgl II(401)
SgrA I(442)
Sph I(598)

Dra III(5127)

f1 origin (4903-5358)

Pvu I(4426)
Sgf I(4426)

Sma I(4300)

Kan (3995-4807)

Cla I(4117)
Nru I(4083)

pET-28a(+)
(5369bp)

Mlu I(1123)
Bcl I(1137)

lacI (773-1852)

BstE II(1304)

Apa I(1334)

Eco57 I(3772)

AlwN I(3640)

BssH II(1534)
EcoR V(1573)
Hpa I(1629)

ori (3286)

BssS I(3397)

BspLU11 I(3224)
Sap I(3108)
Bst1107 I(2995)
Tth111 I(2969)

PshA I(1968)

Bgl I(2187)
Fsp I(2205)
Psp5 II(2230)

T7 promoter primer #69348-3

pET upstream primer #69214-3
Bgl II
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA
T7 promoter   lac operator   Xba I   rbs

Nco I   His·Tag   Nde I   Nhe I   T7·Tag
TATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACAGCAA
       MetGlySerSerHisHisHisHisHisHisSerSerGlyLeuValProArgGlySerHisMetAlaSerMetThrGlyGlyGlnGln

                                                              Eag I
          BamH I EcoR I   Sac I   Sal I   Hind III Not I   Xho I   thrombin His·Tag
ATGGGTCGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCC   pET-28a(+)
MetGlyArgGlySerGluPheGluLeuArgArgGlnAlaCysGlyArgThrArgAlaProProProProProLeuArgSerGlyCysEnd

...GGTCGGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCC   pET-28b(+)
...GlyArgAspProAsnSerSerSerValAspLysLeuAlaAlaAlaLeuGluHisHisHisHisHisHisEnd

...GGTCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCC   pET-28c(+)
...GlyArgIleArgIleArgAlaProSerThrSerLeuArgProHisSerSerThrThrThrThrThrGluIleArgLeuLeuThrLysPro...

Bpu1102 I   T7 terminator
GAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG

T7 terminator primer #69337-3

pET-28a-c(+) cloning/expression region

Figure 172. The vector map of pET-28b. (Novagen)

pET-31b(+) cloning/expression region

Figure 173. *The vector map of pET-31b(+). (Novagen)*

Figure 174. The vector map of pJ411. (AtumBio)

8.3 Organic synthesis

All solvents and reagents were used as obtained from commercial vendors. Purification steps were carried out using HPLC-grade solvents.
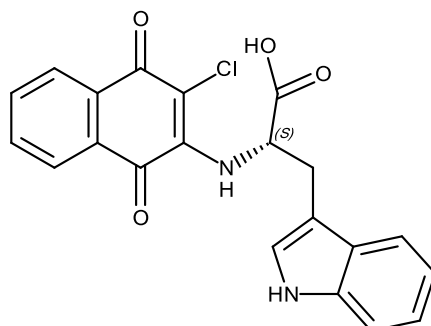
Compound 6(6A)

To a stirred solution of 2,3-dichloro-1,4-naphthoquinone (113.52 mg, 0.5 mmol) in 19 mL absolute EtOH was added dropwise (2S)-2-amino-3-(1H-indol-3-yl)propanoic acid (62.5 mg, 0.3 mmol) which was previously dissolved in 1 mL of 1 M KOH in distilled H_2O. The mixture was stirred at 30°C overnight with a mounted condenser. After incubation, the sample was acidified dropwise by addition of 2 M HCl until pH = 2, as determined with pH paper indicator. Following this, the compound was purified by automated flash chromatography (Combiflash Rf, Silica 40g RediSep column) eluting with 0 to 100% ethyl acetate in heptane. Pure fractions, as determined by HPLC-MS, were further purified by reverse phase automated flash chromatography (Combiflash Rf, C18 50g Gold RediSep column) eluting with 10 to 100% acetonitrile in pH 4 water. Freeze-drying yielded red powder of **(2S)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(1H-indol-3-yl)propanoic acid.**

Yield: 67.35 mg or 55.74%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d$_6$) δ 13.35 (s, 1H), 10.96 (d, J = 2.4 Hz, 1H), 8.00 – 7.93 (m, 1H), 7.89 (d, J = 7.7 Hz, 1H), 7.83 (td, J = 7.5, 1.4 Hz, 1H), 7.75 (td, J = 7.5, 1.4 Hz, 1H), 7.51 (d, J = 7.9 Hz, 1H), 7.30 (d, J = 8.1 Hz, 1H), 7.22 (d, J = 2.4 Hz, 1H), 7.04 (ddd, J = 8.2, 7.0, 1.2 Hz, 1H), 6.98 – 6.89 (m, 1H), 6.58 (d, J = 24.1 Hz, 1H), 5.33 (d, J = 7.2 Hz, 1H), 3.40 (t, J = 5.2 Hz, 2H). **HPLC-MS**: ($C_{21}H_{15}ClN_2O_4$) 394.98 [M+H]$^+$; RT 1.083. **HRMS-ESI:** calculated for $C_{21}H_{15}ClN_2O_4$: 394.072, found 394.0745

The following derivatives were prepared using analogous protocol.
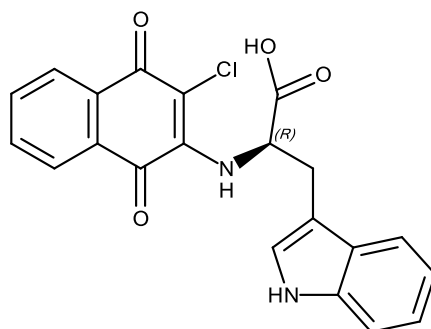
R-enantiomer of Compound 6A

(2R)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(1H-indol-3-yl)propanoic acid. Obtained from 2,3-dichloro-1,4-naphthoquinone and (2R)-2-amino-3-(1H-indol-3-yl)propanoic acid. **Yield:** 55 mg or 46.51%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d6) δ 13.33 (s, 1H), 10.97 (d, J = 2.5 Hz, 1H), 7.96 (d, J = 7.6 Hz, 1H), 7.89 (d, J = 7.5 Hz, 1H), 7.83 (td, J = 7.5, 1.4 Hz, 1H), 7.74 (td, J = 7.5, 1.4 Hz, 1H), 7.51 (d, J = 7.9 Hz, 1H), 7.33 − 7.26 (m, 1H), 7.22 (d, J = 2.4 Hz, 1H), 7.03 (ddd, J = 8.1, 7.0, 1.2 Hz, 1H), 6.98 − 6.89 (m, 1H), 6.56 (s, 1H), 5.33 (d, J = 7.3 Hz, 1H), 3.40 (t, J = 5.6 Hz, 2H). **HPLC-MS:** ($C_{21}H_{15}ClN_2O_4$) 394.98 [M+H]$^+$; RT 1.105. **HRMS-ESI:** calculated for $C_{21}H_{15}ClN_2O_4$: 394.077, found 394.0752.
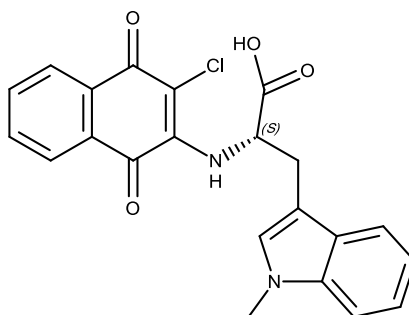
Compound 6B

(2S)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(1-methylindol-3-yl)propanoic acid. Obtained from 2,3-dichloro-1,4-naphthoquinone and (2S)-2-amino-3-(1-methylindol-3-yl)propanoic acid. **Yield**: 45 mg or 36.63%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d6) δ 13.31 (s, 1H), 7.87 (d, J = 7.6 Hz, 1H), 7.75 (td, J = 7.4, 1.4 Hz, 1H), 7.72 (s, 1H), 7.66 (td, J = 7.5, 1.3 Hz, 1H), 7.43 (d, J = 7.9 Hz, 1H), 7.22 (d, J = 8.2 Hz, 1H), 7.10 (s, 1H), 7.05 − 6.97 (m, 1H), 6.94 − 6.85 (m, 1H), 6.50 (s, 1H), 5.26 (dt, J = 8.4, 5.8 Hz, 1H), 3.57 (s, 3H), 3.30 (d, J = 5.7 Hz, 2H). **HPLC-MS**: ($C_{22}H_{17}ClN_2O_4$) 408.84 [M+H]$^+$ RT 1.22. **HRMS-ESI**: calculated for $C_{22}H_{17}ClN_2O_4$: 408.0877, found: 408.0894.
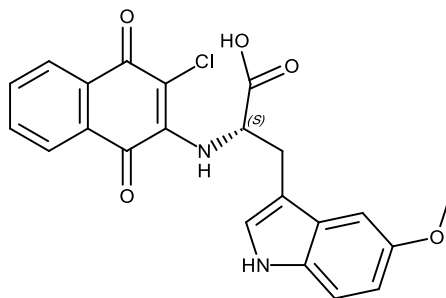
Compound 6C

(2S)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(5-methoxy-1H-indol-3-yl) propanoic acid. Obtained from 2,3-dichloro-1,4-naphthoquinone and (2S)-2-amino-3-(5-methoxy-1H-indol-3-yl)propanoic acid**. Yield:** 73 mg or 57.14%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d$_6$) δ 13.35 (s, 1H), 10.80 (d, J = 2.6 Hz, 1H), 7.96 (d, J = 7.6 Hz, 1H), 7.88 (d, J = 7.6 Hz, 1H), 7.83 (td, J = 7.6, 1.4 Hz, 1H), 7.74 (td, J = 7.5, 1.4 Hz, 1H), 7.24 – 7.07 (m, 2H), 6.98 (d, J = 2.4 Hz, 1H), 6.65 (dd, J = 8.7, 2.4 Hz, 2H), 5.60 – 5.10 (m, 1H), 3.69 (s, 3H), 3.37 (d, J = 6.8 Hz, 2H). **HPLC-MS:** (C$_{22}$H$_{17}$ClN$_2$O$_5$) 424.99 [M+H]$^+$; RT 1.08. **HRMS-ESI:** calculated for C$_{22}$H$_{17}$ClN$_2$O$_5$: 424.0826, found 424.0855.
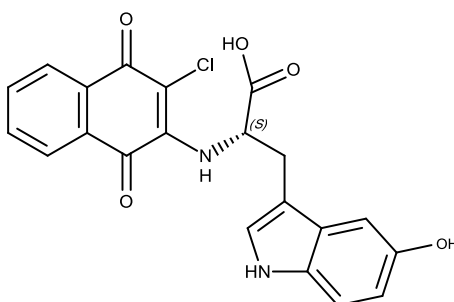
Compound 6D

(2S)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(5-hydroxy-1H-indol-3-yl)propanoic acid. Obtained from 2,3-dichloro-1,4-naphthoquinone and (2S)-2-amino-3-(5-hydroxy-1H-indol-3-yl)propanoic acid. **Yield**: 49.7 mg or 40.34%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d$_6$) δ 13.30 (s, 1H), 10.65 (d, J = 2.5 Hz, 1H), 8.62 (s, 1H), 7.96 (d, J = 7.6 Hz, 1H), 7.89 (s, 1H), 7.83 (td, J = 7.5, 1.4 Hz, 1H), 7.74 (td, J = 7.5, 1.4 Hz, 1H), 7.14 – 7.05 (m, 2H), 6.83 (d, J = 2.3 Hz, 1H), 6.57 (dd, J = 8.6, 2.3 Hz, 2H), 5.31 – 5.23 (m, 1H), 3.30 (d, J = 5.8 Hz, 2H). **HPLC-MS:** (C$_{21}$H$_{15}$ClN$_2$O$_5$) 410.97 [M+H]$^+$; RT 0.93. **HRMS-ESI:** calculated for C$_{21}$H$_{15}$ClN$_2$O$_5$: 410.0669, found 410.0736
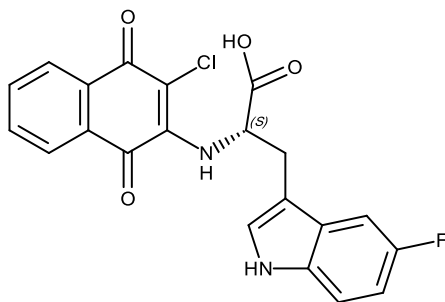
Compound 6E

(2S)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(5-fluoro-1H-indol-3-yl)propanoic acid. Obtained from 2,3-dichloro-1,4-naphthoquinone and (2S)-2-amino-3-(5-fluoro-1H-indol-3-yl)propanoic acid. **Yield:** 41.55 mg or 33.54%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d$_6$) δ 13.28 (s, 1H), 11.06 (d, J = 2.5 Hz, 1H), 7.96 (dd, J = 7.7, 1.3 Hz, 1H), 7.89 (d, J = 7.7 Hz, 1H), 7.83 (td, J = 7.5, 1.4 Hz, 1H), 7.75 (td, J = 7.5, 1.4 Hz, 1H), 7.29 (d, J = 2.7 Hz, 1H), 7.28 − 7.25 (m, 1H), 7.23 (dd, J = 10.1, 2.5 Hz, 1H), 6.86 (td, J = 9.2, 2.5 Hz, 1H), 6.60 (d, J = 36.0 Hz, 1H), 5.33 (q, J = 6.3 Hz, 1H), 3.41 (d, J = 14.9 Hz, 2H). **HPLC-MS:** (C$_{21}$H$_{14}$ClFN$_2$O$_4$) 412.99 [M+H]$^+$; RT 1.11. **HRMS-ESI:** calculated for C$_{21}$H$_{14}$ClFN$_2$O$_4$: 412.0626, found 412.0698.
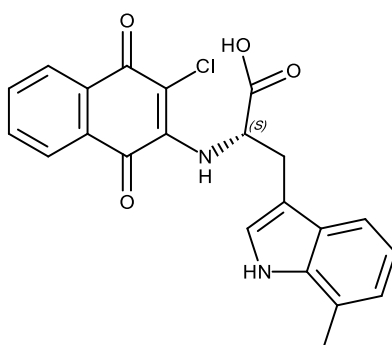
Compound 6F

(2S)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(7-methyl-1H-indol-3-yl)propanoic acid. Obtained from 2,3-dichloro-1,4-naphthoquinone and (2S)-2-amino-3-(7-methyl-1H-indol-3-yl)propanoic acid. **Yield:** 59 mg or 48.07%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d$_6$) δ 13.37 − 13.31 (m, 1H), 10.92 (d, J = 2.5 Hz, 1H), 7.95 (d, J = 7.7 Hz, 1H), 7.90 − 7.78 (m, 2H), 7.73 (td, J = 7.6, 1.3 Hz, 1H), 7.32 (dd, J = 7.3, 1.7 Hz, 1H), 7.19 (d, J = 2.5 Hz, 1H), 6.89 − 6.78 (m, 2H), 6.61 (s, 1H), 5.32 (d, J = 7.2 Hz, 1H), 3.38 (d, J = 5.7 Hz, 2H), 2.37 (s, 3H). **HPLC-MS:** (C$_{22}$H$_{17}$ClN$_2$O$_4$) 408.99 [M+H]$^+$; RT 1.16. **HRMS-ESI:** calculated for C$_{22}$H$_{17}$ClN$_2$O$_4$: 408.0877, found 408.0907
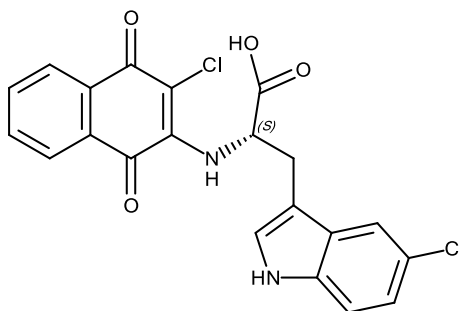
Compound 6G

(2S)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(5-chloro-1H-indol-3-yl)propanoic acid. Obtained from 2,3-dichloro-1,4-naphthoquinone and (2S)-2-amino-3-(5-chloro-1H-indol-3-yl)propanoic acid. **Yield:** 63.5 mg or 49.37%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d$_6$) δ 13.39 (s, 1H), 11.16 (d, J = 2.5 Hz, 1H), 7.96 (dd, J = 7.6, 1.3 Hz, 1H), 7.89 (d, J = 7.6 Hz, 1H), 7.83 (td, J = 7.5, 1.4 Hz, 1H), 7.75 (td, J = 7.5, 1.3 Hz, 1H), 7.51 (d, J = 2.1 Hz, 1H), 7.29 (d, J = 1.3 Hz, 1H), 7.28 (d, J = 4.9 Hz, 1H), 7.00 (dd, J = 8.6, 2.0 Hz, 1H), 6.61 (d, J = 45.3 Hz, 1H), 5.34 (d, J = 7.1 Hz, 1H), 3.45 − 3.35 (m, 2H). **HPLC-MS:** (C$_{21}$H$_{14}$Cl$_2$N$_2$O$_4$) 428.97 [M+H]$^+$; RT 1.16. **HRMS-ESI:** calculated for C$_{21}$H$_{14}$Cl$_2$N$_2$O$_4$: 428.0331, found 428.0403
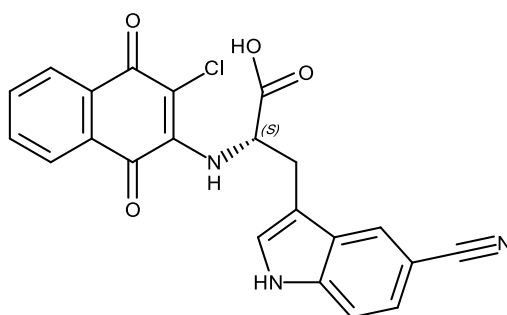
Compound 6H

(2S)-2-[(3-chloro-1,4-dioxonaphthalen-2-yl)amino]-3-(5-cyano-1H-indol-3-yl)propanoic acid. Obtained from 2,3-dichloro-1,4-naphthoquinone and (2S)-2-amino-3-(5-cyano-1H-indol-3-yl)propanoic acid. **Yield:** 69 mg or 54.94%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d$_6$) δ 13.41 (s, 1H), 11.53 (d, J = 2.4 Hz, 1H), 8.03 (d, J = 1.5 Hz, 1H), 7.96 (dd, J = 7.7, 1.3 Hz, 1H), 7.89 (d, J = 7.6 Hz, 1H), 7.83 (td, J = 7.5, 1.4 Hz, 1H), 7.75 (td, J = 7.5, 1.4 Hz, 1H), 7.47 − 7.41 (m, 2H), 7.35 (dd, J = 8.5, 1.5 Hz, 1H), 6.70 (s, 1H), 5.36 (q, J = 6.6 Hz, 1H), 3.53 − 3.38 (m, 2H). **HPLC-MS:** (C$_{22}$H$_{14}$ClN$_3$O$_4$) 419.99 [M+H]$^+$; RT 1.03. **HRMS-ESI:** calculated for C$_{22}$H$_{14}$ClN$_3$O$_4$: 419.0673, found 419.0701.
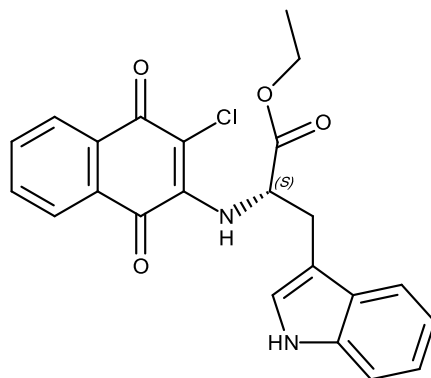
Ethyl ester of compound 6A

To a stirred solution of 2,3-dichloro-1,4-naphthoquinone (113.52 mg, 0.5 mmol) in 5 mL distilled H_2O was added ethyl (2S)-2-amino-3-(1H-indol-3-yl)propanoate (134.37 mg, 0.5 mmol) and trimethylamine (0.21 mL, 1.5 mmol). The mixture was stirred at 30°C for 3 hours with a mounted condenser. After incubation, the product was noted to aggregate on stirring bar, as determined by HPLC-MS. 7 mL of absolute EtOH was added to the mixture and stirred for additional 2 hours which additionally solubilized the reacted and unreacted material for better reaction efficiency. Following this, the compound was purified by automated flash chromatography (Combiflash Rf, Silica 40g RediSep column) eluting with 0 to 100% ethyl acetate in hexane. Pure fractions, as determined by HPLC-MS, were further purified by reverse phase automated flash chromatography (Combiflash Rf, C18 50g Gold RediSep column) eluting with 10 to 100% acetonitrile in water. Freeze-drying yielded orange-red powder of ethyl (2S)-2-[(3-chloro-1,4-dioxo-1,4-dihydronaphthalen-2-yl)amino]-3-(1H-indol-3-yl)propanoate.

Ethyl (2S)-2-[(3-chloro-1,4-dioxo-1,4-dihydronaphthalen-2-yl)amino]-3-(1H-indol-3-yl) propanoate. Obtained from 2,3-dichloro-1,4-naphthoquinone and ethyl (2S)-2-amino-3-(1H-indol-3-yl)propanoate. **Yield**: 135 mg or 64.1%. **NMR:** ^{1}H NMR (399 MHz, DMSO-d$_6$) δ 11.00 (d, J = 2.5 Hz, 1H), 7.96 (dd, J = 7.5, 1.3 Hz, 1H), 7.90 (d, J = 7.6 Hz, 1H), 7.83 (td, J = 7.5, 1.4 Hz, 1H), 7.75 (td, J = 7.5, 1.4 Hz, 1H), 7.47 (d, J = 7.9 Hz, 1H), 7.31 (dt, J = 8.1, 0.9 Hz, 1H), 7.27 (d, J = 2.4 Hz, 1H), 7.05 (ddd, J = 8.2, 7.0, 1.2 Hz, 1H), 6.97 (ddd, J = 8.0, 7.0, 1.1 Hz, 1H), 6.64 (s, 1H), 5.34 (d, J = 7.0 Hz, 1H), 4.14 (qd, J = 7.1, 0.9 Hz, 2H), 3.49 – 3.36 (m, 2H), 1.18 (t, J = 7.1 Hz, 3H). **HPLC-MS**: ($C_{23}H_{19}ClN_2O_4$) 422.87 [M+H]$^+$; RT 1.38. **HRMS-ESI**: calculated for $C_{23}H_{19}ClN_2O_4$: 422.1033, found: 422.1034.

8.4 NMR pulse sequences

bdLOGSYesgpdf.rh

```
# 1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh"
;bdLOGSYesgpdf
;based on zgesgp
;water suppression using excitation sculpting with gradients
;T.-L. Hwang & A.J. Shaka, J. Magn. Reson., Series A 112 275-279 (1995)

prosol relations=<std_triple>

# 1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=


# 169 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.10.2.2 2013/08/30 09:43:33 ber Exp $
# 9 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh" 2

# 1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl  -  include file for Gradient Spectroscopy
;   for TCU3
;
;avance-version (07/01/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 27 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Grad.incl"
define list<gradient> EA=<EA>


# 31 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Grad.incl"
;$Id: Grad2.incl,v 1.14 2012/01/31 17:49:21 ber Exp $
# 10 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh" 2


"p2=p1*2"
"d12=20u"
"p18=d8*0.3"
"p19=d8*0.5"
"p20=d8*0.2"
# 1 "mc_line 18 file /opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh dc-measurement
inserted automatically"

   dccorr
# 18 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh"
# 18 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh"
1 ze
2 d12 setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
```

```
d1 pl1:f1
50u setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
(p1 ph1):f1              ;water selection
p16:gp3
d16 pl0:f1
(p28:sp28 ph2:r):f1
p16:gp3
d16 pl1:f1
(p1 ph1):f1             ;NOE transfer
p18:gp4
d16
(p2 ph1):f1
p19:gp5
d16
(p2 ph1):f1
p20:gp6
d16 pl0:f1
(p29:sp29 ph5:r):f1
d12 pl1:f1
(p1 ph6):f1            ;excitation sculpting
p16:gp1
d16 pl0:f1
(p12:sp1 ph7:r):f1
4u
d12 pl1:f1
(p2 ph8):f1
4u
p16:gp1
d16
50u
p16:gp2
d16 pl0:f1
(p12:sp1 ph9:r):f1
4u
d12 pl1:f1
(p2 ph10):f1
4u
p16:gp2
d16
go=2 ph31
wr #0
4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
exit

ph1=0
ph2=0 3 2 1
ph5=2 0
ph6=0
ph7=2
ph8=0
ph9=2 2 2 2 1 1 1 1 0 0 0 0 3 3 3 3
ph10=0 0 0 0 3 3 3 3 2 2 2 2 1 1 1 1
ph31=0 2 0 2 2 0 2 0

# 75 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh"
;pl0 : 120dB
;pl1 : f1 channel - power level for pulse (default)
```

;sp1 : excitation sculpting shape power level
;sp2 : water selection shaped 180 power level
;sp29 : flip back power level
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p12: excitation sculpting (Squa100.1000)   [2 msec]
;p28: water selection 180 degree pulse [25ms]
;p29: flip back pulse [4ms]
;p16: excitation sculpting gradients
;p17: water selection gradients
;p18: Tm/2 gradient
;d1 : relaxation delay; 1-5 * T1
;d8 : NOE mixing period
;d12: power switching   [20 usec]
;d16: gradient recovery
;NS: 8 * n
;DS: 4

;use gradient ratio: gp 1 : gp 2
;   31 :   11
99 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh"
;for z-only gradients:
;gpz1: 31% first excitation sculpting gradient
;gpz2: 11% second excitation sculpting gradient
;gpz3: 2.4% water selection gradients
;gpz4: 0.3% NOE maintenance gradient
;gpz5: 0.4% NOE maintenance gradient

;use gradient files:
;gpnam1: SINE.100 first excitation sculpting gradient
;gpnam2: SINE.100 second excitation sculpting gradient
;gpnam3: SINE.100 water selection gradients
;gpnam4: SQR.10 NOE maintenance gradient
;gpnam5: SQR.10 NOE maintenance gradient

114 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdLOGSYesgpdf.rh"
;$Id: zgesgp,v 1.3 2000/10/06 09:09:37 ber Exp $

bdcpmgesdf.rh

1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh"
;bdcpmgesdf
;avance-version (00/10/05)
;excitation sculpted 1D with CPMG filter

prosol relations=<std_triple>

1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

```
# 169 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.10.2.2 2013/08/30 09:43:33 ber Exp $
# 7 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh" 2

# 1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl  -  include file for Gradient Spectroscopy
;   for TCU3
;
;avance-version (07/01/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 27 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Grad.incl"
define list<gradient> EA=<EA>

# 31 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Grad.incl"
;$Id: Grad2.incl,v 1.14 2012/01/31 17:49:21 ber Exp $
# 8 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh" 2

# 11 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh"
"p2=p1*2"
"d12=20u"
"p5=d21-d16"

# 1 "mc_line 15 file /opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh dc-measurement inserted
automatically"

   dccorr
# 15 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh"
1 ze
2 d12 setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
 d1 pl1:f1 setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
 p1 ph1

3 p5:gp3
 d16
 (p2 ph6):f1
 p5:gp3
 d16
 lo to 3 times l1

 50u
 p16:gp1
 d16 pl0:f1
 (p12:sp1 ph2:r):f1
 4u
 d12 pl1:f1
# 34 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh"
 p2 ph3

 4u
 p16:gp1
 d16
 50u
```

```
  p16:gp2
  d16 pl0:f1
  (p12:sp1 ph4:r):f1
  4u
  d12 pl1:f1

  p2 ph5
# 48 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh"
  4u
  p16:gp2
  d16

  go=2 ph31
  wr #0
  4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
exit

# 58 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh"
ph1=0
ph2=0 1
ph3=2 3
ph4=0 0 1 1
ph5=2 2 3 3
ph6=1
ph31=0 2 2 0


;pl0 : 120dB
;pl1 : f1 channel - power level for pulse (default)
;sp1 : f1 channel - shaped pulse 180 degree
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p12: f1 channel - 180 degree shaped pulse (Squa100.1000)   [2 msec]
;p16: homospoil/gradient pulse
;d1 : relaxation delay; 1-5 * T1
;d12: delay for power switching                 [20 usec]
;d16: delay for homospoil/gradient recovery
;NS: 8 * n
;DS: 4
;l1: CPMG delay = 2*L1*D21
# 82 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh"
;use gradient ratio:   gp 1 : gp 2
;               31 :   11

;for z-only gradients:
;gpz1: 31%
;gpz2: 11%
;gpz3: cpmg trim gradient

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100
# 96 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bdcpmgesdf.rh"
;$Id: zgesgp,v 1.3 2000/10/06 09:09:37 ber Exp $
```

bdSTDesgpdf

```
# 1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh"
;bdSTDesgpdf
;avance-version (00/10/05)
;1D sequence
;water suppression using excitation sculpting with gradients
;T.-L. Hwang & A.J. Shaka, J. Magn. Reson.,
;   Series A 112 275-279 (1995)
;prosol compatible sort of


prosol relations=<std_triple>
# 1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=


# 169 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.9.8.3 2013/08/30 09:44:49 ber Exp $
# 13 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh" 2

# 1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl  -  include file for Gradient Spectroscopy
;   for TCU3
;
;avance-version (07/01/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 27 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl"
define list<gradient> EA=<EA>


# 31 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl"
;$Id: Grad2.incl,v 1.13.8.1 2012/01/31 17:56:17 ber Exp $
# 14 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh" 2


# 18 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh"
"p2=p1*2"
"d12=20u"
"l2=l1-1"

# 1 "mc_line 22 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh dc-measurement inserted
automatically"

   dccorr
# 22 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh"
1 ze
2 d12 setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
```

```
  d1 pl0:f1
  d12 fq1:f1          ;frequency switch
3 d20
  p28:sp28:f1 ph1:r      ;selective 180
  d20
  lo to 3 times l2
  d20
  p28:sp28:f1 ph1:r
  d20 pl1:f1 fq1:f1     ;frequency switch
  p1 ph1

  50u setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
  p16:gp1
  d16 pl0:f1
  (p12:sp1 ph2:r):f1
  4u
  d12 pl1:f1

  p2 ph3
# 44 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh"
  4u
  p16:gp1
  d16
  50u
  p16:gp2
  d16 pl0:f1
  (p12:sp1 ph4:r):f1
  4u
  d12 pl1:f1

  p2 ph5

  4u
  p16:gp2
  d16
# 60 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh"
  go=2 ph31
  wr #0
  d12 ip31*2 zd
  lo to 2 times l5
  4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
exit


ph1=0
ph2=0 1
ph3=2 3
ph4=0 0 1 1
ph5=2 2 3 3
ph6=1
ph31=0 2 2 0
# 77 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh"
;pl0 : 120dB
;pl1 : f1 channel - power level for pulse (default)
;sp28 : f1 channel - shaped pulse ca. 135 degree for saturation
;sp1 : excitation sculpting shaped pulse
;p1 : f1 channel -  90 degree high power pulse
```

;p2 : f1 channel - 180 degree high power pulse
;p28: f1 channel - saturation shaped pulse [25ms]
;p12: f1 channel - 180 degree shaped pulse (Squa100.1000)   [2 msec]
;p16: homospoil/gradient pulse
;d1 : relaxation delay; 1-5 * T1
;d12: delay for power switching   [20 usec]
;d16: delay for homospoil/gradient recovery
;NS: 8 * n
;DS: 4

;use gradient ratio: gp 1 : gp 2
;   31 :   11
96 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh"
;for z-only gradients:
;gpz1: 31%
;gpz2: 11%

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100

106 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bdSTDesgpdf.rh"
;$Id: zgesgp,v 1.3 2000/10/06 09:09:37 ber Exp $

bbCPMG_bb19F_dec_phase_opt.awp

1 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
;bbCPMG_bb19F_dec.awp
;19F T2 relaxation with broadband CPMG sequence
;1D sequence for F-19 observe with inverse gated H-1 decoupling
;for QNP-operation (F-19 via X-QNP output of switchbox)
;avance-version (02/05/31)
;CF 25.4.05
;AOF 30/06/2015: BURBOP Broadband Refocusing Pulse Implemented
;BURBOP pulses from Tony Reinsperger and Burkhard Luy (KIT)
;AWP 04/12/2015: BURBOP 90 Degree Pulse Implemented

1 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance2.incl
;   for TCU3
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

145 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance2.incl,v 1.23.8.3 2013/08/30 09:44:49 ber Exp $
11 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp" 2

14 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
"d11=30m"
"d12=20u"

```
"l5=d2/(d20+d20+p46)"

"acqt0=-d12"
# 21 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
d9
  d11 setnmr3^3
# 1 "mc_line 23 file /opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp exp.
def. part of mc cmd. before ze"
define delay MCWRK
define delay MCREST
"MCREST = 30m - 30m"
"MCWRK = 0.333333*30m"

   dccorr
# 23 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
1 ze
# 1 "mc_line 23 file /opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp exp.
def. of mc cmd. after ze"
# 24 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
  d11 pl12:f2
# 1 "mc_line 25 file /opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp exp.
start label for mc cmd."
2 MCWRK  * 2 do:f2
LBLF0, MCWRK
  MCREST
# 26 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
  d1
  d12 setnmr3|2
  (p45:sp3 ph1):f1

3 d20              ;start CPMG
  (p46:sp4 ph2):f1
  d20
  lo to 3 times l5    ;end CPMG

  d12 setnmr3^2
  go=2 ph31 cpd2:f2
# 1 "mc_line 37 file /opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp exp.
mc cmd. in line"
  MCWRK do:f2 wr #0
  MCWRK zd
  lo to LBLF0 times td0

  MCWRK
# 38 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
exit


ph1=0 2 2 0 1 3 3 1
ph2=1 3 3 1 2 0 0 2
ph31=0 2 2 0 1 3 3 1
# 46 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
;pl12: f2 channel - power level for CPD/BB decoupling
;sp3 : f1 channel - power level 90 degree shaped excitation pulse (power level should be set to 600W)
;sp4 : f1 channel - power level broadband refocusing CPMG pulse (corresponds to power of regular 180 pulse)
;p45: f1 channel - 90 degree shaped excitation pulse (600 us): HIGH POWER PULSE (check cryo probe limits!)
;p46: f1 channel - 180 broadband refocusing CPMG pulse (1 ms): LONG PULSE (check cryo probe limits!)
```

;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O (30 ms)
;d12: delay for power switching (20 us)
;d2 : CPMG duration (20-400 ms)
;d20 : CPMG delay (10-40 ms)
;l5 : Number of CPMG cycles (1-10)
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

62 "/opt/topspin3.2pl7/exp/stan/nmr/lists/pp/user/bbCPMG_bb19F_dec_phase_opt.awp"
;$Id: zgfhigqn,v 1.7 2002/06/12 09:05:22 ber Exp $

hk_19Fcpmg_screen_adia

1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia"
;19Fcpmg_2d.adia October 2016
;avance-version (16/04/29)
;pseudo 2D sequence
;T2 measurement with CPMG and water suppression using excitation sculpting with gradients
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=

prosol relations=<default>
1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

169 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.9.8.3 2013/08/30 09:44:49 ber Exp $
16 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia" 2

1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl - include file for Gradient Spectroscopy
;   for TCU3
;
;avance-version (07/01/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

27 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl"
define list<gradient> EA=<EA>

31 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl"

```
;$Id: Grad2.incl,v 1.13.8.1 2012/01/31 17:56:17 ber Exp $
# 17 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia" 2


# 1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl  -  include file for commonly used delays
;
;version (13/08/07)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 9 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8
define delay DELTA9
define delay DELTA10
define delay DELTA11
define delay DELTA12
define delay DELTA13
define delay DELTA14
define delay DELTA15
define delay DELTA16

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5
define delay TAU6
define delay TAU7
define delay TAU8
define delay TAU9
# 40 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define delay INCR1
define delay INCR2
define delay INCR3
define delay INCR4
define delay INCR5
define delay INCR6


;delays for centering pulses
# 50 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
```

```
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2


;loop counters
# 69 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3
define loopcounter FACTOR4
define loopcounter FACTOR5
define loopcounter FACTOR6


# 80 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
;$Id: Delay.incl,v 1.13.8.2 2013/08/30 09:44:49 ber Exp $
# 18 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia" 2


;cpmg1d
;avance-version (04/11/23)
 ;1D experiment with
 ;   T2 filter using Carr-Purcell-Meiboom-Gill sequence
 ;
 ;$CLASS=HighRes
 ;$DIM=1D
 ;$TYPE=
 ;$SUBTYPE=
 ;$COMMENT=
# 1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=


# 165 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl"
# 169 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.9.8.3 2013/08/30 09:44:49 ber Exp $
# 31 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia" 2


"p2=p1*2"
```

```
 "d11=30m"
 "d12=12u"
 "d21=d20-p13/2"
 "d23=d20-p13/2-de-1u"
 "acqt0=0"
 ;"d2=l4*(d20*2)"
 ;"l4=d2/(d20*2)"
 ;"d2=l4*(d20*2)"
 ;baseopt_echo
# 1 "mc_line 43 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia exp. def. part of
mc cmd. before ze"
define delay MCWRK
define delay MCREST
"MCREST = 30m - 30m"
"MCWRK = 0.200000*30m"

   dccorr
# 43 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia"
 1 ze
# 1 "mc_line 43 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia exp. def. of mc
cmd. after ze"
# 44 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia"
# 1 "mc_line 44 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia exp. start label
for mc cmd."
2 MCWRK  * 2 do:f2
LBLF0, MCWRK  * 2
LBLF1, MCWRK
 MCREST
# 45 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia"
  d1 pl1:f1 pl12:f2
  20u
  p1 ph1

 3 d21
  ;(p2 ph2):f1          ;hard 180-pulse
  (p13:sp5 ph2):f1                  ; adiabatic 180-pulse
  d21
  lo to 3 times c
  1u
  go=2 ph31 cpd2:f2
# 1 "mc_line 56 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia exp. mc cmd. in
line"
  MCWRK do:f2 wr #0
  MCWRK zd
  lo to LBLF0 times td0
  MCWRK  if #0 MCWRK zd ivc
  lo to LBLF1 times td1
  MCWRK
# 58 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia"
 ;d2
 exit
# 61 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/hk_19Fcpmg_screen_adia"
ph1=0 0 2 2 1 1 3 3
 ph2=1 3 1 3 0 2 0 2
 ph31=0 0 2 2 1 1 3 3
```

;sp5:wvm:cawurst-20(cnst1 ppm, 2ms)
;cnst1: 19F sweep width in ppm
;pl1 : f1 channel - power level for pulse (default)
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O   [30 msec]
;d20: d20<< 1/J ,but >(50 * P2) echo time [1-2 msec]
;fixed echo time to allow elimination of J-mod. effects
;d2: total cpmg time d2=l4*(d20*2+p2)
;l4: loop for T2 filter   [4 - 20]
;NS: 1 * n, total number of scans: NS * TD0
;DS: 16

;$Id: cpmg1d,v 1.4 2005/11/10 12:16:59 ber Exp $

bd_zgigcpmg2d

1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bd_zgigcpmg2d"
;bd_zgigcpmg2d
;based on zgig and cpmg1d
;avance-version (12/01/11)
;1D sequence with inverse gated decoupling
;T2 filter using Carr-Purcell-Meiboom-Gill sequence
;
;$CLASS=HighRes
;$DIM=1D
;$TYPE=
;$SUBTYPE=
;$COMMENT=

1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=
169 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.10.2.2 2013/08/30 09:43:33 ber Exp $
14 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bd_zgigcpmg2d" 2

1 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl - include file for commonly used delays
;
;version (13/08/07)
;
;$CLASS=HighRes Incl
;$COMMENT=

9 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Delay.incl"
;general delays

define delay DELTA

```
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8
define delay DELTA9
define delay DELTA10
define delay DELTA11
define delay DELTA12
define delay DELTA13
define delay DELTA14
define delay DELTA15
define delay DELTA16

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5
define delay TAU6
define delay TAU7
define delay TAU8
define delay TAU9
# 40 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Delay.incl"
define delay INCR1
define delay INCR2
define delay INCR3
define delay INCR4
define delay INCR5
define delay INCR6


;delays for centering pulses
# 50 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Delay.incl"
define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2


;loop counters
# 69 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Delay.incl"
```

```
define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3
define loopcounter FACTOR4
define loopcounter FACTOR5
define loopcounter FACTOR6


# 80 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/Delay.incl"
;$Id: Delay.incl,v 1.14.2.1 2013/08/30 09:43:33 ber Exp $
# 15 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bd_zgigcpmg2d" 2


# 18 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bd_zgigcpmg2d"
"p2=p1*2"
"d11=30m"

"TAU=2*(d20+p1)*l4"
"acqt0=-p1*2/3.1416"

# 1 "mc_line 25 file /opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bd_zgigcpmg2d dc-measurement inserted
automatically"

  dccorr
# 25 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bd_zgigcpmg2d"
# 25 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bd_zgigcpmg2d"
1 ze
 d11 pl12:f2
2 30m do:f2
 d1
 p1 ph1
3 d20
 p2 ph2
 d20
 lo to 3 times c
 go=2 ph31 cpd2:f2
 30m do:f2 wr #0 if #0 ivc
 lo to 1 times td1
exit


ph1=0 2 2 0 1 3 3 1
ph2=1 3 1 3 0 2 0 2
ph31=0 2 2 0 1 3 3 1
# 45 "/opt/topspin3.6.2/exp/stan/nmr/lists/pp/user/bd_zgigcpmg2d"
;pl1 : f1 channel - power level for pulse (default)
;pl12: f2 channel - power level for CPD/BB decoupling
;p1 : f1 channel -  high power pulse
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O                     [30 msec]
;ns: 1 * n, total number of scans: NS * TD0
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence
```

Chsqc.rh

1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
;hsqcetgpsp
;avance-version (12/01/11)
;HSQC
;2D H-1/X correlation via double inept transfer
;phase sensitive using Echo/Antiecho-TPPI gradient selection
;with decoupling during acquisition
;using trim pulses in inept transfer
;using shaped pulses for inversion on f2 - channel
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=

1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=
169 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.9.8.3 2013/08/30 09:44:49 ber Exp $
17 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh" 2

1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl - include file for Gradient Spectroscopy
;   for TCU3
;
;avance-version (07/01/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

27 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl"
define list<gradient> EA=<EA>

31 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl"
;$Id: Grad2.incl,v 1.13.8.1 2012/01/31 17:56:17 ber Exp $
18 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh" 2

1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl - include file for commonly used delays
;
;version (13/08/07)
;

;$CLASS=HighRes Incl
;$COMMENT=

9 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8
define delay DELTA9
define delay DELTA10
define delay DELTA11
define delay DELTA12
define delay DELTA13
define delay DELTA14
define delay DELTA15
define delay DELTA16

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5
define delay TAU6
define delay TAU7
define delay TAU8
define delay TAU9
40 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define delay INCR1
define delay INCR2
define delay INCR3
define delay INCR4
define delay INCR5
define delay INCR6

;delays for centering pulses
50 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3

```
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2


;loop counters
# 69 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3
define loopcounter FACTOR4
define loopcounter FACTOR5
define loopcounter FACTOR6


# 80 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
;$Id: Delay.incl,v 1.13.8.2 2013/08/30 09:44:49 ber Exp $
# 19 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh" 2


# 22 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
"p2=p1*2"
"p4=p3*2"
"d4=1s/(cnst2*4)"
"d11=30m"


"p22=p21*2"
# 34 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
"in0=inf1/2"
"d0=in0*0.5"


"DELTA1=d4-p16-larger(p2,p14)/2-de-8u"
"DELTA2=d4-larger(p2,p14)/2"
"DELTA3=d4-larger(p2,p14)/2-p1*2/PI"
# 43 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
"DELTA=p16+d16+larger(p2,p22)"


# 49 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
"acqt0=0"
baseopt_echo


# 1 "mc_line 53 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh exp. def. part of mc cmd. before
ze"
define delay MCWRK
define delay MCREST
define delay d0orig
"d0orig=d0"
define loopcounter t1loop
"t1loop=0"
define loopcounter ph1loop
"ph1loop=0"
```

```
define loopcounter ST1CNT
"ST1CNT = td1 / 2"
"MCREST = d1 - d1"
"MCWRK = 0.076923*d1"

  dccorr
# 53 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
1 ze
# 1 "mc_line 53 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh exp. def. of mc cmd. after ze"
    "d0=d0orig + t1loop * in0 "
    "phval3 = (t1loop * 180)%360"
    MCWRK ip3 + phval3
    "phval6 = (t1loop * 180)%360"
    MCWRK ip6 + phval6
    "phval31 = (t1loop * 180)%360"
    MCWRK ip31 + phval31
    MCWRK
    "phval0 = ph1loop * 1"
    MCWRK setgrad EA
# 54 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
  d11 pl12:f2
# 1 "mc_line 55 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh exp. start label for mc cmd."
2 MCWRK  * 6 do:f2
LBLF1, MCWRK  * 6
LBLST1, MCWRK
  MCREST
# 56 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
3 (p1 ph1)
  DELTA2 pl0:f2
  4u
  (center (p2 ph1) (p14:sp3 ph6):f2 )
  4u
  DELTA2 pl2:f2 setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
  p28 ph1
  4u
  (p1 ph2) (p3 ph3):f2
  d0
# 68 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
  (center (p2 ph5) (p22 ph1):f3 )


# 73 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
  d0
  p16:gp1*EA
  d16
  (p4 ph4):f2
  DELTA
  (ralign (p1 ph1) (p3 ph4):f2 )
  DELTA3 pl0:f2
  (center (p2 ph1) (p14:sp3 ph1):f2 )
  4u
  p16:gp2
  DELTA1 pl12:f2
  4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
  go=2 ph31 cpd2:f2
# 1 "mc_line 86 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh exp. mc cmd. in line"
  MCWRK  do:f2 wr #0 if #0 zd
```

```
   "ph1loop+=1"
     "d0=d0orig + t1loop * in0 "
     "phval3 = (t1loop * 180)%360"
     MCWRK ip3 + phval3
     "phval6 = (t1loop * 180)%360"
     MCWRK ip6 + phval6
     "phval31 = (t1loop * 180)%360"
     MCWRK ip31 + phval31
     MCWRK
     "phval0 = ph1loop * 1"
     MCWRK setgrad EA
  lo to LBLF1 times 2
  MCWRK

  "t1loop+=1"
     "d0=d0orig + t1loop * in0 "
     "phval3 = (t1loop * 180)%360"
     MCWRK ip3 + phval3
     "phval6 = (t1loop * 180)%360"
     MCWRK ip6 + phval6
     "phval31 = (t1loop * 180)%360"
     MCWRK ip31 + phval31
     MCWRK
     "phval0 = ph1loop * 1"
     MCWRK setgrad EA
  lo to LBLST1 times ST1CNT
  MCWRK
  "t1loop=0"
  "ph1loop=0"
  MCWRK
# 88 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 0 0 2 2 2 2
ph5=0 0 2 2
ph6=0
ph31=0 2 0 2 2 0 2 0

# 100 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
;pl0 : 0W
;pl1 : f1 channel - power level for pulse (default)
;pl2 : f2 channel - power level for pulse (default)
;pl3 : f3 channel - power level for pulse (default)
;pl12: f2 channel - power level for CPD/BB decoupling
;sp3: f2 channel - shaped pulse 180 degree
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel -  90 degree high power pulse
;p4 : f2 channel - 180 degree high power pulse
;p14: f2 channel - 180 degree shaped pulse for inversion
;p16: homospoil/gradient pulse
;p22: f3 channel - 180 degree high power pulse
```

;p28: f1 channel - trim pulse
;d0 : incremented delay (2D)   [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I/O   [30 msec]
;d16: delay for homospoil/gradient recovery
;cnst2: = J(XH)
;inf1: 1/SW(X) = 2 * DW(X)
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;ns: 1 * n
;ds: >= 16
;td1: number of experiments
;FnMODE: echo-antiecho
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

;use gradient ratio: gp 1 : gp 2
;   80 : 20.1   for C-13
;   80 : 8.1   for N-15
135 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
;for z-only gradients:
;gpz1: 80%
;gpz2: 20.1% for C-13, 8.1% for N-15

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100

144 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
                        ;preprocessor-flags-start
;1: for C-13 and N-15 labeled samples start experiment with
;   option -DLABEL_CN (eda: ZGOPTNS)
                        ;preprocessor-flags-end

151 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/Chsqc.rh"
;$Id: hsqcetgpsp,v 1.6.2.1.4.1 2012/01/31 17:56:32 ber Exp $

bd-ledbpgppr2s

1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
;ledbpgppr2s
;avance-version (12/01/11)
;2D sequence for diffusion measurement using stimulated
; echo and LED
;using bipolar gradient pulses for diffusion
;using 2 spoil gradients
;with presaturation during relaxation delay
;
;D. Wu, A. Chen & C.S. Johnson Jr.,
; J. Magn. Reson. A 115, 260-264 (1995).
;
;$CLASS=HighRes
;$DIM=2D

```
;$TYPE=
;$SUBTYPE=
;$COMMENT=


# 1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance3.incl
;   for AV III
;
;avance-version (13/06/17)
;
;$CLASS=HighRes Incl
;$COMMENT=
# 169 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Avance.incl"
;$Id: Avance3.incl,v 1.9.8.3 2013/08/30 09:44:49 ber Exp $
# 19 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s" 2

# 1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl  -  include file for Gradient Spectroscopy
;   for TCU3
;
;avance-version (07/01/17)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 27 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl"
define list<gradient> EA=<EA>


# 31 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Grad.incl"
;$Id: Grad2.incl,v 1.13.8.1 2012/01/31 17:56:17 ber Exp $
# 20 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s" 2

# 1 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl  -  include file for commonly used delays
;
;version (13/08/07)
;
;$CLASS=HighRes Incl
;$COMMENT=

# 9 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8
define delay DELTA9
define delay DELTA10
define delay DELTA11
```

```
define delay DELTA12
define delay DELTA13
define delay DELTA14
define delay DELTA15
define delay DELTA16

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5
define delay TAU6
define delay TAU7
define delay TAU8
define delay TAU9
# 40 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define delay INCR1
define delay INCR2
define delay INCR3
define delay INCR4
define delay INCR5
define delay INCR6


;delays for centering pulses
# 50 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2


;loop counters
# 69 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3
define loopcounter FACTOR4
define loopcounter FACTOR5
define loopcounter FACTOR6


# 80 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/Delay.incl"
```

```
;$Id: Delay.incl,v 1.13.8.2 2013/08/30 09:44:49 ber Exp $
# 21 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s" 2


# 24 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
define list<gradient> diff=<Difframp>


"p2=p1*2"
"d11=30m"
"d12=20u"
# 32 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
"DELTA1=d20-p1*2-p2-p30*2-d16*2-p19-d16-d12-4u"
"DELTA2=d21-p19-d16-d12-8u"


"acqt0=-p1*2/3.1416"
# 1 "mc_line 39 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s exp. def. part of mc cmd.
before ze"
define delay MCWRK
define delay MCREST
define loopcounter t1loop
"t1loop=0"
define loopcounter ph1loop
"ph1loop=0"
define loopcounter ST1CNT
"ST1CNT = td1"
"MCREST = d11 - d11"
"MCWRK = 0.250000*d11"


    dccorr
# 39 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
# 39 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
1 ze
# 1 "mc_line 39 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s exp. def. of mc cmd. after
ze"
    MCWRK
    "phval0 = t1loop * 1"
    MCWRK setgrad diff
# 40 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
# 1 "mc_line 40 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s exp. start label for mc
cmd."
2 MCWRK  * 3
LBLST1, MCWRK
 MCREST
# 41 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
3 d12 pl9:f1
 d1 cw:f1 ph29
 50u do:f1 setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
 d12 pl1:f1
 p1 ph1
 p30:gp6*diff
 d16
 p2 ph1
 p30:gp6*-1*diff
 d16
 p1 ph2
```

```
   p19:gp7
   d16 pl9:f1
   DELTA1 cw:f1 ph29
   4u do:f1
   d12 pl1:f1
   p1 ph3
   p30:gp6*diff
   d16
   p2 ph1
   p30:gp6*-1*diff
   d16
   p1 ph4
   p19:gp8
   d16 pl9:f1
   DELTA2 cw:f1 ph29
   4u do:f1
   d12 pl1:f1
   4u setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
   p1 ph5
   go=2 ph31
# 1 "mc_line 72 file /opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s exp. mc cmd. in line"
   MCWRK  wr #0 if #0 zd

   "t1loop+=1"
      MCWRK
      "phval0 = t1loop * 1"
      MCWRK setgrad diff
   lo to LBLST1 times ST1CNT
   MCWRK
   "t1loop=0"
   MCWRK
# 73 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
exit


ph1= 0
ph2= 0 0 2 2
ph3= 0 0 0 0 2 2 2 2  1 1 1 1 3 3 3 3
ph4= 0 2 0 2 2 0 2 0  1 3 1 3 3 1 3 1
ph5= 0 0 0 0 2 2 2 2  1 1 1 1 3 3 3 3
ph29=0 0 0 0 0 0 0 0  1 1 1 1 1 1 1 1
ph31=0 2 2 0 2 0 0 2  3 1 1 3 1 3 3 1
# 85 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
;pl1 : f1 channel - power level for pulse (default)
;pl9 : f1 channel - power level for presaturation
;p1  : f1 channel -  90 degree high power pulse
;p2  : f1 channel - 180 degree high power pulse
;p19: gradient pulse 2 (spoil gradient)
;p30: gradient pulse (little DELTA * 0.5)
;d1  : relaxation delay; 1-5 * T1
;d11: delay for disk I/O                [30 msec]
;d12: delay for power switching         [20 usec]
;d16: delay for gradient recovery
;d20: diffusion time (big DELTA)
;d21: eddy current delay (Te)           [5 ms]
;ns: 16 * n
;ds: 4 * m
```

;td1: number of experiments
;FnMODE: QF
;   use xf2 and DOSY processing

;use gradient ratio: gp 6 : gp 7 : gp 8
;   100 : -17.13 : -13.17
107 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
;for z-only gradients:
;gpz6: 100%
;gpz7: -17.13% (spoil)
;gpz8: -13.17% (spoil)

;use gradient files:
;gpnam6: SMSQ10.100
;gpnam7: SMSQ10.100
;gpnam8: SMSQ10.100

;use AU-program dosy to calculate gradient ramp-file Difframp
121 "/opt/topspin3.2.5/exp/stan/nmr/lists/pp/user/bd-ledbpgppr2s"
;$Id: ledbpgppr2s,v 1.3.8.1 2012/01/31 17:56:33 ber Exp $

9 List of Abbreviations

AD – Alzheimer's disease

AS – affinity screen

ASO – anti-sense oligonucleotide

BCA - bicinchoninic acid

BLS – binding level screen

BSA – bovine serum albumin

CD – circular dichroism

CMC – critical micelle concentration

CNBr – cyanogen bromide

CPMG - Carr-Purcell-Meiboom-Gill

CS – clean screen

CSA – chemical shift anisotropy

CSP – chemical shift perturbation

DLS – dynamic light scattering

DMSO – dimethyl sulfoxide

DOSY – diffusion ordered spectroscopy

DSS - 4,4-dimethyl-4-silapentane-1-sulfonic acid

DTT- dithiothreitol

EK – enterokinase

ESI-MS – electrospray ionization mass spectroscopy

FBLD – fragment-based ligand discovery

FRET – fluorescence resonance energy transfer

HASPA – hydrophilic acylated surface protein A

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HMQC – heteronuclear multiple quantum coherence

HPLC-MS – high pressure liquid chromatography-mass spectroscopy

HSQC – heteronuclear single quantum coherence

HTS – high throughput screening

IDP – intrinsically disordered protein

IDR – intrinsically disordered region

IMAC – immobilized metal affinity chromatography

ITC – isothermal titration calorimetry

KSI – ketosteroid isomerase

LC-MS – liquid chromatography-mass spectrometry

LE – ligand efficiency

LES – local elementary structure

LMW – low molecular weight

MAPT – microtubule-associated protein tau

MB – methylene blue

MoRF – molecular recognition feature

MS – mass spectroscopy

MST – microscale thermophoresis

MTG – 1-thioglycerol

MW – molecular weight

MWCO – molecular weight cut-off

NGM – NMR-guided modelling

NMR – nuclear magnetic resonance

NN – near neighour

NOESY – nuclear Overhauser effect spectroscopy

NTD – neglected tropical disease

OPA - *ortho*-phthalaldehyde

PAINS – pan-assay interference compounds

PDB – protein data bank

PHF – paired helical filament

pI – isoelectric point

PMF – peptide mass fingerprint

PPI – protein-protein interaction

PTM – post-translational modification

PTP1B – protein-tyrosine phosphatase 1B

QC – quality control

RT – room temperature

RU – response unit

SA – streptavidin

SAR – structure-activity relationship

SBLD – structure-based ligand discovery

SDM – site-directed mutagenesis

SDS-PAGE – sodium dodecyl sulfate - polyacrylamide gel electrophoresis

SEC – size exclusion chromatography

SEC-MALS – size exclusion chromatography coupled with multiple angle light scattering

SHERP – small hydrophilic endoplasmic reticulum-associated protein

SF-HMQC – SOFAST heteronuclear multiple quantum coherence

SPR – surface plasmon resonance

STD – saturation transfer difference

TCEP - tris(2-carboxyethyl)phosphine

TCI – targeted covalent inhibitor

TEV– tobacco etch virus

Thr – thrombin

TOCSY – total correlation spectroscopy

Tris-HCl - tris(hydroxymethyl)aminomethane hydrochloride

WHO – World Health Organization

wLOGSY – water-ligand observed via gradient spectroscopy

10 References

[1]   P. E. Wright and H. J. Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm," *J. Mol. Biol.*, vol. 293, no. 2, pp. 321–331, Oct. 1999.

[2]   A. L. Hopkins and C. R. Groom, "The druggable genome," *Nat. Rev. Drug Discov.*, vol. 1, no. 9, pp. 727–730, Sep. 2002.

[3]   A. E. Mirsky and L. Pauling, "On the Structure of Native, Denatured, and Coagulated Proteins," *Proc. Natl. Acad. Sci.*, vol. 22, no. 7, pp. 439–447, Jul. 1936.

[4]   M. L. Anson, "Protein Denaturation and the Properties of Protein Groups," in *Advances in Protein Chemistry*, vol. 2, no. C, 1945, pp. 361–386.

[5]   L. Bragg, J. C. Kendrew, and M. F. Perutz, "Polypeptide chain configurations in crystalline proteins," *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.*, vol. 203, no. 1074, pp. 321–357, Oct. 1950.

[6]   J. C. Venter *et al.*, "The Sequence of the Human Genome," *Science (80-.).*, vol. 291, no. 5507, pp. 1304–1351, Feb. 2001.

[7]   B. Modrek and C. Lee, "A genomic view of alternative splicing," *Nat. Genet.*, vol. 30, no. 1, pp. 13–19, 2002.

[8]   "The Protein Data Bank," 2020. [Online]. Available: rcsb.org. [Accessed: 20-Jan-2020].

[9]   H. M. Berman, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, Jan. 2000.

[10]   R. A. Pullen, J. A. Jenkins, I. J. Tickle, S. P. Wood, and T. L. Blundell, "The relation of polypeptide hormone structure and flexibility to receptor binding: The relevance of X-ray studies on insulins, glucagon and human placental lactogen," *Mol. Cell. Biochem.*, vol. 8, no. 1, pp. 5–20, Jul. 1975.

[11]   J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones, "Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life," *J. Mol. Biol.*, vol. 337, no. 3, pp. 635–645, 2004.

[12]   J. Yan, A. K. Dunker, V. N. Uversky, and L. Kurgan, "Molecular recognition features (MoRFs) in three domains of life.," *Mol. Biosyst.*, vol. 12, no. 3, pp. 697–710, Mar. 2016.

[13]   Y. Cheng, C. J. Oldfield, J. Meng, P. Romero, V. N. Uversky, and A. K. Dunker, "Mining α-Helix-Forming Molecular Recognition Features with Cross Species Sequence Alignments †," *Biochemistry*, vol. 46, no. 47, pp. 13468–13477, Nov. 2007.

[14]   A. C. Bloomer, J. N. Champness, G. Bricogne, R. Staden, and A. Klug, "Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits," *Nature*, vol. 276, no. 5686, pp. 362–368, Nov. 1978.

[15]   S. W. Muchmore *et al.*, "X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death," *Nature*, vol. 381, no. 6580, pp. 335–341, May 1996.

[16]   J. A. Wells and C. L. McClendon, "Reaching for high-hanging fruit in drug discovery at protein–protein interfaces," *Nature*, vol. 450, no. 7172, pp. 1001–1009, Dec. 2007.

[17]   Y. Zheng *et al.*, "Temporal regulation of EGF signalling networks by the scaffold protein Shc1," *Nature*, vol. 499, no. 7457, pp. 166–171, Jul. 2013.

[18]   A. K. Dunker and V. N. Uversky, "Signal transduction via unstructured protein conduits.," *Nat. Chem. Biol.*, vol. 4, no. 4, pp. 229–230, 2008.

[19] C. Schulenburg and D. Hilvert, "Protein conformational disorder and enzyme catalysis," in *Topics in Current Chemistry*, vol. 337, no. 1, 2013, pp. 41–68.

[20] C. A. Galea, Y. Wang, S. G. Sivakolundu, and R. W. Kriwacki, "Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits," *Biochemistry*, vol. 47, no. 29, pp. 7598–7609, 2008.

[21] C. Querfurth, A. C. R. Diernfellner, E. Gin, E. Malzahn, T. Höfer, and M. Brunner, "Circadian Conformational Change of the Neurospora Clock Protein FREQUENCY Triggered by Clustered Hyperphosphorylation of a Basic Domain," *Mol. Cell*, vol. 43, no. 5, pp. 713–722, Sep. 2011.

[22] P. Nash *et al.*, "Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication," *Nature*, vol. 414, no. 6863, pp. 514–521, Nov. 2001.

[23] M. Borg, T. Mittag, T. Pawson, M. Tyers, J. D. Forman-Kay, and H. S. Chan, "Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity," *Proc. Natl. Acad. Sci.*, vol. 104, no. 23, pp. 9650–9655, Jun. 2007.

[24] A. Bah and J. D. Forman-Kay, "Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications," *J. Biol. Chem.*, vol. 291, no. 13, pp. 6696–6705, Mar. 2016.

[25] P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 1, pp. 18–29, 2014.

[26] A. K. Dunker, C. J. Brown, and Z. Obradovic, "Identification and functions of usefully disordered proteins," *Adv. Protein Chem.*, vol. 62, pp. 25–49, 2002.

[27] H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser, "The ensemble nature of allostery," *Nature*, vol. 508, no. 7496, pp. 331–339, Apr. 2014.

[28] E. Monsellier and F. Chiti, "Prevention of amyloid-like aggregation as a driving force of protein evolution," *EMBO Rep.*, vol. 8, no. 8, pp. 737–742, Aug. 2007.

[29] J. Gsponer, M. E. Futschik, S. A. Teichmann, and M. M. Babu, "Tight regulation of unstructured proteins: from transcript synthesis to protein degradation," *Science (80-.).*, vol. 322, no. 5906, pp. 1365–1368, 2008.

[30] A. K. Dunker *et al.*, "Intrinsically disordered protein," *J. Mol. Graph. Model.*, vol. 19, no. 1, pp. 26–59, 2001.

[31] V. N. Uversky, "Protein folding revisited. A polypeptide chain at the folding - Misfolding - Nonfolding cross-roads: Which way to go?," *Cell. Mol. Life Sci.*, vol. 60, no. 9, pp. 1852–1871, 2003.

[32] V. Iešmantavičius, J. Dogan, P. Jemth, K. Teilum, and M. Kjaergaard, "Helical Propensity in an Intrinsically Disordered Protein Accelerates Ligand Binding," *Angew. Chemie Int. Ed.*, vol. 53, no. 6, pp. 1548–1551, Feb. 2014.

[33] A. V. Follis *et al.*, "PUMA binding induces partial unfolding within BCL-xL to disrupt p53 binding and promote apoptosis," *Nat. Chem. Biol.*, vol. 9, no. 3, pp. 163–168, Mar. 2013.

[34] C. J. Oldfield and A. K. Dunker, "Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions," *Annu. Rev. Biochem.*, vol. 83, no. 1, pp. 553–584, Jun. 2014.

[35] B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky, and A. K. Dunker, "Predicting intrinsic disorder in proteins: an overview," *Cell Res.*, vol. 19, no. 8, pp. 929–949, Aug. 2009.

[36] C. J. Brown, A. K. Johnson, and G. W. Daughdrill, "Comparing Models of Evolution for Ordered

and Disordered Proteins," *Mol. Biol. Evol.*, vol. 27, no. 3, pp. 609–621, 2010.

[37]   P. Sarkar, C. Reichman, T. Saleh, R. B. Birge, and C. G. Kalodimos, "Proline cis-trans Isomerization Controls Autoinhibition of a Signaling Protein," *Mol. Cell*, vol. 25, no. 3, pp. 413–426, 2007.

[38]   M. H. Glickman and A. Ciechanover, "The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction," *Physiol. Rev.*, vol. 82, no. 2, pp. 373–428, Apr. 2002.

[39]   A. Steward, S. Adhya, and J. Clarke, "Sequence Conservation in Ig-like Domains: The Role of Highly Conserved Proline Residues in the Fibronectin Type III Superfamily," *J. Mol. Biol.*, vol. 318, no. 4, pp. 935–940, 2002.

[40]   V. N. Uversky, "Natively unfolded proteins: A point where biology waits for physics," *Protein Sci.*, vol. 11, no. 4, pp. 739–756, Apr. 2002.

[41]   M. R. Arkin and J. A. Wells, "Small-molecule inhibitors of protein–protein interactions: progressing towards the dream," *Nat. Rev. Drug Discov.*, vol. 3, no. 4, pp. 301–317, Apr. 2004.

[42]   W. L. DeLano, "Convergent Solutions to Binding at a Protein-Protein Interface," *Science (80-.).*, vol. 287, no. 5456, pp. 1279–1283, Feb. 2000.

[43]   M. Fuxreiter, P. Tompa, and I. Simon, "Local structural disorder imparts plasticity on linear motifs," *Bioinformatics*, vol. 23, no. 8, pp. 950–956, 2007.

[44]   H. Dinkel *et al.*, "The eukaryotic linear motif resource ELM: 10 years and counting," *Nucleic Acids Res.*, vol. 42, no. D1, pp. D259–D266, Jan. 2014.

[45]   A. R. Dinner, A. Šali, L. J. Smith, C. M. Dobson, and M. Karplus, "Understanding protein folding via free-energy surfaces from theory and experiment," *Trends Biochem. Sci.*, vol. 25, no. 7, pp. 331–339, Jul. 2000.

[46]   C. M. Dobson, "Protein folding and misfolding," *Nature*, vol. 426, no. 6968, pp. 884–890, Dec. 2003.

[47]   F. Khan, J. I. Chuang, S. Gianni, and A. R. Fersht, "The kinetic pathway of folding of barnase," *J. Mol. Biol.*, vol. 333, no. 1, pp. 169–186, 2003.

[48]   V. Daggett and A. Fersht, "Opinion: The present view of the mechanism of protein folding," *Nat. Rev. Mol. Cell Biol.*, vol. 4, no. 6, pp. 497–502, 2003.

[49]   C. Levinthal, "How to fold graciously," *Mössbauer Spectrosc. Biol. Syst. Proc.*, vol. 24, no. 41, pp. 22–24, 1969.

[50]   R. Broglia, Y. Levy, and G. Tiana, "HIV-1 protease folding and the design of drugs which do not create resistance," *Curr. Opin. Struct. Biol.*, vol. 18, no. 1, pp. 60–66, Feb. 2008.

[51]   M. R. Pincus, "Identification of structured peptide segments in folding proteins," *Biopolymers*, vol. 32, no. 4, pp. 347–351, 1992.

[52]   S. K. Wandinger, K. Richter, and J. Buchner, "The Hsp90 chaperone machinery," *J. Biol. Chem.*, vol. 283, no. 27, pp. 18473–18477, 2008.

[53]   F. U. Hartl, "Molecular chaperones in cellular protein folding," *Nature*, vol. 381, no. 6583, pp. 571–580, Jun. 1996.

[54]   S. Jones and J. M. Thornton, "Principles of protein protein interactions," *Proc. Natl. Acad. Sci.*, vol. 93, no. 1, pp. 13–20, 1996.

[55]   A. C. Cheng *et al.*, "Structure-based maximal affinity model predicts small-molecule druggability," *Nat Biotech*, vol. 25, no. 1, pp. 71–75, 2007.

[56]   R. D. Smith, L. Hu, J. A. Falkner, M. L. Benson, J. P. Nerothin, and H. A. Carlson, "Exploring protein-ligand recognition with Binding MOAD," *J. Mol. Graph. Model.*, vol. 24, no. 6, pp. 414–425, 2006.

[57]   J. Chen, N. Sawyer, and L. Regan, "Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area," *Protein Sci.*, vol. 22, no. 4, pp. 510–515, Apr. 2013.

[58]   F. B. Sheinerman, R. Norel, and B. Honig, "Electrostatic aspects of protein-protein interactions," *Curr. Opin. Struct. Biol.*, vol. 10, no. 2, pp. 153–159, 2000.

[59]   A. M. Davis and S. J. Teague, "Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis," *Angew. Chemie - Int. Ed.*, vol. 38, no. 6, pp. 736–749, 1999.

[60]   A. Borgia *et al.*, "Extreme disorder in an ultrahigh-affinity protein complex," *Nature*, vol. 555, no. 7694, pp. 61–66, Mar. 2018.

[61]   A. C. Joerger and A. R. Fersht, "Structure–function–rescue: the diverse nature of common p53 cancer mutants," *Oncogene*, vol. 26, no. 15, pp. 2226–2242, Apr. 2007.

[62]   L. T. Vassilev *et al.*, "In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2.," *Science (80-.).*, vol. 303, no. 5659, pp. 844–848, 2004.

[63]   S. Shangary *et al.*, "Temporal activation of p53 by a specific MDM2 inhibitor is selectively toxic to tumors and leads to complete tumor growth inhibition," *Proc. Natl. Acad. Sci.*, vol. 105, no. 10, pp. 3933–3938, Mar. 2008.

[64]   J. A. Toretsky *et al.*, "Oncoprotein EWS-FLI1 Activity Is Enhanced by RNA Helicase A," *Cancer Res.*, vol. 66, no. 11, pp. 5574–5581, Jun. 2006.

[65]   H. V Erkizan *et al.*, "A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma," *Nat Med*, vol. 15, no. 7, pp. 750–756, 2009.

[66]   T. Berg *et al.*, "Small-molecule antagonists of Myc/Max dimerization inhibit Myc-induced transformation of chicken embryo fibroblasts.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, pp. 3830–3835, 2002.

[67]   H. Wang *et al.*, "Improved low molecular weight Myc-Max inhibitors," *Mol. Cancer Ther.*, vol. 6, no. 9, pp. 2399–2408, Sep. 2007.

[68]   A. V. Follis, D. I. Hammoudeh, A. T. Daab, and S. J. Metallo, "Small-molecule perturbation of competing interactions between c-Myc and Max," *Bioorganic Med. Chem. Lett.*, vol. 19, no. 3, pp. 807–810, 2009.

[69]   A. V. Follis, D. I. Hammoudeh, H. Wang, E. V. Prochownik, and S. J. Metallo, "Structural Rationale for the Coupled Binding and Unfolding of the c-Myc Oncoprotein by Small Molecules," *Chem. Biol.*, vol. 15, no. 11, pp. 1149–1155, Nov. 2008.

[70]   S. Couty, I. Westwood, A. Kalusa, and C. Cano, "The discovery of potent ribosomal S6 kinase inhibitors by high- throughput screening and structure-guided drug design," *Oncotarget*, vol. 4, no. 10, pp. 1647–1661, 2013.

[71]   B. Shen, "A New Golden Age of Natural Products Drug Discovery," *Cell*, vol. 163, no. 6, pp. 1297–1300, 2015.

[72] J. W. Blunt, B. R. Copp, R. A. Keyzers, M. H. G. Munro, and M. R. Prinsep, "Marine natural products," *Nat. Prod. Rep.*, vol. 32, no. 2, pp. 116–211, 2015.

[73] B. J. Walters *et al.*, "High-Throughput Screening Reveals Alsterpaullone, 2-Cyanoethyl as a Potent p27Kip1 Transcriptional Inhibitor," *PLoS One*, vol. 9, no. 3, p. e91173, Mar. 2014.

[74] M. M. Hann and T. I. Oprea, "Pursuing the leadlikeness concept in pharmaceutical research," *Curr. Opin. Chem. Biol.*, vol. 8, no. 3, pp. 255–263, Jun. 2004.

[75] P. J. Hajduk *et al.*, "Discovery of Potent Nonpeptide Inhibitors of Stromelysin Using SAR by NMR," *J. Am. Chem. Soc.*, vol. 119, no. 25, pp. 5818–5827, Jun. 1997.

[76] S. B. Shuker, P. J. Hajduk, R. P. Meadows, and S. W. Fesik, "Discovering High-Affinity Ligands for Proteins: SAR by NMR," *Science (80-.).*, vol. 274, no. 5292, pp. 1531–1534, Nov. 1996.

[77] M. J. Harner, A. O. Frank, and S. W. Fesik, "Fragment-based drug discovery using NMR spectroscopy," *J. Biomol. NMR*, vol. 56, no. 2, pp. 65–75, 2013.

[78] M. J. Hartshorn, C. W. Murray, A. Cleasby, M. Frederickson, I. J. Tickle, and H. Jhoti, "Fragment-based lead discovery using X-ray crystallography," *J. Med. Chem.*, vol. 48, no. 2, pp. 403–413, 2005.

[79] M. M. Hann, A. R. Leach, and G. Harper, "Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery," *J. Chem. Inf. Comput. Sci.*, vol. 41, no. 3, pp. 856–864, May 2001.

[80] N. Baurin *et al.*, "Design and Characterization of Libraries of Molecular Fragments for Use in NMR Screening against Protein Targets," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 6, pp. 2157–2166, Nov. 2004.

[81] H. Jhoti, G. Williams, D. C. Rees, and C. W. Murray, "The 'rule of three' for fragment-based drug discovery: where are we now?," *Nat. Rev. Drug Discov.*, vol. 12, no. 8, pp. 644–644, 2013.

[82] M. Congreve, R. Carr, C. Murray, and H. Jhoti, "A 'Rule of Three' for fragment-based lead discovery?," *Drug Discov. Today*, vol. 8, no. 19, pp. 876–877, Oct. 2003.

[83] B. Lamoree and R. E. Hubbard, "Current perspectives in fragment-based lead discovery (FBLD)," *Essays Biochem.*, vol. 61, no. 5, pp. 453–464, Nov. 2017.

[84] R. E. Hubbard and J. B. Murray, "Experiences in Fragment-Based Lead Discovery," in *Methods in Enzymology*, 1st ed., vol. 493, Elsevier Inc., 2011, pp. 509–531.

[85] M. Mayer and B. Meyer, "Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy," *Angew. Chemie Int. Ed.*, vol. 38, no. 12, pp. 1784–1788, Jun. 1999.

[86] C. Dalvit, P. Pevarello, M. Tatò, M. Veronesi, A. Vulpetti, and M. Sundström, "Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water.," *J. Biomol. NMR*, vol. 18, no. 1, pp. 65–68, Sep. 2000.

[87] P. J. Hajduk, E. T. Olejniczak, and S. W. Fesik, "One-Dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules," *J. Am. Chem. Soc.*, vol. 119, no. 50, pp. 12257–12261, Dec. 1997.

[88] A. Vulpetti and C. Dalvit, "Design and Generation of Highly Diverse Fluorinated Fragment Libraries and their Efficient Screening with Improved 19 F NMR Methodology," *ChemMedChem*, vol. 8, no. 12, pp. 2057–2069, Dec. 2013.

[89] M. Sattler and S. W. Fesik, "Use of deuterium labeling in NMR: overcoming a sizeable problem,"

Structure, vol. 4, no. 11, pp. 1245–1249, Nov. 1996.

[90]   B. Krishnarjuna, G. Jaipuria, A. Thakur, P. D'Silva, and H. S. Atreya, "Amino acid selective unlabeling for sequence specific resonance assignments in proteins," *J. Biomol. NMR*, vol. 49, no. 1, pp. 39–51, Jan. 2011.

[91]   D. A. Erlanson, B. J. Davis, and W. Jahnke, "Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures," *Cell Chem. Biol.*, vol. 26, no. 1, pp. 9–15, Jan. 2019.

[92]   J. B. Murray *et al.*, "Establishing Drug Discovery and Identification of Hit Series for the Anti-apoptotic Proteins, Bcl-2 and Mcl-1," *ACS Omega*, vol. 4, no. 5, pp. 8892–8906, May 2019.

[93]   GE Healthcare, "Biacore Assay Handbook," 2012. [Online]. Available: https://www.gelifesciences.co.jp/contact/pdf/BiacoreAssayHandbook.pdf. [Accessed: 25-Feb-2020].

[94]   I. Navratilova and A. L. Hopkins, "Fragment Screening by Surface Plasmon Resonance," *ACS Med. Chem. Lett.*, vol. 1, no. 1, pp. 44–48, Apr. 2010.

[95]   J. B. Murray, S. D. Roughley, N. Matassova, and P. A. Brough, "Off-Rate Screening (ORS) By Surface Plasmon Resonance. An Efficient Method to Kinetically Sample Hit to Lead Chemical Space from Unpurified Reaction Products," *J. Med. Chem.*, vol. 57, no. 7, pp. 2845–2850, Apr. 2014.

[96]   R. A. Copeland, D. L. Pompliano, and T. D. Meek, "Drug-target residence time and its implications for lead optimization," *Nat. Rev. Drug Discov.*, vol. 5, no. 9, pp. 730–739, 2006.

[97]   M. Jerabek-Willemsen *et al.*, "MicroScale Thermophoresis: Interaction analysis and beyond," *J. Mol. Struct.*, vol. 1077, pp. 101–113, Dec. 2014.

[98]   P. Linke *et al.*, "An Automated Microscale Thermophoresis Screening Approach for Fragment-Based Lead Discovery," *J. Biomol. Screen.*, vol. 21, no. 4, pp. 414–421, Apr. 2016.

[99]   S. D. Bembenek, B. A. Tounge, and C. H. Reynolds, "Ligand efficiency and fragment-based drug discovery," *Drug Discov. Today*, vol. 14, no. 5–6, pp. 278–283, Mar. 2009.

[100]   M. R. Arkin, Y. Tang, and J. A. Wells, "Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality," *Chem. Biol.*, vol. 21, no. 9, pp. 1102–1114, 2014.

[101]   C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug Deliv. Rev.*, vol. 64, no. SUPPL., pp. 4–17, Dec. 2012.

[102]   F. Giordanetto and J. Kihlberg, "Macrocyclic drugs and clinical candidates: What can medicinal chemists learn from their properties?," *J. Med. Chem.*, vol. 57, no. 2, pp. 278–295, 2014.

[103]   C. A. Lipinski, "Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions," *Adv. Drug Deliv. Rev.*, vol. 101, pp. 34–41, 2016.

[104]   T. Mohamed, T. Hoang, M. Jelokhani-Niaraki, and P. P. N. Rao, "Tau-Derived-Hexapeptide 306 VQIVYK 311 Aggregation Inhibitors: Nitrocatechol Moiety as A Pharmacophore In Drug Design," *ACS Chem. Neurosci.*, vol. 4, no. 12, pp. 1559–1570, Dec. 2013.

[105]   N. Krishnan *et al.*, "Targeting the disordered C terminus of PTP1B with an allosteric inhibitor," *Nat. Chem. Biol.*, vol. 10, no. 7, pp. 558–566, 2014.

[106] D.-P. Hong, A. L. Fink, and V. N. Uversky, "Structural Characteristics of α-Synuclein Oligomers Stabilized by the Flavonoid Baicalein," *J. Mol. Biol.*, vol. 383, no. 1, pp. 214–223, Oct. 2008.

[107] P. Bernado, D. I. Svergun, P. Bernadó, and D. I. Svergun, "Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering," *Mol Biosyst*, vol. 8, no. 1, pp. 151–167, 2012.

[108] "Neglected tropical diseases," 2018. [Online]. Available: http://www.who.int/neglected_diseases/diseases/en/. [Accessed: 16-Jan-2020].

[109] IOM (Institute of Medicine), *The Causes and Impacts of Neglected Tropical and Zoonotic Diseases*. Washington, D.C.: National Academies Press, 2011.

[110] B. L. Herwaldt, "Leishmaniasis," *Lancet*, vol. 354, no. 9185, pp. 1191–1199, Oct. 1999.

[111] P. D. Ready, "Biology of Phlebotomine Sand Flies as Vectors of Disease Agents," *Annu. Rev. Entomol.*, vol. 58, no. 1, pp. 227–250, Jan. 2013.

[112] V. S. Amato, F. F. Tuon, H. A. Bacha, V. A. Neto, and A. C. Nicodemo, "Mucosal leishmaniasis. Current scenario and prospects for treatment," *Acta Trop.*, vol. 105, no. 1, pp. 1–9, 2008.

[113] T. S. Tiuman, A. O. Santos, T. Ueda-Nakamura, B. P. D. Filho, and C. V. Nakamura, "Recent advances in leishmaniasis treatment," *Int. J. Infect. Dis.*, vol. 15, no. 8, pp. e525–e532, Aug. 2011.

[114] R. Killick-Kendrick, "The biology and control of Phlebotomine sand flies," *Clin. Dermatol.*, vol. 17, no. 3, pp. 279–289, May 1999.

[115] H. M. Flinn and D. F. Smith, "Genomic organisation and expression of a differentially-regulated gene family from Leishmania major," *Nucleic Acids Res.*, vol. 20, no. 4, pp. 755–762, 1992.

[116] P. G. McKean, R. Delahay, P. F. P. Pimenta, and D. F. Smith, "Characterisation of a second protein encoded by the differentially regulated LmcDNA16 gene family of Leishmania major," *Mol. Biochem. Parasitol.*, vol. 85, no. 2, pp. 221–231, Apr. 1997.

[117] R. M. R. Coulson and D. F. Smith, "Isolation of genes showing increased or unique expression in the infective promastigotes of Leishmania major," *Mol. Biochem. Parasitol.*, vol. 40, no. 1, pp. 63–75, Apr. 1990.

[118] P. W. Denny, S. Gokool, D. G. Russell, M. C. Field, and D. F. Smith, "Acylation-dependent Protein Export in Leishmania," *J. Biol. Chem.*, vol. 275, no. 15, pp. 11017–11025, Apr. 2000.

[119] S. McKenna, "Functional and structural studies of the hydrophilic acylated surface proteins from Leishmania donovani," University of York, 2015.

[120] J. S. P. Doehl *et al.*, "Leishmania HASP and SHERP Genes Are Required for In Vivo Differentiation, Parasite Transmission and Virulence Attenuation in the Host," *PLOS Pathog.*, vol. 13, no. 1, p. e1006130, Jan. 2017.

[121] L. M. MacLean *et al.*, "Trafficking and release of Leishmania metacyclic HASPB on macrophage invasion," *Cell. Microbiol.*, vol. 14, no. 5, pp. 740–761, May 2012.

[122] P. G. McKean, P. W. Denny, E. Knuepfer, J. K. Keen, and D. F. Smith, "Phenotypic changes associated with deletion and overexpression of a stage-regulated gene family in Leishmania," *Cell. Microbiol.*, vol. 3, no. 8, pp. 511–523, 2001.

[123] R. J. Spears *et al.*, "Site-selective C–C modification of proteins at neutral pH using organocatalyst-mediated cross aldol ligations," *Chem. Sci.*, vol. 9, no. 25, pp. 5585–5593, 2018.

[124] Y. Shi *et al.*, "Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding," *Protein Sci.*, vol. 21, no. 8, pp. 1197–1209, Aug. 2012.

[125] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, and T. Gray, "How to measure and predict the molar absorption coefficient of a protein," *Protein Sci.*, vol. 4, no. 11, pp. 2411–2423, Nov. 1995.

[126] D. Zhu, A. Saul, S. Huang, L. B. Martin, L. H. Miller, and K. M. Rausch, "Use of o-phthalaldehyde assay to determine protein contents of Alhydrogel-based vaccines," *Vaccine*, vol. 27, no. 43, pp. 6054–6059, Oct. 2009.

[127] A. E. Derome, *Modern NMR techniques for chemistry research*, vol. 6. Elsevier, 2013.

[128] D. V. Waterhous and W. C. Johnson, "Importance of Environment in Determining Secondary Structure in Proteins," *Biochemistry*, vol. 33, no. 8, pp. 2121–2128, 1994.

[129] C. Krittanai and W. C. Johnson, "The relative order of helical propensity of amino acids changes with solvent environment," *Proteins Struct. Funct. Genet.*, vol. 39, no. 2, pp. 132–141, 2000.

[130] R. Orlando, S.-C. Jao, M. G. Zagorski, K. Ma, and J. Talafous, "Trifluoroacetic acid pretreatment reproducibly disaggregates the amyloid β-peptide," *Amyloid*, vol. 4, no. 4, pp. 240–252, 1997.

[131] S. D. Wolff and R. S. Balaban, "NMR imaging of labile proton exchange," *J. Magn. Reson.*, vol. 86, no. 1, pp. 164–169, Jan. 1990.

[132] S. R. LaPlante *et al.*, "Compound Aggregation in Drug Discovery: Implementing a Practical NMR Assay for Medicinal Chemists," *J. Med. Chem.*, vol. 56, no. 12, pp. 5142–5150, Jun. 2013.

[133] B. Fuglestad, N. E. Kerstetter, S. Bédard, and A. J. Wand, "Extending the Detection Limit in Fragment Screening of Proteins Using Reverse Micelle Encapsulation," *ACS Chem. Biol.*, vol. 14, no. 10, pp. 2224–2232, Oct. 2019.

[134] World Health Organization, "Cancer." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer. [Accessed: 05-Jan-2020].

[135] M. A. Pierotti, M. Frattini, F. Molinari, G. Sozzi, and C. M. Croce, "Oncogenes," in *Holland-Frei Cancer Medicine*, John Wiley & Sons, Inc., 2016, pp. 1–22.

[136] E. M. Blackwood and R. N. Eisenman, "Max: A Helix-Loop-Helix Zipper Protein That Forms a Sequence-Specific DNA-Binding Complex with Myc," *Source Sci. New Ser.*, vol. 251, no. 4998, pp. 1211–1217, 1991.

[137] C. V Dang *et al.*, "Function of the c-Myc oncogenic transcription factor.," *Exp. Cell Res.*, vol. 253, no. 1, pp. 63–77, 1999.

[138] H. Wang *et al.*, "c-Myc depletion inhibits proliferation of human tumor cells at various stages of the cell cycle," *Oncogene*, vol. 27, no. 13, pp. 1905–1915, Mar. 2008.

[139] C. E. Nesbit, J. M. Tersak, and E. V Prochownik, "MYC oncogenes and human neoplastic disease," *Oncogene*, vol. 18, no. 19, pp. 3004–3016, May 1999.

[140] T. Oltersdorf *et al.*, "An inhibitor of Bcl-2 family proteins induces regression of solid tumours," *Nature*, vol. 435, no. 7042, pp. 677–681, Jun. 2005.

[141] M. J. Smith, D. C. Charron-Prochownik, and E. V Prochownik, "The leucine zipper of c-Myc is required for full inhibition of erythroleukemia differentiation.," *Mol. Cell. Biol.*, vol. 10, no. 10, pp. 5333–5339, Oct. 1990.

[142] X. Yin, C. Giap, J. S. Lazo, and E. V Prochownik, "Low molecular weight inhibitors of Myc-Max

interaction and function.," *Oncogene*, vol. 22, no. 40, pp. 6151–6159, 2003.

[143] I. Müller *et al.*, "Targeting of the MYCN Protein with Small Molecule c-MYC Inhibitors," *PLoS One*, vol. 9, no. 5, pp. 1–12, May 2014.

[144] J. L. Yap *et al.*, "Pharmacophore identification of c-Myc inhibitor 10074-G5," *Bioorg. Med. Chem. Lett.*, vol. 23, no. 1, pp. 370–374, Jan. 2013.

[145] K.-Y. Jung *et al.*, "Perturbation of the c-Myc–Max Protein–Protein Interaction via Synthetic α-Helix Mimetics," *J. Med. Chem.*, vol. 58, no. 7, pp. 3002–3024, Apr. 2015.

[146] J. Wanner *et al.*, "Reversible linkage of two distinct small molecule inhibitors of myc generates a dimeric inhibitor with improved potency that is active in myc over-expressing cancer cell lines," *PLoS One*, vol. 10, no. 4, pp. 1–17, 2015.

[147] J. R. Hart *et al.*, "Inhibitor of MYC identified in a Krohnke pyridine library," *Proc. Natl. Acad. Sci.*, pp. 2–7, 2014.

[148] A. Castell *et al.*, "A selective high affinity MYC-binding compound inhibits MYC:MAX interaction and MYC-dependent tumor cell proliferation," *Sci. Rep.*, vol. 8, no. 1, pp. 1–17, Dec. 2018.

[149] H. Wang *et al.*, "Direct inhibition of c-Myc-Max heterodimers by celastrol and celastrol-inspired triterpenoids," *Oncotarget*, vol. 6, no. 32, pp. 32380–32395, Oct. 2015.

[150] I. Coin, M. Beyermann, and M. Bienert, "Solid-phase peptide synthesis: From standard procedures to the synthesis of difficult sequences," *Nature Protocols*, vol. 2, no. 12. pp. 3247–3256, 2007.

[151] M. Milburn *et al.*, "Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins," *Science (80-.).*, vol. 247, no. 4945, pp. 939–945, Feb. 1990.

[152] S. H. Shahravan, X. Qu, I.-S. Chan, and J. A. Shin, "Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag," *Protein Expr. Purif.*, vol. 59, no. 2, pp. 314–319, Jun. 2008.

[153] I. Čipáková, J. Gašperík, and E. Hostinová, "Expression and purification of human antimicrobial peptide, dermcidin, in Escherichia coli," *Protein Expr. Purif.*, vol. 45, no. 2, pp. 269–274, 2006.

[154] Y. A. Andreev, S. A. Kozlov, A. A. Vassilevski, and E. V. Grishin, "Cyanogen bromide cleavage of proteins in salt and buffer solutions," *Anal. Biochem.*, vol. 407, no. 1, pp. 144–146, 2010.

[155] V. S. Dobrev, L. M. Fred, K. P. Gerhart, and S. J. Metallo, "Characterization of the Binding of Small Molecules to Intrinsically Disordered Proteins," in *Methods in Enzymology*, 1st ed., vol. 611, Elsevier Inc., 2018, pp. 677–702.

[156] Y. An, J. W. G. Bloom, and S. E. Wheeler, "Quantifying the π-Stacking Interactions in Nitroarene Binding Sites of Proteins," *J. Phys. Chem. B*, vol. 119, no. 45, pp. 14441–14450, Nov. 2015.

[157] M. H. Forouzanfar *et al.*, "Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015," *Lancet*, vol. 388, no. 10053, pp. 1659–1724, Oct. 2016.

[158] J. Ferlay *et al.*, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. E359–E386, Mar. 2015.

[159] WHO, "Obesity and overweight," 2017. [Online]. Available: https://www.who.int/en/news-

room/fact-sheets/detail/obesity-and-overweight. [Accessed: 02-Jan-2020].

[160] R. R. A. Bourne *et al.*, "Causes of vision loss worldwide, 1990–2010: a systematic analysis," *Lancet Glob. Heal.*, vol. 1, no. 6, pp. e339–e349, Dec. 2013.

[161] N. Sarwar *et al.*, "Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102 prospective studies," *Lancet*, vol. 375, no. 9733, pp. 2215–2222, 2010.

[162] A. Chaudhury *et al.*, "Clinical Review of Antidiabetic Drugs: Implications for Type 2 Diabetes Mellitus Management," *Front. Endocrinol. (Lausanne).*, vol. 8, no. JAN, pp. 1–12, Jan. 2017.

[163] J. Schlessinger, "Cell Signaling by Receptor Tyrosine Kinases," *October*, vol. 103, no. 2, pp. 211–225, 2000.

[164] N. K. Tonks, "Protein tyrosine phosphatases: from genes, to function, to disease," *Nat Rev Mol Cell Biol*, vol. 7, no. 11, pp. 833–846, 2006.

[165] J. J. Hornberg *et al.*, "Principles behind the multifarious control of signal transduction," *FEBS J.*, vol. 272, no. 1, pp. 244–258, Dec. 2004.

[166] M. Bentires-Alj *et al.*, "Activating Mutations of the Noonan Syndrome-Associated SHP2/PTPN11 Gene in Human Solid Tumors and Adult Acute Myelogenous Leukemia," *Cancer Res.*, vol. 64, no. 24, pp. 8816–8820, Dec. 2004.

[167] M. Elchebly *et al.*, "Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene.," *Science*, vol. 283, no. 5407, pp. 1544–1548, 1999.

[168] S. G. Julien *et al.*, "Protein tyrosine phosphatase 1B deficiency or inhibition delays ErbB2-induced mammary tumorigenesis and protects from lung metastasis.," *Nat. Genet.*, vol. 39, no. 3, pp. 338–346, 2007.

[169] N. K. Tonks, "Protein tyrosine phosphatases - From housekeeping enzymes to master regulators of signal transduction," *FEBS J.*, vol. 280, no. 2, pp. 346–378, 2013.

[170] A. K. Tamrakar, C. K. Maurya, and A. K. Rai, "PTP1B inhibitors for type 2 diabetes treatment: a patent review (2011 – 2014)," *Expert Opin. Ther. Pat.*, vol. 24, no. 10, pp. 1101–1115, Oct. 2014.

[171] S. ZHANG and Z. ZHANG, "PTP1B as a drug target: recent developments in PTP1B inhibitor discovery," *Drug Discov. Today*, vol. 12, no. 9–10, pp. 373–381, May 2007.

[172] R. He, L.-F. Zeng, Y. He, and Z.-Y. Zhang*, "Chapter 6. Recent Advances in PTP1B Inhibitor Development for the Treatment of Type 2 Diabetes and Obesity," in *Drugs of the Future*, vol. 29, no. 12, 2012, pp. 142–176.

[173] S. Shrestha, B. R. Bhattarai, H. Cho, J.-K. Choi, and H. Cho, "PTP1B inhibitor Ertiprotafib is also a potent inhibitor of IκB kinase β (IKK-β)," *Bioorg. Med. Chem. Lett.*, vol. 17, no. 10, pp. 2728–2730, May 2007.

[174] T. Jin, H. Yu, and X.-F. Huang, "Selective binding modes and allosteric inhibitory effects of lupane triterpenes on protein tyrosine phosphatase 1B," *Sci. Rep.*, vol. 6, no. 1, pp. 1–14, Feb. 2016.

[175] C. Wiesmann *et al.*, "Allosteric inhibition of protein tyrosine phosphatase 1B.," *Nat Struct Mol Biol*, vol. 11, no. 8, pp. 730–737, 2004.

[176] M. S. Choy *et al.*, "Conformational Rigidity and Protein Dynamics at Distinct Timescales Regulate PTP1B Activity and Allostery.," *Mol. Cell*, vol. 65, no. 4, pp. 644-658.e5, Feb. 2017.

[177] Isis Pharmaceuticals, "Isis Pharmaceuticals Reports Data From Phase 2 Study of ISIS-PTP1B Rx in Patients With Type 2 Diabetes." [Online]. Available: https://ir.ionispharma.com/news-releases/news-release-details/isis-pharmaceuticals-reports-data-phase-2-study-isis-ptp1b-rx. [Accessed: 19-Oct-2019].

[178] L. Hao, T. Tiganis, N. K. Tonks, and H. Charbonneau, "The noncatalytic C-terminal segment of the T cell protein tyrosine phosphatase regulates activity via an intramolecular mechanism," *J. Biol. Chem.*, vol. 272, no. 46, pp. 29322–29329, 1997.

[179] J. V. Frangioni, A. Oda, M. Smith, E. W. Salzman, and B. G. Neel, "Calpain-catalyzed cleavage and subcellular relocation of protein phosphotyrosine phosphatase 1B (PTP-1B) in human platelets.," *EMBO J.*, vol. 12, no. 12, pp. 4843–4856, Dec. 1993.

[180] J. V. Frangioni, P. H. Beahm, V. Shifrin, C. A. Jost, and B. G. Neel, "The nontransmembrane tyrosine phosphatase PTP-1B localizes to the endoplasmic reticulum via its 35 amino acid C-terminal sequence," *Cell*, vol. 68, no. 3, pp. 545–560, Feb. 1992.

[181] M. N. Rao *et al.*, "Aminosterols from the dogfish shark Squalus acanthias," *J. Nat. Prod.*, vol. 63, no. 5, pp. 631–635, 2000.

[182] L. Lu and M. Zhu, "Protein Tyrosine Phosphatase Inhibition by Metals and Metal Complexes," *Antioxid. Redox Signal.*, vol. 20, no. 14, pp. 2210–2224, May 2014.

[183] N. Krishnan, K. F. Konidaris, G. Gasser, and N. K. Tonks, "A potent, selective, and orally bioavailable inhibitor of the protein-tyrosine phosphatase PTP1B improves insulin and leptin signaling in animal models.," *J. Biol. Chem.*, vol. 293, no. 5, pp. 1517–1525, Feb. 2018.

[184] Z. Qin *et al.*, "Functional properties of Claramine: A novel PTP1B inhibitor and insulin-mimetic compound," *Biochem. Biophys. Res. Commun.*, vol. 458, no. 1, pp. 21–27, 2015.

[185] M. J. Suskiewicz, J. L. Sussman, I. Silman, and Y. Shaul, "Context-dependent resistance to proteolysis of intrinsically disordered proteins," *Protein Sci.*, vol. 20, no. 8, pp. 1285–1297, 2011.

[186] J. Gowrishankar and R. Harinarayanan, "Why is transcription coupled to translation in bacteria?," *Mol. Microbiol.*, vol. 54, no. 3, pp. 598–603, Sep. 2004.

[187] D. Esposito and D. K. Chatterjee, "Enhancement of soluble protein expression through the use of fusion tags," *Curr. Opin. Biotechnol.*, vol. 17, no. 4, pp. 353–358, 2006.

[188] H. Tegel, J. Ottosson, and S. Hober, "Enhancing the protein production levels in Escherichia coli with a strong promoter.," *FEBS J.*, vol. 278, no. 5, pp. 729–39, Mar. 2011.

[189] F. W. Studier, "Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system," *J. Mol. Biol.*, vol. 219, no. 1, pp. 37–44, 1991.

[190] M. Perni *et al.*, "Multistep Inhibition of α-Synuclein Aggregation and Toxicity in Vitro and in Vivo by Trodusquemine," *ACS Chem. Biol.*, vol. 13, no. 8, pp. 2308–2319, 2018.

[191] F. Leroux, P. Jeschke, and M. Schlosser, "α-Fluorinated Ethers, Thioethers, and Amines: Anomerically Biased Species," *Chem. Rev.*, vol. 105, no. 3, pp. 827–856, Mar. 2005.

[192] M. Morgenthaler *et al.*, "Predicting and Tuning Physicochemical Properties in Lead Optimization: Amine Basicities," *ChemMedChem*, vol. 2, no. 8, pp. 1100–1115, Aug. 2007.

[193] K. Muller, C. Faeh, and F. Diederich, "Fluorine in Pharmaceuticals: Looking Beyond Intuition," *Science (80-.).*, vol. 317, no. 5846, pp. 1881–1886, Sep. 2007.

[194] A. Vulpetti, N. Schiering, and C. Dalvit, "Combined use of computational chemistry, NMR screening, and X-ray crystallography for identification and characterization of fluorophilic protein environments," *Proteins Struct. Funct. Bioinforma.*, vol. 78, no. 16, pp. 3281–3291, 2010.

[195] C. Dalvit, "Ligand- and substrate-based 19F NMR screening: Principles and applications to drug discovery," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 51, no. 4, pp. 243–271, Nov. 2007.

[196] H. S. Andersen *et al.*, "Discovery and SAR of a Novel Selective and Orally Bioavailable Nonpeptide Classical Competitive Inhibitor Class of Protein-Tyrosine Phosphatase 1B," *J. Med. Chem.*, vol. 45, no. 20, pp. 4443–4459, Sep. 2002.

[197] J. Singh, R. C. Petter, T. A. Baillie, and A. Whitty, "The resurgence of covalent drugs," *Nat. Rev. Drug Discov.*, vol. 10, no. 4, pp. 307–317, 2011.

[198] Y.-L. Zhang, Y.-F. Keng, Y. Zhao, L. Wu, and Z.-Y. Zhang, "Suramin Is an Active Site-directed, Reversible, and Tight-binding Inhibitor of Protein-tyrosine Phosphatases," *J. Biol. Chem.*, vol. 273, no. 20, pp. 12281–12287, May 1998.

[199] A. K. Larsen, "Suramin: an anticancer drug with unique biological effects," *Cancer Chemother. Pharmacol.*, vol. 32, no. 2, pp. 96–98, 1993.

[200] D. F. McCain *et al.*, "Suramin Derivatives as Inhibitors and Activators of Protein-tyrosine Phosphatases," *J. Biol. Chem.*, vol. 279, no. 15, pp. 14713–14725, Apr. 2004.

[201] P. Stenlund, Å. Frostell-Karlsson, and O. P. Karlsson, "Studies of small molecule interactions with protein phosphatases using biosensor technology," *Anal. Biochem.*, vol. 353, no. 2, pp. 217–225, 2006.

[202] S. S. Gustafsson, L. Vrang, Y. Terelius, and U. H. Danielson, "Quantification of interactions between drug leads and serum proteins by use of 'binding efficiency,'" *Anal. Biochem.*, vol. 409, no. 2, pp. 163–175, Feb. 2011.

[203] R. Karlsson, O. Karlsson, and P. Belcher, "Chapter 7. Fragment and Low Molecular Weight Compound Analysis," in *Handbook of Surface Plasmon Resonance*, Cambridge: Royal Society of Chemistry, 2017, pp. 255–294.

[204] A. Schuffenhauer, S. Ruedisser, A. Marzinzik, W. Jahnke, P. Selzer, and E. Jacoby, "Library Design for Fragment Based Screening," *Curr. Top. Med. Chem.*, vol. 5, no. 8, pp. 751–762, 2005.

[205] A. M. Giannetti, B. D. Koch, and M. F. Browner, "Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors," *J. Med. Chem.*, vol. 51, no. 3, pp. 574–580, 2008.

[206] M. G. Spillantini, T. D. Bird, and B. Ghetti, "Frontotemporal dementia and Parkinsonism linked to chromosome 17: a new group of tauopathies.," *Brain Pathol.*, vol. 8, no. 2, pp. 387–402, Apr. 1998.

[207] A. Delacourte *et al.*, "Specific Pathological Tau Protein Variants Characterize Pick's Disease," *J. Neuropathol. Exp. Neurol.*, vol. 55, no. 2, pp. 159–168, Feb. 1996.

[208] M. Baker *et al.*, "Association of an Extended Haplotype in the Tau Gene with Progressive Supranuclear Palsy," *Hum. Mol. Genet.*, vol. 8, no. 4, pp. 711–715, Apr. 1999.

[209] E. M. Mandelkow and E. Mandelkow, "Tau in Alzheimer's disease.," *Trends Cell Biol.*, vol. 8, no. 11, pp. 425–427, Nov. 1998.

[210] J. Q. Trojanowski, M. L. Schmidt, R.-W. Shin, G. T. Bramblett, D. Rao, and V. M. -. Lee, "Altered

Tau and Neurofilament Proteins in Neuro-Degenerative Diseases: Diagnostic Implications for Alzheimer's Disease and Lewy Body Dementias," *Brain Pathol.*, vol. 3, no. 1, pp. 45–54, Jan. 1993.

[211] L. Reiniger *et al.*, "Tau, prions and Aβ: the triad of neurodegeneration," *Acta Neuropathol.*, vol. 121, no. 1, pp. 5–20, Jan. 2011.

[212] WHO, "Dementia," 2017. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia. [Accessed: 24-Oct-2019].

[213] I. Grundke-Iqbal, K. Iqbal, M. Quinlan, Y. C. Tung, M. S. Zaidi, and H. M. Wisniewski, "Microtubule-associated protein tau. A component of Alzheimer paired helical filaments.," *J. Biol. Chem.*, vol. 261, no. 13, pp. 6084–6089, May 1986.

[214] L. Buée, T. Bussière, V. Buée-Scherrer, A. Delacourte, and P. R. Hof, "Tau protein isoforms, phosphorylation and role in neurodegenerative disorders.," *Brain Res. Brain Res. Rev.*, vol. 33, no. 1, pp. 95–130, Aug. 2000.

[215] V. M.-Y. Lee, M. Goedert, and J. Q. Trojanowski, "Neurodegenerative Tauopathies," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1121–1159, Mar. 2001.

[216] M. Goedert, M. G. Spillantini, R. Jakes, D. Rutherford, and R. A. Crowther, "Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease," *Neuron*, vol. 3, no. 4, pp. 519–526, Oct. 1989.

[217] C. Liu, X. Song, R. Nisbet, and J. Götz, "Co-immunoprecipitation with Tau Isoform-specific Antibodies Reveals Distinct Protein Interactions and Highlights a Putative Role for 2N Tau in Disease," *J. Biol. Chem.*, vol. 291, no. 15, pp. 8173–8188, Apr. 2016.

[218] B. L. Goode *et al.*, "Functional interactions between the proline-rich and repeat regions of tau enhance microtubule binding and assembly.," *Mol. Biol. Cell*, vol. 8, no. 2, pp. 353–365, Feb. 1997.

[219] H. Kadavath *et al.*, "Tau stabilizes microtubules by binding at the interface between tubulin heterodimers," *Proc. Natl. Acad. Sci.*, vol. 112, no. 24, pp. 7501–7506, Jun. 2015.

[220] A. Abraha *et al.*, "C-terminal inhibition of tau assembly in vitro and in Alzheimer's disease.," *J. Cell Sci.*, vol. 113 Pt 21, no. 21, pp. 3737–3745, Nov. 2000.

[221] M. von Bergen *et al.*, "Mutations of Tau Protein in Frontotemporal Dementia Promote Aggregation of Paired Helical Filaments by Enhancing Local β-Structure," *J. Biol. Chem.*, vol. 276, no. 51, pp. 48165–48174, Dec. 2001.

[222] M. von Bergen, P. Friedhoff, J. Biernat, J. Heberle, E.-M. Mandelkow, and E. Mandelkow, "Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif (306VQIVYK311) forming beta structure," *Proc. Natl. Acad. Sci.*, vol. 97, no. 10, pp. 5129–5134, May 2000.

[223] O. Schweers, E. M. Mandelkow, J. Biernat, and E.-M. Mandelkow, "Oxidation of cysteine-322 in the repeat domain of microtubule-associated protein tau controls the in vitro assembly of paired helical filaments.," *Proc. Natl. Acad. Sci.*, vol. 92, no. 18, pp. 8463–8467, Aug. 1995.

[224] S. Jeganathan, M. von Bergen, E.-M. Mandelkow, and E. Mandelkow, "The Natively Unfolded Character of Tau and Its Aggregation to Alzheimer-like Paired Helical Filaments †," *Biochemistry*, vol. 47, no. 40, pp. 10526–10539, Oct. 2008.

[225] W. Li and V. M.-Y. Lee, "Characterization of two VQIXXK motifs for tau fibrillization in vitro.," *Biochemistry*, vol. 45, no. 51, pp. 15692–15701, Dec. 2006.

[226] V. M.-Y. Lee, K. R. Brunden, M. Hutton, and J. Q. Trojanowski, "Developing Therapeutic Approaches to Tau, Selected Kinases, and Related Neuronal Protein Targets," *Cold Spring Harb. Perspect. Med.*, vol. 1, no. 1, pp. 1–20, Sep. 2011.

[227] B. Bulic, M. Pickhardt, and E. Mandelkow, "Progress and Developments in Tau Aggregation Inhibitors for Alzheimer Disease," *J. Med. Chem.*, vol. 56, no. 11, pp. 4135–4155, Jun. 2013.

[228] B. Bulic, M. Pickhardt, E. M. Mandelkow, and E. Mandelkow, "Tau protein and tau aggregation inhibitors," *Neuropharmacology*, vol. 59, no. 4–5, pp. 276–289, 2010.

[229] J. Wagner *et al.*, "Reducing tau aggregates with anle138b delays disease progression in a mouse model of tauopathies," *Acta Neuropathol.*, vol. 130, no. 5, pp. 619–631, Nov. 2015.

[230] J. Lefèvre *et al.*, "The C terminus of tubulin, a versatile partner for cationic molecules: Binding of tau, polyamines, and calcium," *J. Biol. Chem.*, vol. 286, no. 4, pp. 3065–3078, 2011.

[231] M. Pickhardt *et al.*, "Identification of Small Molecule Inhibitors of Tau Aggregation by Targeting Monomeric Tau As a Potential Therapeutic Approach for Tauopathies," *Curr. Alzheimer Res.*, vol. 12, no. 9, pp. 814–828, Oct. 2015.

[232] J. Wagner *et al.*, "Anle138b: a novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease.," *Acta Neuropathol.*, vol. 125, no. 6, pp. 795–813, Jun. 2013.

[233] M. Von Bergen, S. Barghorn, J. Biernat, E. M. Mandelkow, and E. Mandelkow, "Tau aggregation is driven by a transition from random coil to beta sheet structure," *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1739, no. 2, pp. 158–166, 2005.

[234] A. Crowe *et al.*, "Aminothienopyridazines and Methylene Blue Affect Tau Fibrillization via Cysteine Oxidation," *J. Biol. Chem.*, vol. 288, no. 16, pp. 11024–11037, Apr. 2013.

[235] M. Pickhardt *et al.*, "Anthraquinones Inhibit Tau Aggregation and Dissolve Alzheimer's Paired Helical Filaments in Vitro and in Cells," *J. Biol. Chem.*, vol. 280, no. 5, pp. 3628–3635, Feb. 2005.

[236] M. Pickhardt *et al.*, "Phenylthiazolyl-Hydrazide and Its Derivatives Are Potent Inhibitors of τ Aggregation and Toxicity in Vitro and in Cells †," *Biochemistry*, vol. 46, no. 35, pp. 10016–10023, Sep. 2007.

[237] C. Ballatore *et al.*, "Aminothienopyridazine inhibitors of tau assembly," WO/2011/037985, 2011.

[238] M. Frenkel-Pinter *et al.*, "Cl-NQTrp Alleviates Tauopathy Symptoms in a Model Organism through the Inhibition of Tau Aggregation-Engendered Toxicity," *Neurodegener. Dis.*, vol. 17, no. 2–3, pp. 73–82, 2017.

[239] C. Ballatore *et al.*, "Aminothienopyridazine inhibitors of tau aggregation: Evaluation of structure–activity relationship leads to selection of candidates with desirable in vivo properties," *Bioorg. Med. Chem.*, vol. 20, no. 14, pp. 4451–4461, Jul. 2012.

[240] E. Akoury, M. Pickhardt, M. Gajda, J. Biernat, E. Mandelkow, and M. Zweckstetter, "Mechanistic Basis of Phenothiazine-Driven Inhibition of Tau Aggregation," *Angew. Chemie Int. Ed.*, vol. 52, no. 12, pp. 3511–3515, Mar. 2013.

[241] E. Fisher *et al.*, "Detection and Characterization of Small Molecule Interactions with Fibrillar Protein Aggregates Using Microscale Thermophoresis," *ACS Chem. Neurosci.*, vol. 8, no. 9, pp. 2088–2095, 2017.

[242] R. Kiss *et al.*, "Structural Basis of Small Molecule Targetability of Monomeric Tau Protein," *ACS

Chem. Neurosci., vol. 9, no. 12, pp. 2997–3006, Dec. 2018.

[243] K. Terpe, "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems," *Appl. Microbiol. Biotechnol.*, vol. 72, no. 2, pp. 211–222, Sep. 2006.

[244] F. W. Studier and B. A. Moffatt, "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes," *J. Mol. Biol.*, vol. 189, no. 1, pp. 113–130, May 1986.

[245] F. Baneyx, "Recombinant protein expression in Escherichia coli," *Curr. Opin. Biotechnol.*, vol. 10, no. 5, pp. 411–421, Oct. 1999.

[246] C. Li, E. A. Lutz, K. M. Slade, R. A. S. Ruf, G.-F. Wang, and G. J. Pielak, "19 F NMR Studies of α-Synuclein Conformation and Fibrillation," *Biochemistry*, vol. 48, no. 36, pp. 8578–8584, Sep. 2009.

[247] N. Michael Green, "[5] Avidin and streptavidin," in *Methods in enzymology*, vol. 184, 1990, pp. 51–67.

[248] J. Turková, "Oriented immobilization of biologically active proteins as a tool for revealing protein interactions and function," *J. Chromatogr. B Biomed. Sci. Appl.*, vol. 722, no. 1–2, pp. 11–31, Feb. 1999.

[249] E. A. Bayer and M. Wilchek, "Application of avidin—biotin technology to affinity-based separations," *J. Chromatogr. A*, vol. 510, pp. 3–11, Jun. 1990.

[250] M. Fairhead and M. Howarth, "Site-Specific Biotinylation of Purified Proteins Using BirA," in *Methods in Molecular Biology*, no. 4, 2015, pp. 171–184.

[251] P. Schanda, Ē. Kupče, and B. Brutscher, "SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds," *J. Biomol. NMR*, vol. 33, no. 4, pp. 199–211, Dec. 2005.

[252] P. Barré and D. Eliezer, "Structural transitions in tau k18 on micelle binding suggest a hierarchy in the efficacy of individual microtubule-binding repeats in filament nucleation," *Protein Sci.*, vol. 22, no. 8, pp. 1037–1048, Aug. 2013.

[253] R. Scherzer-Attali, R. Shaltiel-Karyo, Y. Adalist, D. Segal, and E. Gazit, "Generic inhibition of amyloidogenic proteins by two naphthoquinone-tryptophan hybrid molecules," *Proteins Struct. Funct. Bioinforma.*, vol. 80, no. 8, pp. 1962–1973, Aug. 2012.

[254] M. D. Mukrasch, J. Biernat, M. von Bergen, C. Griesinger, E. Mandelkow, and M. Zweckstetter, "Sites of Tau Important for Aggregation Populate β-Structure and Bind to Microtubules and Polyanions," *J. Biol. Chem.*, vol. 280, no. 26, pp. 24978–24986, Jul. 2005.

[255] C. Dalvit, G. P. Fogliatto, A. Stewart, M. Veronesi, and B. Stockman, "WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability," *J. Biomol. NMR*, vol. 21, no. 4, pp. 349–359, 2001.

[256] Y. Ayotte *et al.*, "Exposing Small-Molecule Nanoentities by a Nuclear Magnetic Resonance Relaxation Assay," *J. Med. Chem.*, vol. 62, no. 17, pp. 7885–7896, Sep. 2019.

[257] S. Paynter and D. A. Russell, "Surface plasmon resonance measurement of pH-induced responses of immobilized biomolecules: conformational change or electrostatic interaction effects?," *Anal. Biochem.*, vol. 309, no. 1, pp. 85–95, Oct. 2002.

[258] J. E. Gestwicki, H. V. Hsieh, and J. B. Pitner, "Using Receptor Conformational Change To Detect Low Molecular Weight Analytes by Surface Plasmon Resonance," *Anal. Chem.*, vol. 73, no. 23,

pp. 5732–5737, Dec. 2001.

[259] S. Doose, H. Neuweiler, and M. Sauer, "A Close Look at Fluorescence Quenching of Organic Dyes by Tryptophan," *ChemPhysChem*, vol. 6, no. 11, pp. 2277–2285, Nov. 2005.

[260] C. Valente, R. Moreira, R. C. Guedes, J. Iley, M. Jaffar, and K. T. Douglas, "The 1,4-naphthoquinone scaffold in the design of cysteine protease inhibitors," *Bioorg. Med. Chem.*, vol. 15, no. 15, pp. 5340–5350, Aug. 2007.

[261] M. Zaldini Hernandes, S. T. Melo Cavalcanti, D. M. Rodrigo Moreira, W. Filgueira de Azevedo Junior, and A. Cristina Lima Leite, "Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design," *Curr. Drug Targets*, vol. 11, pp. 303–314, 2010.

[262] J. E. Kohn *et al.*, "Random-coil behavior and the dimensions of chemically unfolded proteins," *Proc. Natl. Acad. Sci.*, vol. 101, no. 34, pp. 12491–12496, Aug. 2004.

[263] J. T. Gerig, "Fluorine NMR of proteins," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 26, no. PART 4, pp. 293–370, Jan. 1994.

[264] V. Římal, H. Štěpánková, and J. Štěpánek, "Analysis of NMR spectra in case of temperature-dependent chemical exchange between two unequally populated sites," *Concepts Magn. Reson. Part A*, vol. 38A, no. 3, pp. 117–127, May 2011.

[265] P. M. Seidler *et al.*, "Structure-based inhibitors of tau aggregation.," *Nat. Chem.*, vol. 10, no. 2, pp. 170–176, Feb. 2018.

[266] J. Yao, X. Gao, W. Sun, T. Yao, S. Shi, and L. Ji, "Molecular Hairpin: A Possible Model for Inhibition of Tau Aggregation by Tannic Acid," *Biochemistry*, vol. 52, no. 11, pp. 1893–1902, Mar. 2013.

[267] N. G. Sharaf and A. M. Gronenborn, "19F-Modified Proteins and 19F-Containing Ligands as Tools in Solution NMR Studies of Protein Interactions," in *Isotope labeling of biomolecules: labeling techniques*, 1st ed., Elsevier Inc., 2015, pp. 67–95.

[268] R. Evans, G. Dal Poggetto, M. Nilsson, and G. A. Morris, "Improving the Interpretation of Small Molecule Diffusion Coefficients," *Anal. Chem.*, vol. 90, no. 6, pp. 3987–3994, 2018.