

Centre Armand-Frappier Santé Biotechnologie

## **IMPLEMENTING METHODS AND PROCESSES TO SUPPORT FRAGMENT-BASED DRUG DISCOVERY INITIATIVES**

Par  
Yann Ayotte

Thèse présentée pour l'obtention du grade de  
Philosophiae Doctor (Ph.D.)  
en Innovation Pharmaceutique

### **Jury d'évaluation**

Président du jury et  
examineur interne

Charles Calmettes  
INRS – Centre Armand-Frappier Santé  
Biotechnologie

Examineur externe

Steve Bourgault  
Département de chimie  
Université du Québec à Montréal

Examineur externe

Pierre Lavigne  
Faculté de médecine et des sciences de  
la santé  
Université de Sherbrooke

Directeur de recherche

Steven Laplante  
INRS – Centre Armand-Frappier Santé  
Biotechnologie



## RÉSUMÉ

Au cours des dernières décennies, il y a eu un intérêt croissant pour les approches de découverte de pistes basées sur des fragments (FBLD, *fragment-based lead discovery*). En raison de leur faible poids moléculaire et de leur complexité, les fragments peuvent sonder une plus grande partie de l'espace chimique, ce qui nécessite donc des chimiothèques beaucoup plus petites, ce qui rend le FBLD beaucoup plus accessible aux petites institutions. Les approches par fragments sont également de plus en plus utilisées contre des cibles difficiles où les poches de liaison moins traitables ne peuvent pas toujours être sondées avec les méthodes classiques.

Cependant, la détection de fragments possédant de faibles affinités de liaison nécessite fréquemment des concentrations élevées de ligand, ce qui peut augmenter les risques d'artefacts dus à l'auto-agrégation. Nous avons donc validé l'utilisation d'une expérience T2-CPMG de résonance magnétique nucléaire (RMN) comme outil de détection d'agrégation. La méthode est versatile et se prête à un haut débit permettant la surveillance du comportement des composés en parallèle au criblage.

Un autre défi dans la découverte de médicaments est de hiérarchiser l'affinité des composés testés afin de guider l'optimisation chimique de ces composés. La RMN est fréquemment utilisée pour l'évaluation de la liaison, car elle est sensible aux événements de liaison de faibles affinités. Cependant, les méthodes RMN standard pour la détermination de constantes d'affinités nécessitent l'utilisation de plusieurs points de titrage, ce qui rend ces approches moins pratiques en contexte d'optimisation chimique. Pour relever ce défi, nous avons développé une plateforme « RMN pour la RSA » (relation structure-activité) qui repose en grande partie sur des méthodes de RMN basés sur la détection de ligand qui contournent le besoin d'utiliser des concentrations élevées de ligand pour détecter de faibles liaisons intermoléculaires. Des scores de liaison sont extraits des données acquises utilisant des points de concentration uniques (plutôt que des titrages), ce qui permet un classement en fonction des affinités de liaison relatives et ainsi guider les efforts de chimie médicinale. Le comportement des composés en solution, de même que la protéine sont aussi soigneusement surveillés en parallèle. Nous démontrons notre plate-forme RMN pour la RSA avec la cible protéique oncogène HRas comme une preuve de concept dans laquelle des ligands de qualité d'affinités de l'ordre du micromolaire / nanomolaire ont été développés à partir d'un fragment d'environ 10 millimolaire.

La détection de ligands par RMN du fluor ( $^{19}\text{F}$ ), en particulier des ligands de faibles affinités, peut nécessiter des interprétations de changements mineurs dans les déplacements chimiques. Ainsi, le référencement du déplacement chimique est essentiel pour l'interprétations de tels changements. Malheureusement, aucune ligne directrice n'existe quant au choix des références chimiques internes en conditions aqueuses. Dix composés candidats potentiels pouvant servir de références chimiques RMN  $^{19}\text{F}$  ont donc été évalués. Plusieurs paramètres ont été systématiquement évalués afin de déterminer la pertinence de ces candidats pour des fins de criblage par RMN  $^{19}\text{F}$ . Il a été conclu qu'il n'y avait pas de candidat idéal parmi les molécules testées, mais cinq composés pourraient tout de même convenir au criblage par RMN.

On a longtemps supposé que les faibles puissances des fragments les rendraient incompatibles avec le criblage cellulaire. Néanmoins, l'utilisation de chimiothèques de fragments pourraient présenter plusieurs avantages en contexte de criblage phénotypique. Nous avons donc appliqué une approche phénotypique de découverte de pistes basée sur des fragments (FPLD, *fragment-based phenotypic lead discovery*) contre divers modèles de maladies tels que des parasites, des bactéries et des virus afin de déterminer si cela aboutit à une identification fructueuse de composés touchés. Des tendances RSA préliminaires ont été observées, soutenant l'idée que les effets ne sont pas simplement dus à la promiscuité et que le FPLD représente une approche intéressante pour lancer des projets de découverte de médicaments, en particulier pour les maladies plus complexes qui ne possèdent pas de cibles moléculaires bien validées.

Mots-clés : Découverte de médicaments, auto-agrégation de composés, criblage par fragment, résonance magnétique nucléaire, plateforme biophysique, Ras, référence de fluor, criblage phénotypique, maladies infectieuses

## ABSTRACT

Fragment-based lead discovery (FBLD) has gained in popularity over recent decades as a drug discovery approach. Due to their lower molecular weight and complexity, fragments can probe greater portion of chemical space and therefore necessitate significantly smaller libraries, making it much more accessible to smaller institutions. Fragments approaches are also increasingly used against difficult targets where less tractable binding pockets cannot always be probed with classical methods.

However, detection of weaker affinity fragment binders frequently requires high ligand concentrations which increases risks of artifacts due to compound self-aggregation. We therefore validated the use of an nuclear magnetic resonance (NMR) T2-CPMG experiment to detect compound aggregation. The method is versatile and amenable to high-throughput, allowing monitoring of solution behavior in parallel to screening.

Another challenge in drug discovery is how to rank compound affinities to guide chemical optimization of these compounds. NMR is frequently used for fragment screening, as it is sensitive to weak affinity binding events. However, standard NMR methods for binding affinity determination require the use several titration points which makes such approaches less practical for chemical optimization. To address this challenge, we have developed an “NMR for SAR” (structure-activity relationship) platform that is centered on ligand-detected NMR methods which circumvent the need to use high ligand concentrations to detect weak binding. Binding scores are extracted from single concentration points, allowing rank ordering of compounds according to their relative binding affinities in order to guide medicinal chemistry efforts. Compound free-state behavior, as well as the protein are surveyed in parallel. We demonstrate our NMR for SAR platform with the oncogenic HRas protein as a proof of concept in which quality micromolar/nanomolar binders were developed from an initial ~10 millimolar fragment screening hit.

The detection of ligand binding by fluorine ( $^{19}\text{F}$ ) NMR, especially those with weak affinities, can require interpretations of minor changes in chemical shifts. Thus, chemical shift referencing is critical for accurate measurements and interpretations. Unfortunately, no guidelines exist as to the choice of internal chemical references in aqueous solutions. Ten potential candidate compounds that could serve as  $^{19}\text{F}$  NMR chemical references were therefore evaluated. Multiple parameters were systematically evaluated to monitor suitability of these candidates for  $^{19}\text{F}$  NMR

screening purposes. It was concluded that there was no ideal candidate, but five compounds were identified that could be suitable for NMR screening.

It has long been assumed that the low potencies of fragments would make them incompatible with cell-based screening. Nonetheless, the use of fragment libraries in phenotypic screening context could hold several advantages. We therefore applied a fragment-based phenotypic lead discovery (FPLD) approach against various disease models such as parasites, bacteria and viruses in order to determine if this resulted in successful identification of hit compounds. Early SAR trends were observed, supporting the idea that the effects are not simply due to promiscuity and that FPLD represents an interesting approach, especially for more complex diseases that often lack well-validated molecular targets.

Keywords : Drug discovery, compound self-aggregation, fragment screening, nuclear magnetic resonance, Ras, fluorine reference, phenotypic screening, infectious diseases

# TABLE OF CONTENTS

<b>RÉSUMÉ .....</b>	<b>III</b>
<b>ABSTRACT .....</b>	<b>V</b>
<b>TABLE OF CONTENTS .....</b>	<b>VII</b>
<b>LIST OF FIGURES .....</b>	<b>IX</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>XI</b>
<b>1 INTRODUCTION.....</b>	<b>1</b>
1.1 MONITORING AGGREGATION IN FRAGMENT-BASED LEAD DISCOVERY.....	9
1.1.1 <i>Implementing a T2-CPMG NMR Experiment as a Method to Detect Aggregation</i> .....	11
1.1.2 <i>Integrating Multiple Assays Into a Practical Protocol to Monitor Nano-entities</i> .....	15
1.2 VALIDATING AN NMR-CENTRIC PLATFORM TO GENERATE NEW CHEMICAL MATTER.....	16
1.2.1 <i>Protein-observed NMR</i> .....	16
1.2.2 <i>Ligand-observed NMR</i> .....	18
1.2.3 <i>Fragment screening</i> .....	22
1.2.4 <i>Fluorine NMR</i> .....	23
1.2.5 <i>HRas</i> .....	25
1.3 CHALLENGES ASSOCIATED WITH FLUORINE NMR REFERENCING.....	28
1.4 EXPANDING FRAGMENT-BASED LEAD DISCOVERY TO PHENOTYPIC SCREENING APPROACHES .....	29
1.4.1 <i>Parasitic disease systems</i> .....	35
1.4.2 <i>Bacterial disease systems</i> .....	36
1.4.3 <i>Viral disease systems</i> .....	36
1.5 SUMMARY OF CHALLENGES, HYPOTHESES, AND RESEARCH OBJECTIVES.....	37
<b>2 ARTICLE 1 – VALIDATION OF AN NMR RELAXATION ASSAY TO MONITOR COMPOUND AGGREGATION.....</b>	<b>41</b>
<b>3 SUPPORTING INFORMATION - ARTICLE 1 .....</b>	<b>70</b>
<b>4 ARTICLE 2 – INTEGRATING MULTIPLE ASSAYS INTO A PRACTICAL PROTOCOL TO MONITOR NANO-ENTITIES .....</b>	<b>89</b>
<b>5 SUPPORTING INFORMATION - ARTICLE 2 .....</b>	<b>134</b>
<b>6 ARTICLE 3 – IMPLEMENTATION AND VALIDATION OF AN NMR-CENTRIC PLATFORM TO GUIDE EARLY LEAD OPTIMIZATION.....</b>	<b>149</b>
<b>7 SUPPORTING INFORMATION – ARTICLE 3 .....</b>	<b>179</b>
<b>8 ARTICLE 4 - PRACTICAL CONSIDERATIONS AND GUIDELINES FOR SPECTRAL REFERENCING FOR FLUORINE NMR LIGAND SCREENING .....</b>	<b>189</b>
<b>9 SUPPORTING INFORMATION – ARTICLE 4 .....</b>	<b>212</b>

<b>10</b>	<b>ARTICLE 5 – PROOF-OF-CONCEPT STUDY OF A FRAGMENT-BASED PHENOTYPIC SCREENING APPROACH ON LEISHMANIA PARASITES .....</b>	<b>221</b>
<b>11</b>	<b>SUPPORTING INFORMATION - ARTICLE 5 .....</b>	<b>243</b>
<b>12</b>	<b>ARTICLE 6 – EXPANSION OF THE FRAGMENT-BASED PHENOTYPIC APPROACH TO OTHER DISEASE SYSTEMS .....</b>	<b>251</b>
<b>13</b>	<b>SUPPORTING INFORMATION - ARTICLE 6 .....</b>	<b>265</b>
<b>14</b>	<b>GENERAL DISCUSSION AND CONCLUSION .....</b>	<b>279</b>
14.1	COMPOUND AGGREGATION .....	279
14.2	“NMR FOR SAR” PLATFORM .....	281
14.3	FLUORINE NMR REFERENCING .....	283
14.4	FRAGMENT-BASED PHENOTYPIC SCREENING .....	283
<b>15</b>	<b>REFERENCES.....</b>	<b>287</b>



## LIST OF FIGURES

FIGURE 1.1	RELATIONSHIP BETWEEN THE SIZE OF A LIGAND AND ITS BINDING PROBABILITY. ....	3
FIGURE 1.2	INTERACTIONS OF AN HTS HIT COMPARED TO A FRAGMENT HIT.....	4
FIGURE 1.3	RELATIONSHIP BETWEEN COMPOUND MOLECULAR WEIGHT AND POTENCY. ....	5
FIGURE 1.4	FRAGMENT LINKING STRATEGY.....	6
FIGURE 1.5	FRAGMENT MERGING STRATEGY.....	7
FIGURE 1.6	FRAGMENT GROWTH STRATEGY. ....	8
FIGURE 1.7	COMPOUNDS AGGREGATES CAN ASSEMBLE INTO ENTITIES OF VARIOUS SIZES.....	10
FIGURE 1.8	NMR DILUTION ASSAY. ....	12
FIGURE 1.9	INFLUENCE OF T2 ON NET MAGNETIZATION. ....	13
FIGURE 1.10	RELATIONSHIP BETWEEN T2 AND CORRELATION TIME. ....	14
FIGURE 1.11	T2 TO MONITOR MOLECULAR SIZE AND CHEMICAL EXCHANGE. ....	15
FIGURE 1.12	PROTEIN CHEMICAL SHIFT PERTURBATIONS.....	17
FIGURE 1.13	CHEMICAL SHIFT PERTURBATIONS PROVIDE INSIGHT AS TO THE BINDING STOICHIOMETRY.....	17
FIGURE 1.14	VARIOUS NMR PARAMETERS CAN REPORT ON LIGAND BINDING TO A PROTEIN. ....	19
FIGURE 1.15	DEPICTION OF A STEREOTYPICAL BINDER MONITORED BY TRADITIONAL 1D NMR EXPERIMENTS. .....	20
FIGURE 1.16	RELATIONSHIP OF PROTEIN-BOUND LIGAND AND LIGAND CONCENTRATION. ....	21
FIGURE 1.17	SCREENING A POOL OF FRAGMENT BY <sup>19</sup> F NMR.....	24
FIGURE 1.18	REGULATION OF RAS NUCLEOTIDE CYCLE.....	26
FIGURE 1.19	DISTRIBUTION OF RAS ISOFORMS ACROSS TYPES OF TUMORS. ....	27
FIGURE 1.21	NEW MOLECULAR ENTITIES AND HOW THEY WERE IDENTIFIED. <b>ERROR! BOOKMARK NOT DEFINED.</b>	
FIGURE 1.22	DISCOVERY OF LOW MOLECULAR WEIGHT DRUGS OVER THE YEARS.....	32
FIGURE 1.23	EXAMPLES OF FRAGMENT-SIZED DRUGS USED FOR THE TREATMENT OF TUBERCULOSIS. ....	34
FIGURE 1.24	CLINICAL MANIFESTATIONS OF LEISHMANIASIS.....	35
FIGURE 6.1	TARGET IDENTIFICATION – LEGACY VERSUS EMERGING THINKING.....	285



## LIST OF ABBREVIATIONS

$^1\text{H}$	Hydrogen
$^{13}\text{C}$	Carbon-13
$^{15}\text{N}$	Nitrogen-15
$^{19}\text{F}$	Fluorine-19
AIE	Aggregation-induced emission
BCG	Bacille Calmette–Guérin
BMM	Bone marrow-derived macrophages
BSA	Bovine serum albumin
CAC	Critical aggregation concentration
CNS	Central nervous system
$\text{CCl}_3\text{F}$	Trichlorofluoromethane
CMC	Critical micelle concentration
CLogP	Partition coefficient
CLSM	Confocal laser scanning microscopy
CPMG	Carr-Purcell-Meiboom-Gill
CSA	Chemical shift anisotropy
CSP	Chemical shift perturbations
$\text{D}_2\text{O}$	Deuterium oxide
DENV	Dengue virus
DLBS	Differential line broadening/shifting
DLS	Dynamic light scattering
DMSO	Dimethyl sulfoxide
DOSY	Diffusion-ordered spectroscopy
DSS	Sodium trimethylsilylpropanesulfonate

EC <sub>50</sub>	Half maximal effective concentration
EM	Electron microscopy
ERETIC	Electronic reference to access in vivo concentrations
GAP	GTPase-activating molecules
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GTP	Guanosine triphosphate
FBLD	Fragment-based drug discovery
FDA	United States Food and Drug Administration
FID	Free-induction decay
FKBP	FK506 binding protein
FPLD	Fragment-based phenotypic lead discovery
GHIT	Global Health Innovative Technology
HSQC	Heteronuclear single quantum coherence
HTS	High-throughput screening
IC <sub>50</sub>	Half maximal inhibitory concentration
ITC	Isothermal titration calorimetry
IUPAC	International Union of Pure and Applied Chemistry
K <sub>D</sub>	Binding affinity
L/P	Ligand-to-protein ratio
LE	Ligand efficiency
MAPK	Mitogen-activated protein kinase
MoA	Mechanism of action
MST	Microscale thermophoresis
NME	New molecular entity
NMR	Nuclear magnetic resonance

NO	Nitric oxide
NTA	Nanoparticle tracking analysis
$p_b$	Protein-bound ligand
PBS	Phosphate-buffered saline
PDD	Phenotypic drug discovery
PLD	Phenotypic lead discovery
PI3K	Phosphatidylinositol 3-kinase
qNMR	Quantitative nuclear magnetic resonance
SAR	Structure-activity relationship
SNER	Structure nano-entity relationship
SOS1	Son of sevenless 1
SPR	Surface plasmon resonance
T2	Transverse relaxation
TB	Tuberculosis
TEM	Transmission electron microscopy
TNF	Tumor necrosis factor
TROSY	transverse relaxation optimized spectroscopy
WHO	World Health Organization
ZIKV	Zika virus

# 1 INTRODUCTION

---

With advancements in automation and recombinant technologies, recent drug discovery has relied heavily on high-throughput screening (HTS) where large libraries of molecules (often millions) are tested against disease targets. This approach aims at identifying large numbers of relatively potent drug leads with binding affinities often ranging from the low micromolar to the nanomolar range <sup>1</sup>. However, such large chemical libraries require significant resources to acquire, maintain and use, which often limits the accessibility of HTS to larger organizations. Moreover, this significant library size will often amount to very large numbers of hit molecules to investigate. For example, a hit rate of 1% using a library of 1 million molecules will result in 10,000 hits to evaluate. Considering that HTS assays are often plagued with false-positives, a large proportion of these hits can be artifactual <sup>2-5</sup>. Therefore, confirming these hits can necessitate considerable time and resources and the aforementioned artifacts often leads to low confirmation rates <sup>6</sup>.

Fragment-based lead discovery (FBLD) has been introduced a little over two decades ago and instead relies on the use of low molecule weight compounds to identify binders to macromolecules such as proteins, nucleic acids, etc. (for the sake of simplicity, these macromolecules will henceforth be referred to as “proteins”). The early concept of FBLD originates from a 1981 paper by Jencks where he introduced the theoretical notion that weak affinity small molecules (fragments) could form good quality interactions and could be optimized to improve their binding affinities <sup>7</sup>. However, it took 15 years before the first real application (and success) of FBLD when a team from Abbott implemented the pioneering method coined “SAR by NMR” (*vide infra*) <sup>8</sup>.

There is no absolute definition as to what constitute a fragment but usually these molecules have molecular weights that are below 300-350 Da. Guidelines have also been published as to what properties increase the chances of finding good quality fragment hits. Analogous to Lipinski’s Rule of Five which provided guidelines for the development of orally bioavailable drugs <sup>9,10</sup>, a team from Astex Pharmaceuticals published “Rule of Three” guidelines for the design of fragment libraries: number of hydrogen bond donors  $\leq 3$ , number of hydrogen bond acceptors  $\leq 3$ , ClogP  $\leq 3$ , number of rotatable bonds  $\leq 3$ , polar surface area  $\leq 60$ ) <sup>11</sup>. Although, it must be highlighted that these rules are guidelines rather than hard-set rules as some non-“Rule of Three” molecules can also yield successful chemical matter <sup>12</sup>.

Some calculations of the size of the chemical universe estimate that  $10^{60}$  molecules larger than 500 Da could exist <sup>13,14</sup>, and although it can be argued that a lot of these molecules may not

represent interesting starting points for drugs, the screening a library of 1 million molecules only probes a minuscule proportion of this available chemical space. In contrast, the fragment chemical universe is significantly smaller with one estimate of the chemical universe for molecules below 160 Da being about 14 million molecules<sup>15</sup>. Therefore, screening a typical-sized fragment library of only a couple hundred to thousand molecules covers considerably more chemical diversity space than a standard HTS<sup>16,17</sup>. This represents one of the main advantages of FBLD since screening can be performed on relatively small libraries, several of which are commercially available, making FBLD much more accessible to smaller institutions such as academia and biotech companies, despite being also part of the toolbox of a lot of large pharmaceutical companies<sup>18</sup>. The smaller library sizes also means that a smaller quantity of hits would need to be sorted through as compared to an HTS screen. For example, a 1% hit rate in a fragment-based screen of 1000 compounds will result in 10 hit compounds, whereas a comparable hit rate from a HTS on 1 million compounds would yield 10,000 hits. Fragment screening therefore provides a significantly more manageable quantity of compounds to follow-up on, especially considering that hits from HTS are often plagued with artifacts, as will be discussed in the next section. Additionally, the lower complexity of these fragments is associated with higher hit rates<sup>19,20</sup>. This is due to the fact that as molecular complexity increases, the probability of generating good complementarity with the protein decreases<sup>19,21</sup>. This is illustrated in Figure 1.1, where although the probability of measuring a binding event increases with the size of a ligand (green curve), this binding tends to become less productive/specific as the molecule becomes larger. This is mostly due to the fact that larger molecules will have more atoms that can interact with a protein, even if these interactions are not efficient. On the other hand, too small of a ligand can also result in unspecific interactions and there should therefore be an intermediate size range (Figure 1.1, purple curve) where the probability of observing a “useful” event is larger.

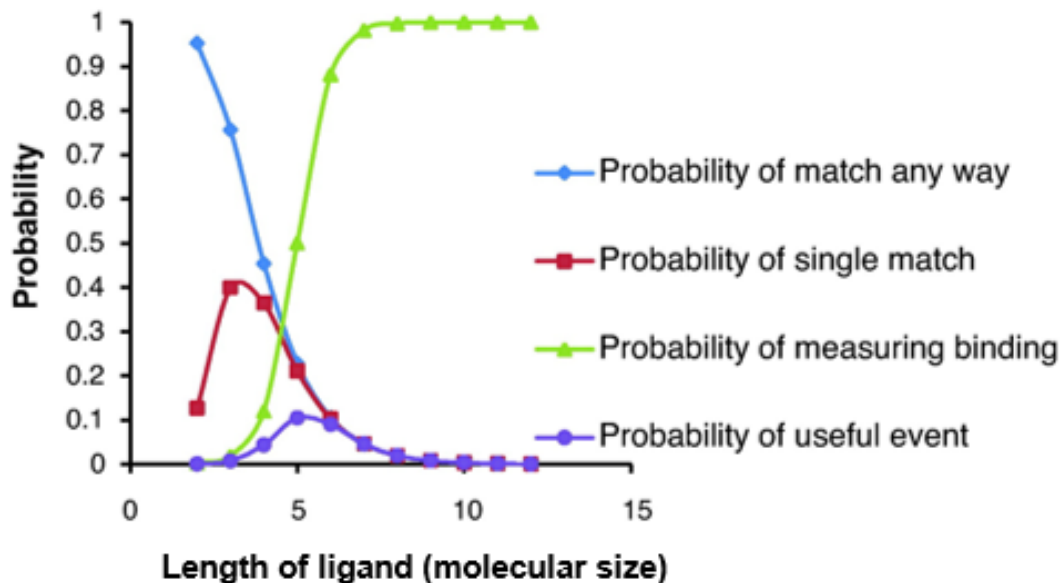
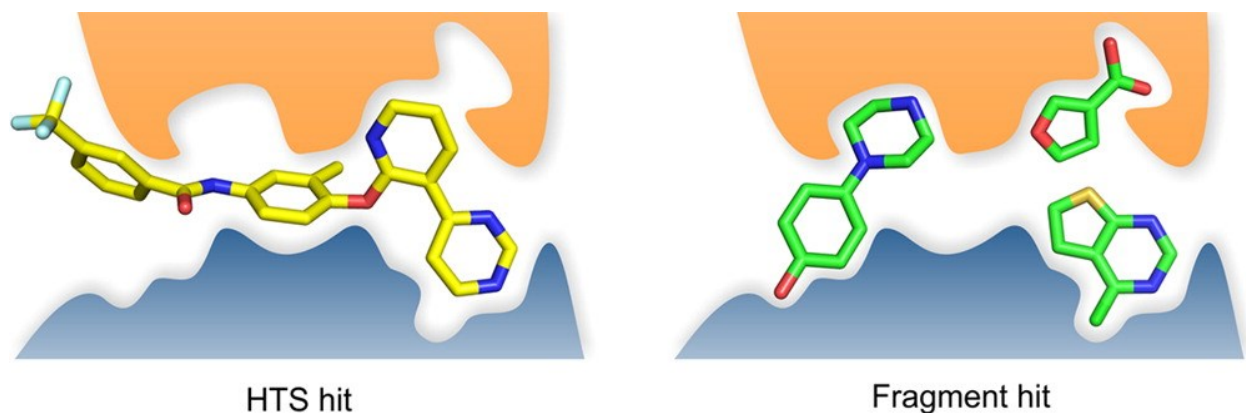


Figure 1.1 Relationship between the size of a ligand and its binding probability.

With increasing size of a ligand, the probability of measuring a binding event between a ligand and a protein increases, as depicted in the green curve. However, the purple curve illustrates that the probability of a “useful” binding event would be limited to a specific range of relatively intermediate molecular size, as depicted by the bell-shaped curve. Adapted with permission from <https://doi.org/10.1016/j.cbpa.2011.05.008>. Copyright 2011 Elsevier.

Along these lines, Figure 1.2 summarizes these aforementioned conceptual differences between HTS and fragment hits. Whereas the molecular size of the average HTS hit is usually larger and can therefore produce more interactions with the target (Figure 1.2, left-hand side), these interactions are expected to be of lower efficiency whereas fragment hits would be expected to exhibit weaker but more efficient interactions due to their smaller sizes (Figure 1.2, right-hand side). The lower fragment complexity also means that molecules derived from FBLD tend to have higher quality interactions which usually translates to better ligand efficiency (LE; binding energy divided by the molecular weight or number of heavy atoms, although other types of LE measures have been proposed) <sup>22-29</sup>, a concept that helps focus on producing drug candidates that are smaller in order to improve properties such as absorption, distribution, metabolism and excretion. This is in line with observations that compounds of larger molecular weights and higher lipophilicity have been associated with higher rates of attrition in the clinic <sup>25,30</sup>.

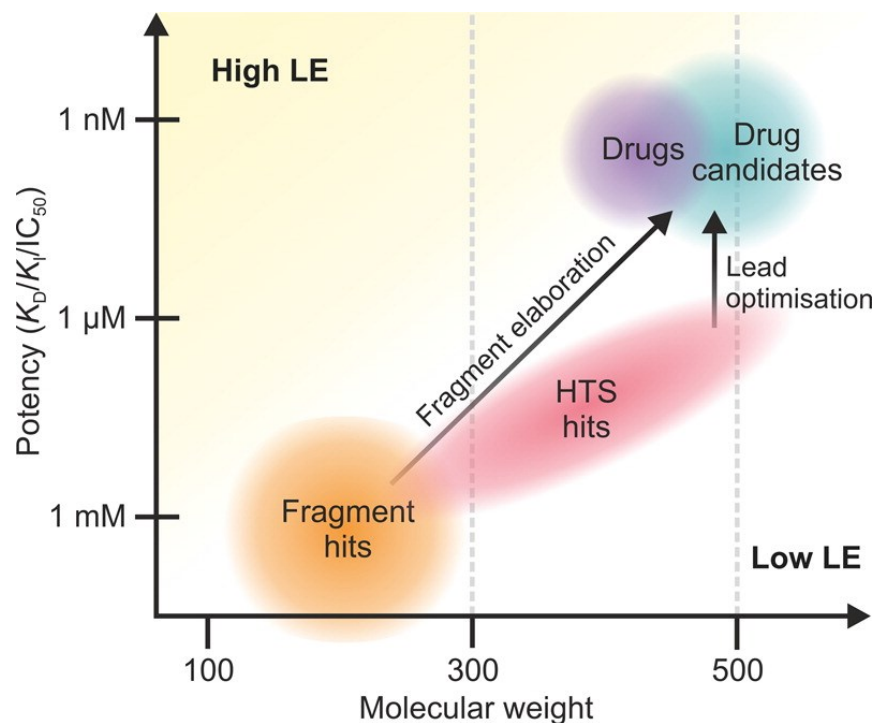




**Figure 1.2 Interactions of an HTS hit compared to a fragment hit.**

**HTS hits usually involve less efficient interaction with the protein target in comparison with fragment hits. HTS: High-throughput screening. Reproduced with permission from <https://doi.org/10.1021/bi3005126>. Copyright 2012 American Chemical Society.**

Regardless of the screening approach used, a molecule will usually need to be optimized by chemical modification/elaboration to improve its affinity/potency in order to get to a point where a drug candidate is obtained. This is illustrated in Figure 1.3 which shows the conceptual differences between hits originating from HTS and FBLD efforts. While HTS hits tend to exhibit better potency from the start, fragments are smaller and weaker binders which need to be elaborated to a greater extent in order to significantly improve their potency in the hope of obtaining a drug candidate. While this simplified depiction makes it appear as if HTS hits would already be closer to the “finish line” due to their better potencies, in practice the less efficient binding interaction of these larger molecules usually require these compounds to be “deconstructed” and regrown in order to improve these interactions. Therefore, FBLD can be considered more of a linear approach as compared to efforts originating from HTS which can be more bidirectional.



**Figure 1.3 Relationship between compound molecular weight and potency.**

Depiction of the molecular optimization for fragment and HTS hits. Fragment hits need to be significantly grown to improve their potency, but they usually exhibit greater ligand efficiency (LE) in comparison with HTS hits that exhibit greater molecular weight and potency on average, at the expense of ligand efficiency. Reproduced with permission from <https://doi.org/10.1021/bi3005126>. Copyright 2012 American Chemical Society.

There are three main approaches for performing fragment elaboration: linking, merging or growing.

As mentioned above, the first successful implementation of FBLD was coined “SAR by NMR” by a team at Abbott to generate chemical leads against FK506 binding protein (FKBP) <sup>8</sup>. The approach they used is depicted in Figure 1.4 where they screened FKBP by 2D NMR and identified a hit for a first binding site. This hit was optimized to molecule **2** with a  $K_D$  of 2  $\mu$ M. They then used **2** to saturate the first binding site in order to identify molecules that bound to a second adjacent binding site. This second hit was then also optimized, resulting in compounds **9** with a  $K_D$  of 100  $\mu$ M. These two molecules were later linked together using various linkers and the resulting molecules exhibited nanomolar affinities for FKBP. This “fragment linking” strategy has had several successes as it relies on a concept that “the whole is greater than the sum of its part” <sup>31,32</sup>, such that the binding affinity of the final linked molecule is expected to be significantly greater

than what would be expected based on the  $K_D$  of the individual compounds before linkage. However, there are significant challenges with this approach. For example, such approach requires various synthetic strategies to generate a “linked” compound with the desired activities as the linking process can alter the binding of the initial fragments <sup>31</sup>. Also, although linking can improve potency, it often leads to larger and less ligand-efficient molecules and can lead to fewer opportunities for optimization <sup>33-36</sup>. Finally, fragment linking usually requires the presence of a second binding site in close proximity with the initial one <sup>1</sup>.

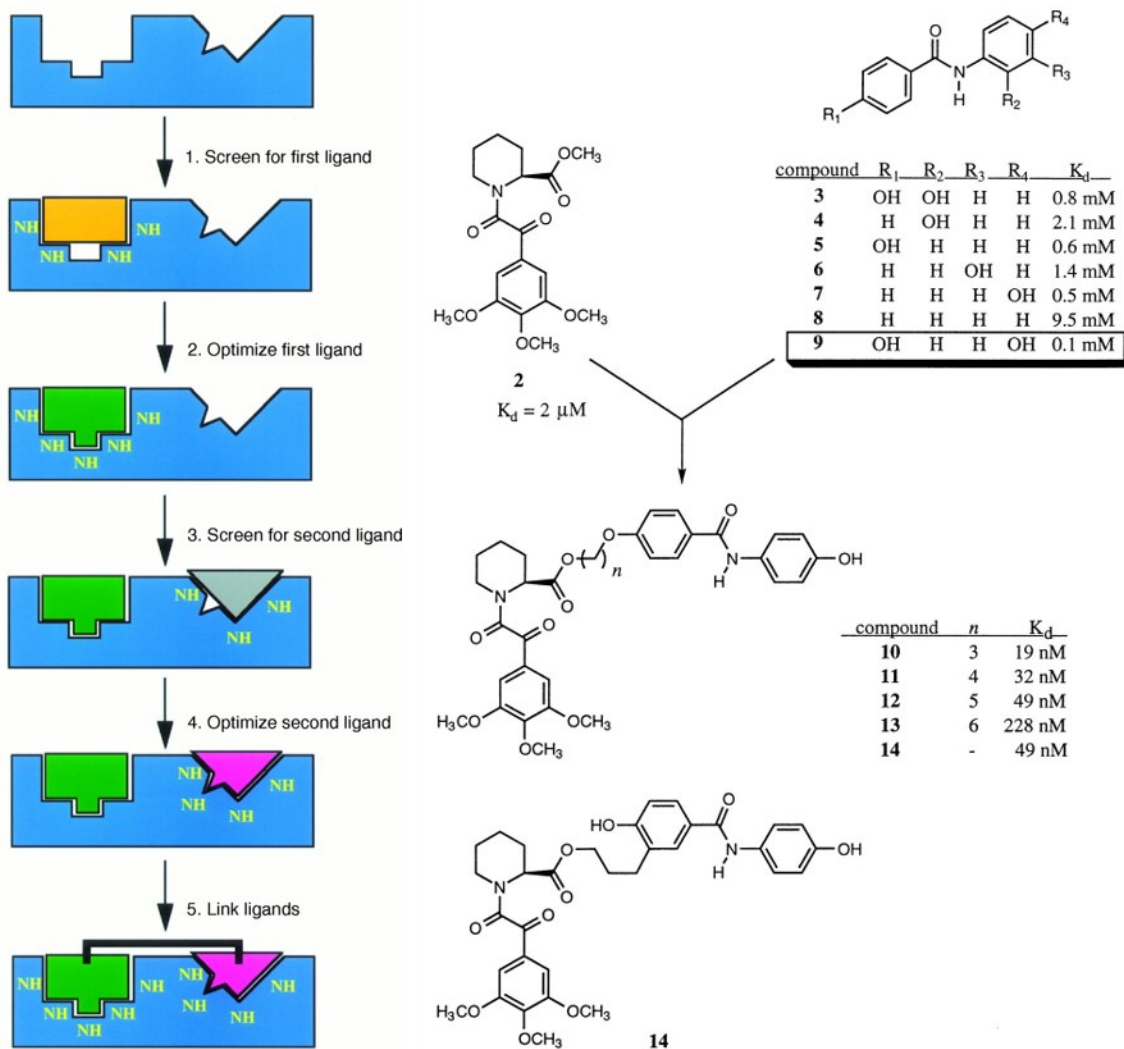
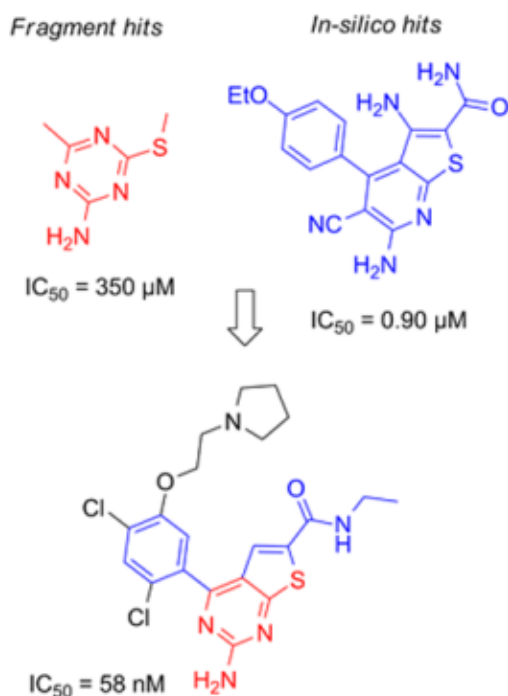


Figure 1.4 Fragment linking strategy.

The first successful lead molecule originating from a fragment-based strategy (SAR by NMR) was developed using a fragment linking strategy to target FK506 binding protein. Adapted with permission from 10.1126/science.274.5292.1531. Copyright 1996 AAAS.

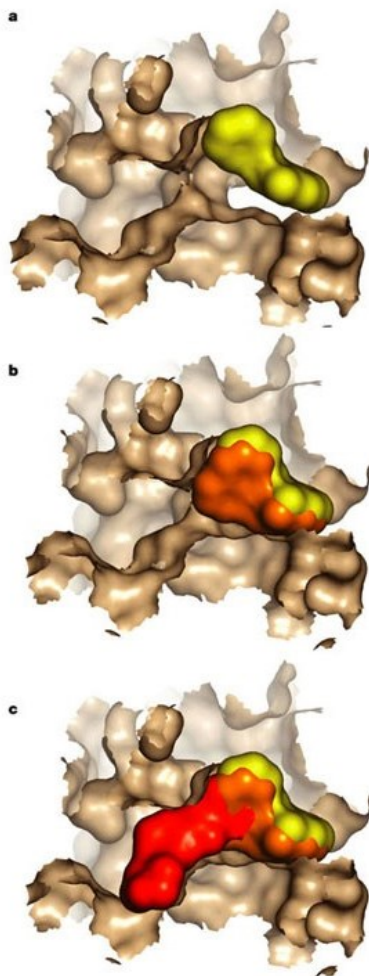
Another fragment optimization strategy is called “fragment merging” and consists into the incorporation of structural components based on the structural overlap of different fragments or known ligands/substrates <sup>37-39</sup>. Figure 1.5 illustrates an example of a fragment hit obtained from a screen at Vernalis against the Heat shock protein 90. The hit scaffold was then combined with the structure of an *in-silico* hit in order to produce a significantly more potent molecule <sup>40</sup>.



**Figure 1.5** Fragment merging strategy.

Example of a fragment merging strategy for a molecule binding the Heat shock protein 90. Adapted with permission from [dx.doi.org/10.1021/bi3005126](https://doi.org/10.1021/bi3005126). Copyright 2012 American Chemical Society.

The most common and usually straightforward optimization strategy is fragment growing (also called evolution or elaboration) where an initial molecule is simply “grown” via chemical synthesis in order to probe additional interactions with the protein and therefore increase its binding affinity or potency. This approach however does not usually result in very large steps in affinity/potency, but rather incremental gains via gradual chemical modifications. Figure 1.6 illustrates a case example of a fragment (depicted in yellow) grown in the ATP-binding site of p38 mitogen-activated protein kinase (growths are depicted in orange and red) <sup>41</sup>.



**Figure 1.6** Fragment growth strategy.

**Surface representation of fragment growth strategy performed against p38 mitogen-activated protein kinase. Reproduced with permissions from 10.1038/nrd1467. Copyright 2004 Nature.**

Obviously, these fragment optimization approaches are not mutually exclusive and hybrid strategies can be used.

Since its inception, a vast number of potent inhibitors against various targets were developed using fragment-based drug design, with several having made their way into clinical trials<sup>1,42,43</sup>. The B-Raf kinase inhibitor Vemurafenib became the first FDA-approved (Food and Drug Administration) drug that emanated from FBLD<sup>44,45</sup>. There are also various examples of FBLD successes against targets that proved to be difficult, or even failures by HTS approaches<sup>34,46-52</sup>. One recent example of such success was the development of Venetoclax, a Bcl-2 inhibitor which was granted a “breakthrough therapy” designation by the FDA for relapsed or refractory chronic

lymphoid leukemia with 17p deletion<sup>31,53</sup>. Considering that FBLD is still a fairly young discipline and the average delay of ~12-18 years between initial screens and drug approval, it is likely that we will see more examples of drugs emanating from this approach in the future, especially given the fact that almost all large pharmaceutical companies use FBLD as part of their drug discovery workflows<sup>1,54-56</sup>.

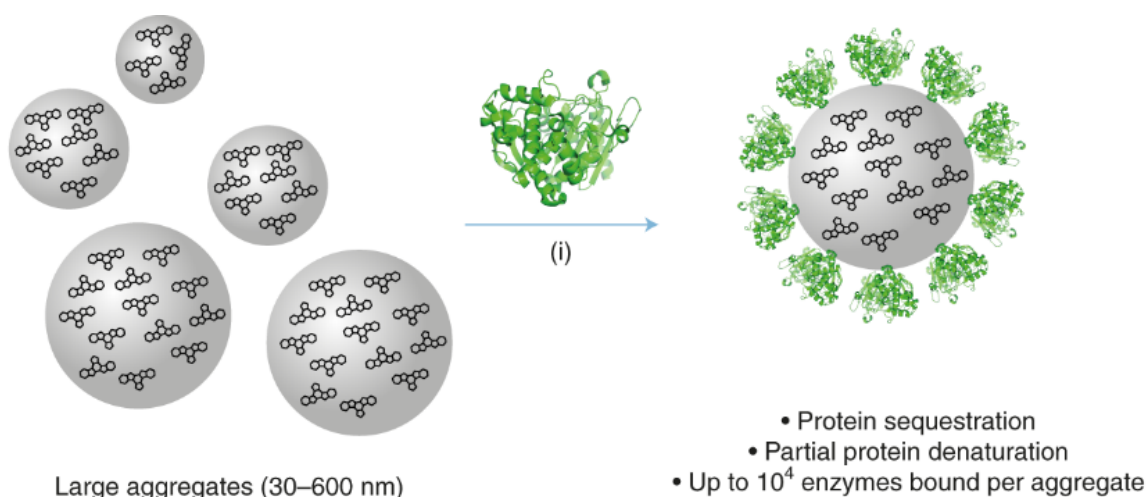
Nevertheless, as with any drug discovery approach, the associated costs as well as the risks of failure are considerable. As such, there is an ongoing need for improving the methods and workflows in order to improve the outcomes. This thesis will therefore describe the implementation of techniques and workflows to monitor compound self-aggregation, the implementation of an NMR-centric platform to discover new drug seeds, guidelines for choice of fluorine NMR references, as well as proof of concepts supporting the viability of combining fragment-based approaches with phenotypic screening.

## 1.1 Monitoring Aggregation in Fragment-based Lead Discovery

Compound self-aggregation is condition-dependent and it has been observed that molecules will adopt a spontaneous assembly into nano-entities above a certain concentration, the critical aggregation concentration (CAC)<sup>2,57,58</sup>, a concept analogous to the critical micelle concentration (CMC) concept used for surfactant molecules. Temperature and aqueous conditions (e.g. salt, pH) also have an effect on the aggregation behaviors of molecules which increases the complexity associated with monitoring or even predicting the phenomenon<sup>59-61</sup>.

Aggregation has been identified as the main cause of artifacts in early-stage drug screenings<sup>62-65</sup>. More specifically, they have been associated with false-positives in functional assays via non-specific adsorption of the enzymes to the aggregate surface<sup>66</sup>, and the binding affinity between some proteins and colloids has been estimated to be in the picomolar range<sup>65,67</sup>. This phenomenon illustrated in Figure 1.7 has been reported to occur via local unfolding events of the enzyme upon non-specific binding to the aggregates, which can result in a loss of enzymatic activity<sup>57,68</sup>. It is estimated that one aggregate can sequester up to 10,000 enzymes and it has also been reported that some aggregates are able to inhibit membrane-bound proteins as well<sup>57,69</sup>. This phenomenon of protein adsorption to colloidal aggregates appears to be surprisingly specific to proteins. Indeed, a study comparing fluorescently labeled DNA (single and double-stranded) did not observe any evidence of adsorption of DNA to colloids<sup>70</sup>. Moreover, that same paper also reported that peptide fragments resulted in significantly lower binding to colloids when

compared to full proteins. Conversely, while aggregates are usually associated with false-positives in functional assays, some aggregates have been shown to result in false-negatives in cell-based assays <sup>71,72</sup>. It was reported that the efficacy of various anticancer drugs was greatly diminished when tested at concentration where they aggregate <sup>71</sup>. This would be due to the large size of these entities which restrict crossing of the compound through cell membranes and has been associated with bell-shaped activity curves in cell-assays <sup>72</sup>. Aggregates could also impact downstream processes in the drug development. For example, aggregates of the calcium channel blocker cilnidipine have been reported to inhibit several cytochrome P450 enzymes which can impact drug metabolism evaluations and can have serious implications for potential drug-drug interactions <sup>73</sup>.



**Figure 1.7** Compounds aggregates can assemble into entities of various sizes.

**Compound aggregates can sequester proteins via partial denaturation. This represents the main mechanism explaining false-positives in biochemical screens. Adapted with permission from <https://www.nature.com/articles/s41557-019-0234-9>. Copyright 2019 Nature.**

Several aggregates can be disrupted upon addition of detergents <sup>2,74</sup>, and this phenomenon is often exploited in various assays which often include some detergents such as Triton X-100 and polysorbate (Tween) 80 in order to mitigate artifacts emanating from the presence of aggregates in the samples <sup>71,75</sup>. However, even by including detergents in screening buffers, there are still significant risks of obtaining artifactual data due to the presence of some aggregates that can remain resistant to the presence of detergent. Moreover, additional artifacts can emerge due to the presence of detergent such as the interaction of non-aggregating compounds with the detergent which could reduce the amount of freely available compound in solution, or the possible

occlusion/alteration of the binding pockets of a protein due the presence of these detergents which could prevent binding of some ligands.

### **1.1.1 Implementing a T2-CPMG NMR Experiment as a Method to Detect Aggregation**

Various strategies have been reported to detect aggregates, the most frequently used one being dynamic light scattering (DLS) since it can provide information regarding to the particle sizes at relatively low concentrations <sup>62,76</sup>. However, DLS has various limitations in terms of sample conditions. For example, it can have difficulty resolving polydisperse samples (e.g. various types of aggregate species, presence of precipitation, etc.) <sup>77,78</sup>. Moreover, compound concentration needs to be optimized and a serial dilution could be needed to determine the optimal concentration <sup>79</sup>, which could have an impact on the properties (e.g. size, distribution) of the observed aggregate species since aggregation is concentration-dependent. Colored/fluorescent samples can also be problematic for DLS and the quality of measurements can also be greatly affected by presence of contaminant such as dust particles <sup>80,81</sup>.

Transmission electron microscopy (TEM) is also a relatively common technique used to detect aggregation <sup>82,83</sup>. Some advantages of TEM include being amenable to samples containing various species of aggregates and being essentially the only techniques allowing a direct, high-resolution visualization of the particles. However, this technique is low throughput and usually requires negative staining as well as dehydration of the sample, which may alter the properties of the aggregate particles.

While useful, one common limitation of both these techniques is that they are sensitive to larger aggregate states. Figure 1.8 illustrates an NMR dilution assay that was previously reported where aggregation is monitored based on abnormal characteristics of the spectra such as changes in chemical shift or peak shapes/number on the NMR spectra of a compound upon dilution <sup>84,85</sup>. Changes in chemical shift would suggest changes in the local environment around the molecule and changes in peak shape would be due to changes in the tumbling rate (size) of a species with a larger entity being expected to exhibit broader peak shapes. Abnormal compound behavior from this dilution assay was found to be highly correlated with promiscuity in off-target assays.



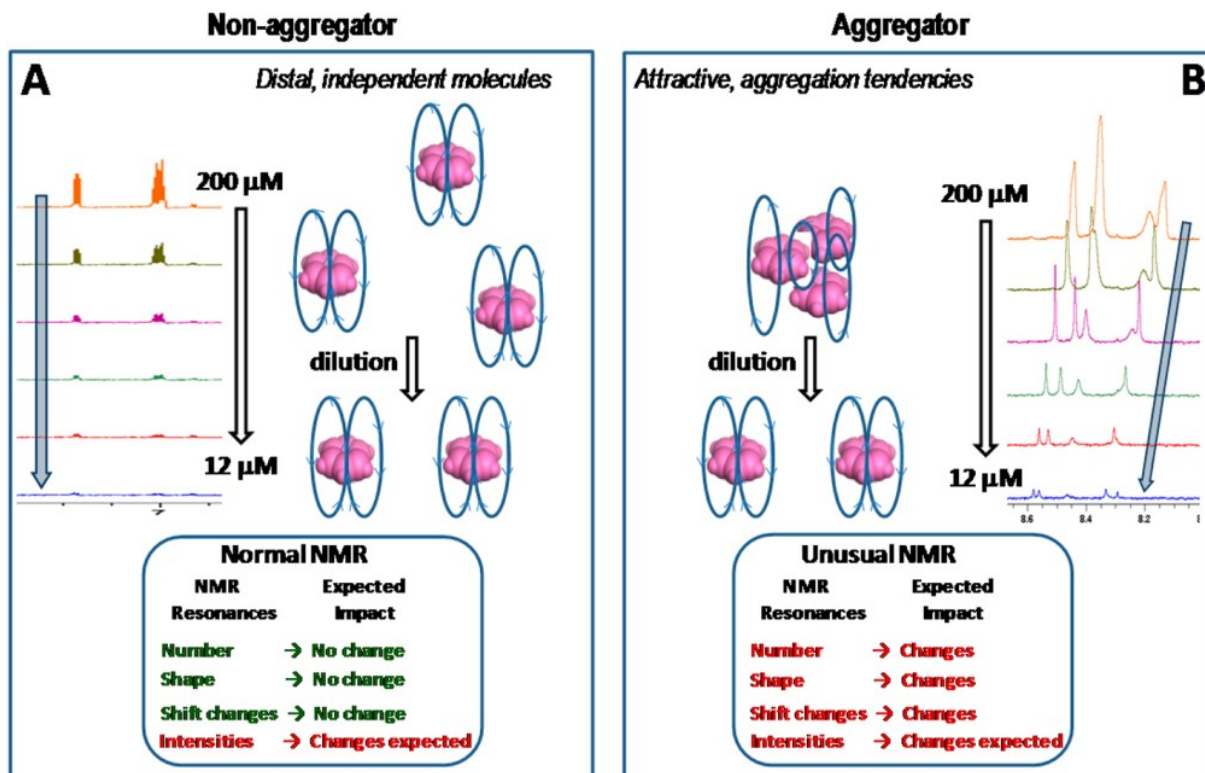


Figure 1.8 NMR dilution assay.

Depiction of the NMR dilution assay for a non-aggregating molecule (A) and an aggregator (B). Upon dilution of a non-aggregator, only changes in signal intensities are expected due to differences in concentrations. However, in the case of an aggregator, changes in the number of peaks, their shapes and chemical shift can be observed. Adapted with permission from [dx.doi.org/10.1021/jm400535b](https://doi.org/10.1021/jm400535b). Copyright 2013 American Chemical Society.

While useful to characterize compound behavior, this dilution assay can be less amenable in a high-throughput setting as it requires the preparation and acquisition of several samples at different compound concentration. There is therefore still a need to develop new methods and strategies that can integrate in a higher-throughput workflow in order to improve understanding and monitoring of compound aggregates.

Relaxation is usually a very sensitive indicator of intermolecular interaction, which is why it has often been used to detect binding between ligand and proteins by NMR<sup>86,87</sup>. Transverse relaxation (T<sub>2</sub> or spin-spin relaxation) has been exploited to study various biological phenomena such as protein dynamics and ligand/receptor interaction<sup>87,88</sup>. The concept of T<sub>2</sub> relaxation is illustrated in Figure 1.9 where at the equilibrium state (A) nuclear spins (depicted by the red arrow) are aligned along the magnetic field in the z axis where they precess randomly. At this equilibrium

state, no signal is detected by the instrument. When a 90° radiofrequency pulse is applied to the sample, these spins become phase-coherent and the net magnetization (green arrow) is then aligned with the x-y plane (B), which allows detection of signal by the instrument. Over time, this coherence will be lost and the net magnetization will diminish, resulting in a gradual loss of signal (C-E). The depiction in Figure 1.9 is oversimplified as it does not include the use of 180° pulses which define the T2-CPMG delay times (delay = (number of echoes/repetition of the 180° pulse \* (time before 180° pulse + time after pulse) = n(t + t)). These refocusing 180° pulses are performed after the initial 90° pulse to remove contributions of magnetic field inhomogeneities on the decoherence, so that most of the decoherence is due to transverse relaxation.

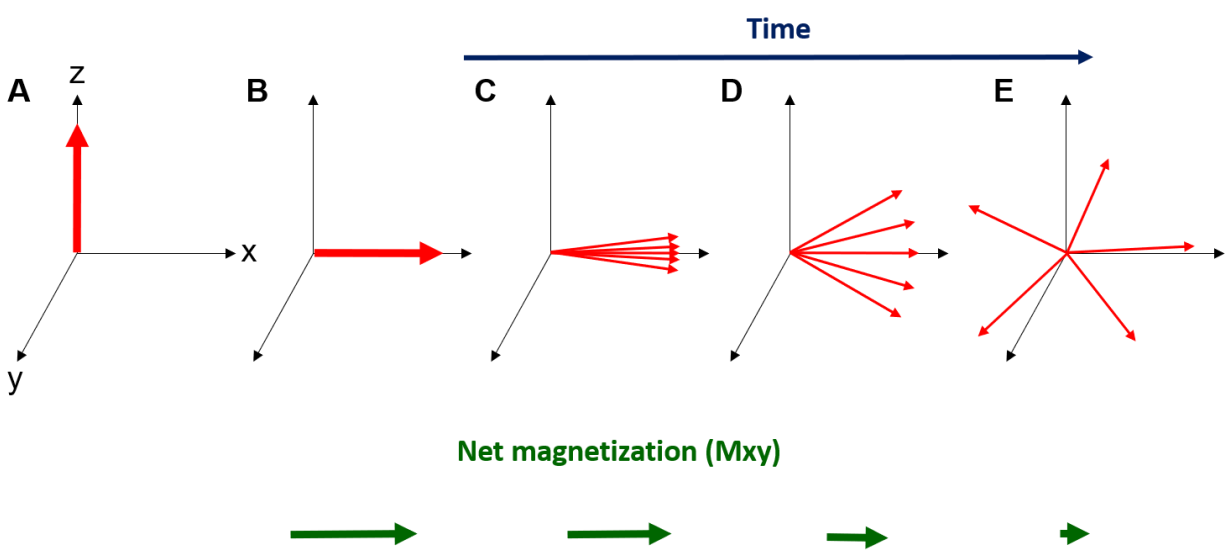


Figure 1.9 Influence of T2 on net magnetization.

Loss of coherence and associated relaxation of magnetization vectors are depicted. Before application of a radiofrequency pulse, the net orientation of the bulk magnetization in the sample is aligned with the magnetic field in the z-axis, which does not result in any magnetization detectable in the x-y-plane, as depicted in A. After a 90° pulse, the magnetization is rotated in the x-y-plane (as depicted in B) which results in detection of signal in the NMR coil. With time, this magnetization loses its phase coherence due to relaxation which results in a decrease in the net magnetization in the x-y-plane depicted in green. This loss of coherence therefore results in a gradual loss of detectable signal. The figure is oversimplified as in a T2-CPMG experiment, a 180° pulse would be used to refocus the loss of coherence to control for the contribution from magnetic field inhomogeneities on the decoherence, so that most of the remaining decoherence is due to spin-spin relaxation. This 180° pulse would normally be applied after a brief period that would fall within the C-D timeframe depicted above. Application of multiple 180° (echoes) is what defines the T2-CPMG delay times.

One interesting property of the transverse relaxation is that there is a close-to-linear relationship between correlation time/molecular tumbling (i.e. molecular size) and  $T_2$ , as illustrated in Figure 1.10. Therefore,  $T_2$  should be a sensitive parameter for the monitoring of compound aggregation due to changes in molecular tumbling.

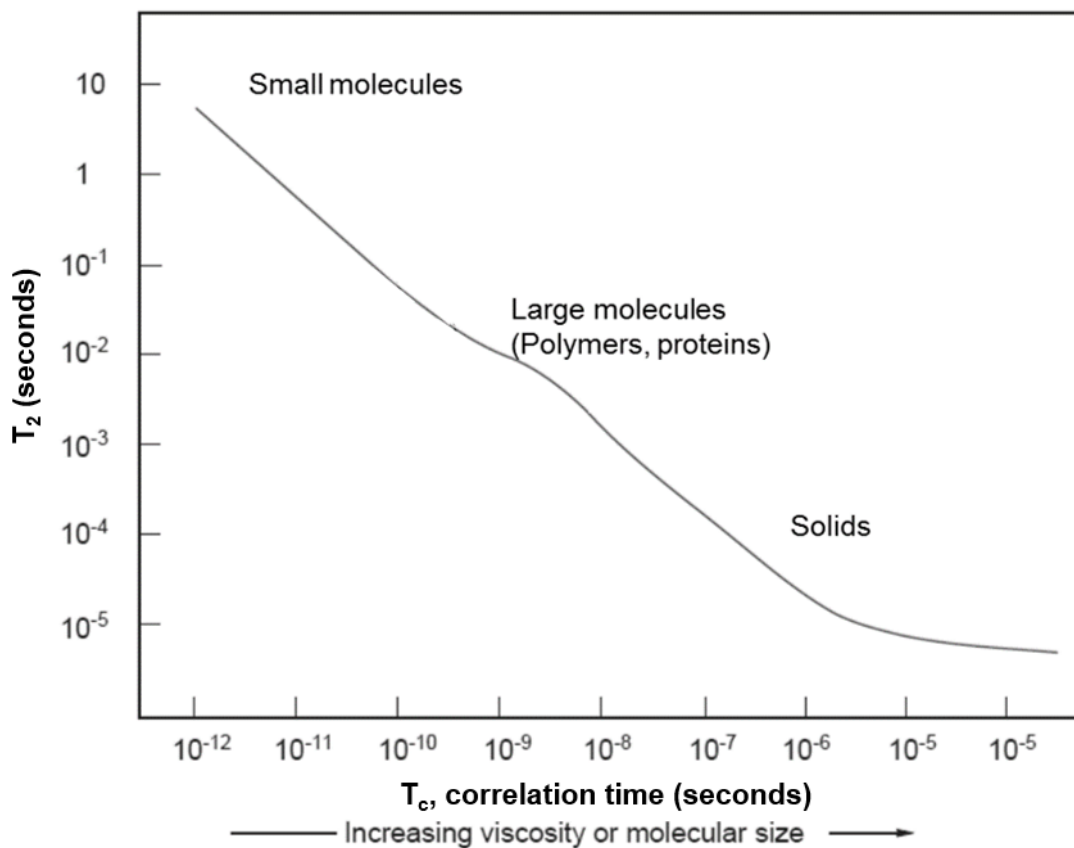


Figure 1.10 Relationship between  $T_2$  and correlation time.

Transverse relaxation ( $T_2$ ) exhibits a close-to-linear relationship with the correlation time of a molecule. This correlation time is dependent of sample viscosity and molecular size. Figure adapted with permission from <https://doi.org/10.1103/PhysRev.73.679>.

Moreover,  $T_2$  is also sensitive to chemical exchanges. Therefore, if a small species such as a monomeric compound is exchanging with a larger aggregate present in solution, this exchange should have an influence on the  $T_2$  of this compound. Figure 1.11 illustrates how the  $T_2$  could report on molecular size (top left), or chemical exchange (top right), with an example of a resulting signal decay (bottom) that could be expected. In reality, contribution from size or chemical exchanges are not mutually exclusive and both can influence the observed  $T_2$  of a molecule.

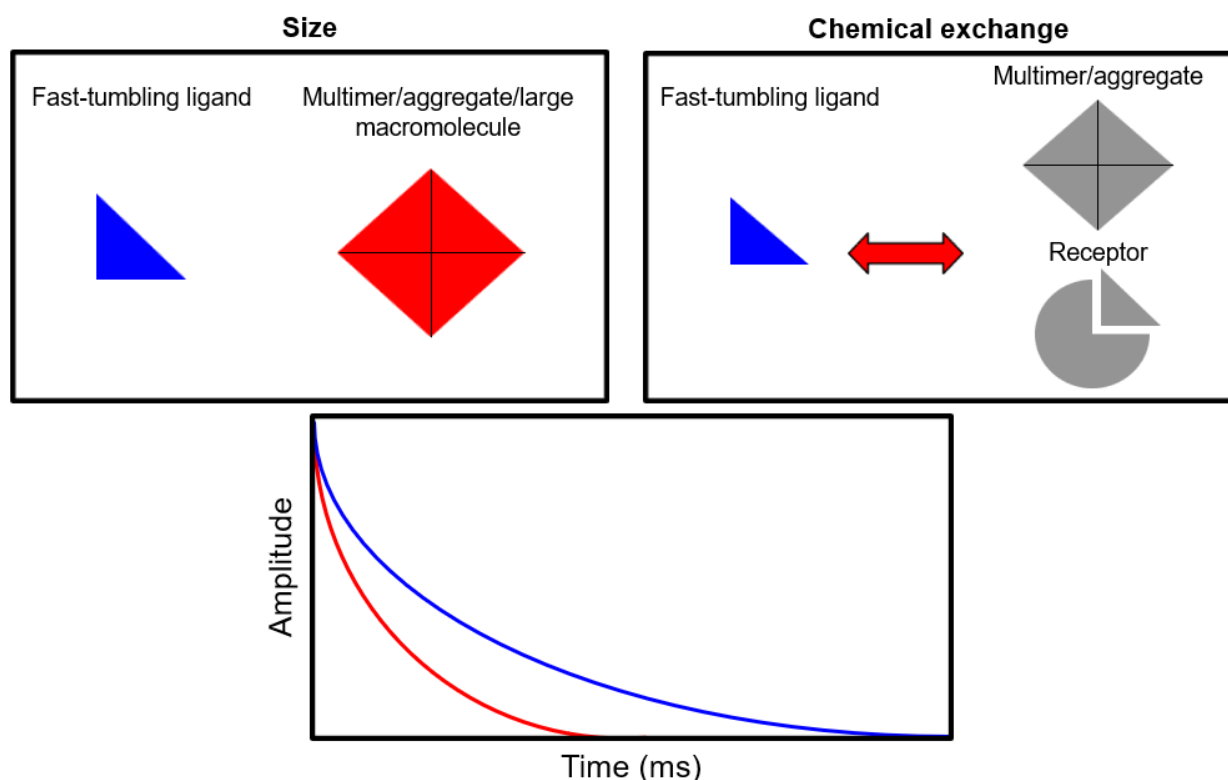


Figure 1.11 T2 to monitor molecular size and chemical exchange.

T2 relaxation can report on molecular size directly via changes in tumbling (top left), or indirectly via chemical exchange between states such as a lone-tumbling small molecule and a larger entity (aggregate or protein receptor). Both size and chemical exchange can contribute to the experimentally observed T2 relaxation and a depiction of the resulting T2-CPMG signal decay over time is illustrated in the bottom box where the lone molecule would be expected to result in a longer relaxation time, as depicted in blue. Conversely, presence of a larger entity would be expected to result in a faster decay of signal over time, such as in the one observed for the red signal decay profile.

The goal of this project was therefore to validate the use of a T2-CPMG experiment in order to monitor compound aggregation.

### 1.1.2 Integrating Multiple Assays Into a Practical Protocol to Monitor Nano-entities

As aforementioned, the various available techniques to monitor aggregation all possess some advantages and limitations. Moreover, these various limitations are not necessarily understood

by everyone involved in drug discovery projects. Therefore, integrating these techniques in a practical workflow that exploits their strength and mitigate their weaknesses should provide a better overall picture of the solution behavior of studied compounds and provide guidance to drug discovery scientists. The goal of this publication was to establish such a protocol to provide a guideline as to what experiments can be performed, in what order, and provide insight as to the interpretation of the data emanating from these various techniques.

## **1.2 Validating an NMR-centric Platform to Generate New Chemical Matter**

Several biophysical methods have been applied to FBLD, such as NMR <sup>89,90</sup>, surface plasmon resonance (SPR) <sup>91,92</sup>, X-ray crystallography <sup>93,94</sup>, isothermal titration calorimetry (ITC) <sup>95</sup>, differential scanning fluorometry <sup>96</sup>, and more recently microscale thermophoresis (MST) <sup>97</sup>. As mentioned above, NMR was the first method to be successfully applied to fragment screening and is to this day still widely used due to its sensitivity to weak binding affinities and its versatility. Indeed, NMR is one of the only techniques that is both label- and immobilization-free, while also providing information on the compound, the protein, and even buffer/solvent impurities.

### **1.2.1 Protein-observed NMR**

Initial SAR by NMR approaches used isotopically labeled (<sup>15</sup>N) protein to identify hits that cause chemical shift perturbations (CSP) on 2D <sup>1</sup>H-<sup>15</sup>N HSQC spectra <sup>8</sup>. These CSPs can be used to extract binding affinity of the molecules by fitting these CSPs with a titration the molecule against the protein as depicted in figure 1.12 (left-hand side). If the protein residues assignments are known, these CSPs can be used to map the binding site of the molecule based on a 3D structure of the protein based on the principle that residues showing significant perturbations should be in close interaction with the molecule. However, ligand binding to the protein can also induce changes further away from the binding site due to allosteric effects or changes in protein conformation so care must be taken when interpreting such data. Oftentimes, most CSPs will be clustered around the same area (as depicted in figure 1.12, right-hand side), which increases confidence about the binding site determination while a random distribution (or absence) of CSPs would be suggestive of nonspecific binding. In the absence of structural binding information (such as a co-crystallized ligand/protein complex), this mapping of the binding site, can be used to restrict docking efforts to the identified pocket in order to help generate medicinal chemistry hypotheses <sup>98</sup>.

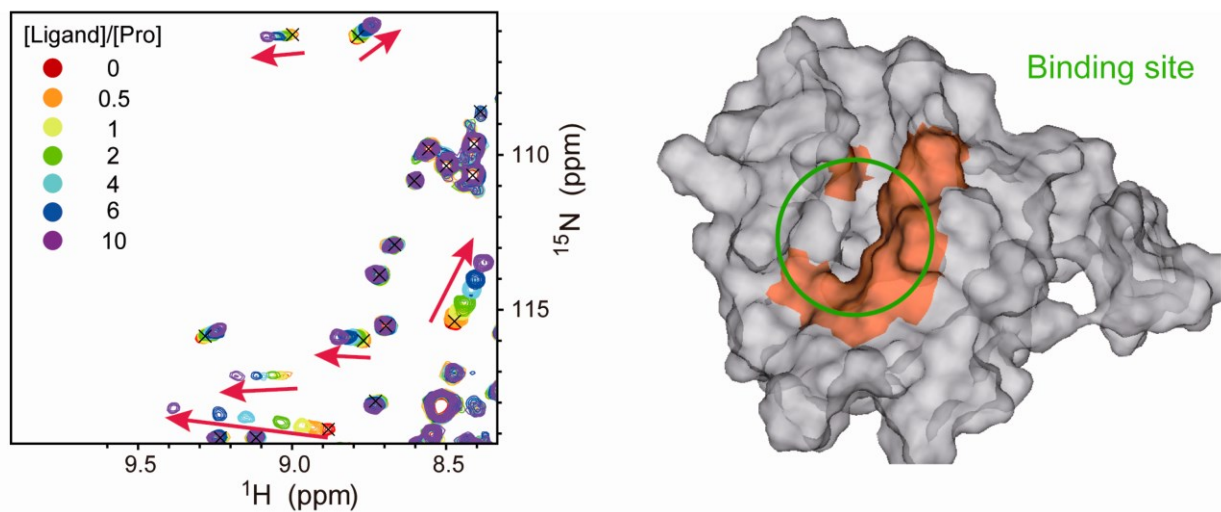


Figure 1.12 Protein chemical shift perturbations.

Chemical shift perturbation of isotopically labeled protein can be used to determine binding affinity from titration of compound (left-hand side). If a 3D structure as well as spectra assignments are available, then these perturbations can also be mapped onto the structure to infer binding site information (right-hand side). Adapted with permission from 10.3390/molecules21070854.

Additionally, the CSP profile upon titration of a molecule can also provide useful information as to the stoichiometry of binding. For examples, on the left-hand side of Figure 1.13, a linear CSP profile is observed which would be consistent with a 1:1 binding mode, whereas the profile on the right-hand side would suggest non-stoichiometric binding (i.e. multiple binding sites).

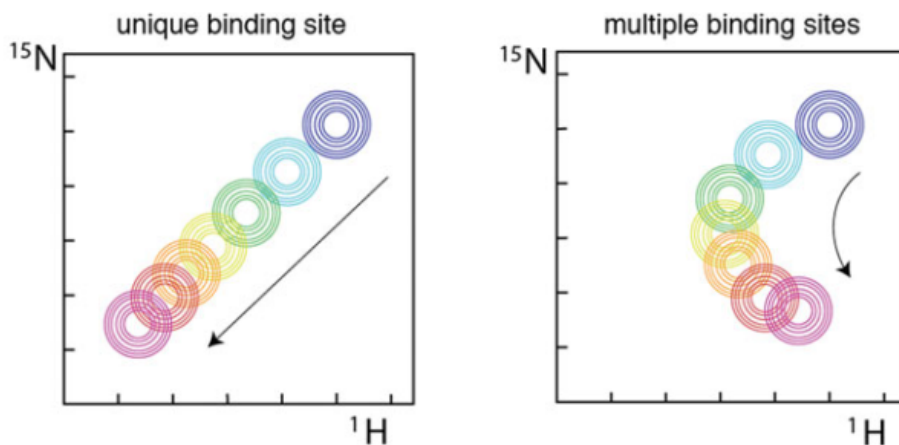


Figure 1.13 Chemical shift perturbations provide insight as to the stoichiometry of binding.

Depicted are protein-observed 2D NMR experiments under fast exchange regime. Titration of a ligand can provide information as to the stoichiometry of binding. On the left-hand spectra, linear changes in chemical shift perturbations are observed which is indicative of a unique binding site. On the right-hand spectra, titration of the ligand results in non-linear changes on the protein resonances, which is indicative of multiple binding sites. Adapted with permission from 10.1002/0471140864.ps1718s81. Copyright 2015 John Wiley & Sons, Inc.

Although this approach is very robust as it can be used to assess binding specificity and stoichiometry, some of its main limitations are the very large quantities of isotopically labeled protein that are required (several hundred milligrams may be required for a screen), as well as the usually longer acquisition times needed for these experiments<sup>99</sup>. Moreover, such typical labeling scheme (e.g. <sup>15</sup>N, <sup>13</sup>C) are usually limited to protein sizes smaller than ~40 kDa due to the increasing loss of signal intensities caused by fast relaxation with increasing protein size. Several methods can be employed to expand this protein size limitation. The most commonly used methods to study larger systems combine labeling strategy and perdeuteration with the use of TROSY (transverse relaxation optimized spectroscopy) experiments<sup>100</sup>. <sup>13</sup>C-methyl labeling of residues such as isoleucine, leucine as well as valine (ILV labeling) is also relatively frequently used<sup>101</sup>. Several other techniques can be adopted and development of new labeling strategies with or without new associated NMR techniques is a very active field of research. However, the use of more “complex” strategies is generally associated with increasing costs of reagents as well as significant decreases in expression yields. Analysis can also become quite time-consuming with several of these strategies which can hinder their use for routine screening purposes. These limitations probably explain why NMR is often used as a secondary assay for validating hits from other methods such as HTS rather than for primary screening<sup>3,102-104</sup>.

### 1.2.2 Ligand-observed NMR

A parallel NMR approach is the monitoring of the ligand instead of the protein. As illustrated in Figure 1.14, there are various properties that can be monitored to detect binding between a small molecule and a protein target. Upon interacting with the protein, a ligand will inherit some of the properties from the protein and these changes in properties will be detected by specific experiments.

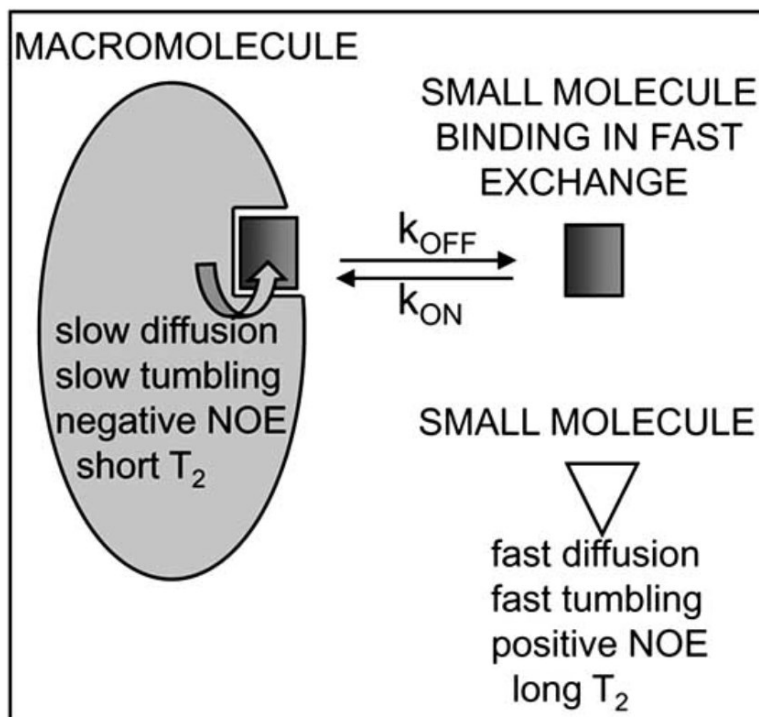


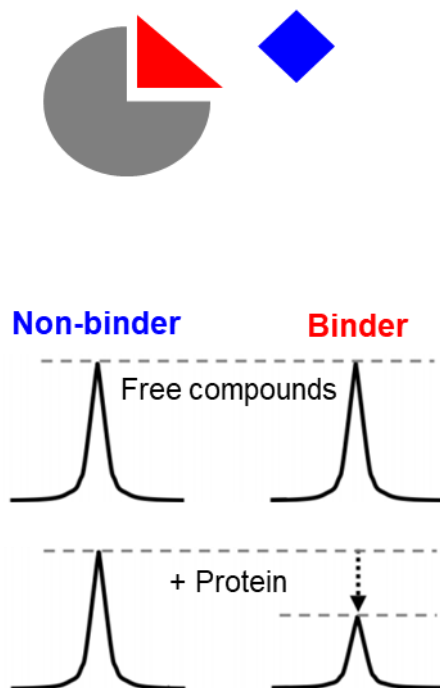
Figure 1.14 Various NMR parameters can report on ligand binding to a protein.

Summary of the NMR parameters that can be used to report on ligand-protein interactions. Upon interaction with a larger macromolecule such as a protein, the small molecule adopts certain properties from the protein which can then be measured using appropriate experiments. Adapted with permission from 10.2174/156802611793611887.

For example, a change in tumbling can result in broadening of the ligand peak which is exemplified in Figure 1.15 where the spectra of the free compounds (a binder and a non-binder) would both be expected to exhibit sharp resonances. However, upon addition of the protein to the compounds, the tumbling of the binder (depicted in red) would slow down due to its interaction with the larger protein, resulting in a broadening of its resonance, as depicted in the lower spectrum. The spectrum of the blue non-binder in this example will not have exhibited any significant change upon addition of the protein.



## Ligand-observed binding



**Figure 1.15** Depiction of a stereotypical binder monitored by traditional 1D NMR experiments.

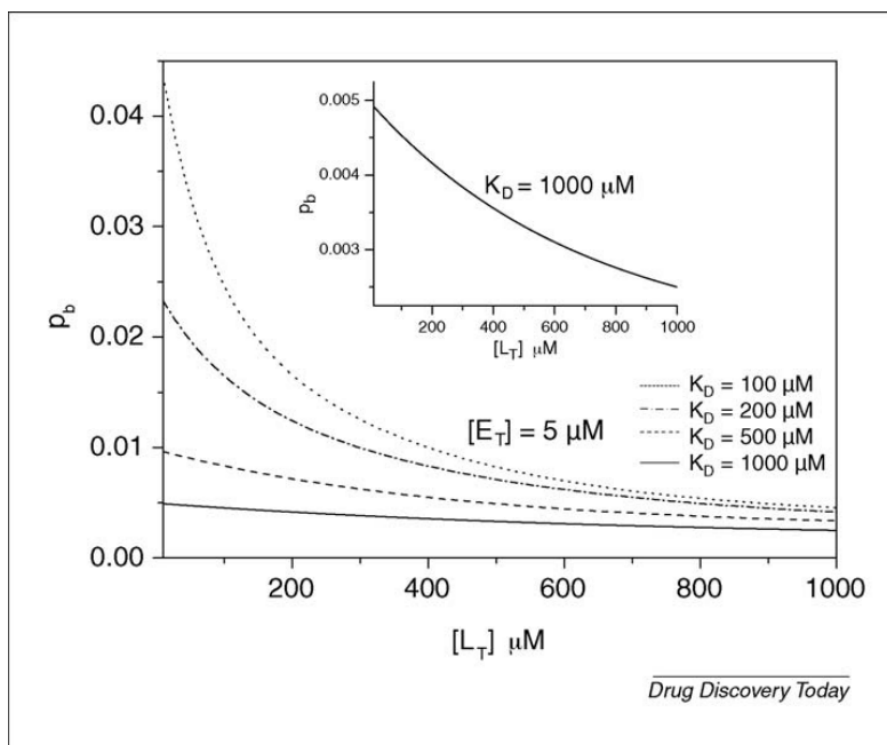
Illustration of the difference in broadening/peak intensity that can be observed for a binder (red molecule) in presence of its target protein. In contrast, a non-binder (depicted in blue) would not exhibit any change on its spectrum in presence of the protein. Copyright 2011 Bentham Science.

While protein-observed methods usually require concentrations of tens to hundreds micromolar of isotopically labeled protein, ligand-observed methods can be used with much lower amounts of proteins (nM to low  $\mu\text{M}$ ). This key difference is due to the concept of fraction of protein-bound ligand ( $p_b$ ,  $f_b$ ,  $X_b$ ), which is described by the following equation<sup>105</sup>:

$$p_b = \frac{[P]_T}{([L]_T + K_D)}$$

where  $[P]_T$  represents the total protein concentration,  $[L]_T$  is the total ligand concentration and  $K_D$  is the binding affinity of the complex. Therefore, for fixed  $K_D$  and protein values, increasing ligand concentration actually decreases the fraction of protein-bound ligands. This is exemplified in Figure 1.16 where the relationship between fraction bound ( $p_b$ ) (depicted on the Y axis) and total ligand concentration (X axis) are simulated based on various  $K_D$  values at a fixed protein

concentration. It becomes evident that as ligand concentration increases,  $p_b$  values decrease in an asymptotic fashion.



**Figure 1.16** Relationship of protein-bound ligand and ligand concentration.

Depicted are simulations illustrating the fraction of protein-bound ligand ( $p_b$ ) at varying ligand concentration ( $L_T$ ). The protein concentration ( $E_T$ ) used for the simulation was  $5 \mu\text{M}$  and curves were generated using four different  $K_D$  values ( $100$ ,  $200$ ,  $500$  and  $1000 \mu\text{M}$ ). A zoomed in depiction of the  $1000 \mu\text{M}$   $K_D$  curve is also visible. Reproduced with permission from 10.1016/j.drudis.2009.07.013. Copyright 2009 Elsevier.

Therefore, in contrast to protein-observed NMR and most other screening techniques, weak affinity ligands such as fragments do not need to be evaluated at very high concentrations. This has several implications for ligand-based NMR approaches. Firstly, because the screening of weak binders does not require the use of very high ligand concentrations, it reduces the risks of artifacts such as aggregation which, as mentioned above, is a concentration-dependent phenomenon. It also allows the use and development of screening libraries that include compounds with relatively lower solubility limits (e.g. tens-hundreds micromolar) than what would often be included in fragment libraries designed to be screened by other biophysical approaches. Another major advantage is that with ligand-based NMR approaches, there is no theoretical limit

as to the protein size that can be screened and the ligand-detected binding even tend to be greater as the size of the protein increases since the tumbling rate of the ligand would consequently be further affected from its interaction with a larger protein.

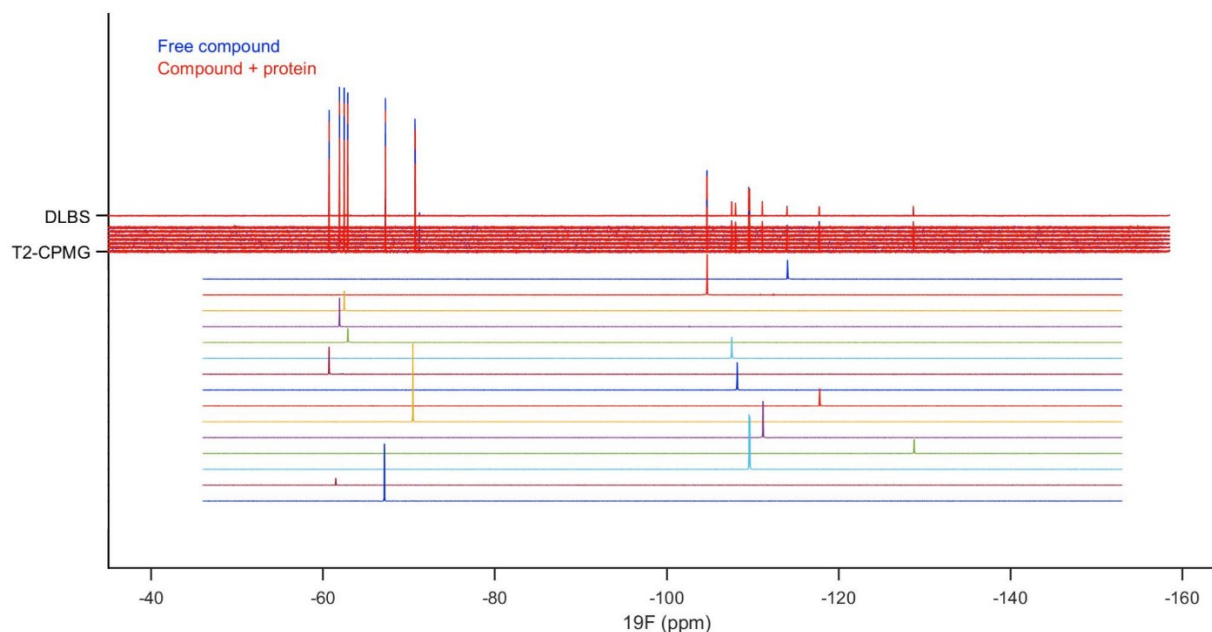
### **1.2.3 Fragment screening**

Screening of fragments is frequently done as pools (mixtures) of compounds in order to increase throughput. While screening fragment libraries as pools can also be done with most techniques, ligand-detected NMR is one of the few that can directly monitor which molecule is binding within a pool, which significantly helps reducing deconvolution time. Only the selected binders need to be retested as singleton molecules in order to validate the observed binding. The choice of how many compounds will be pooled together is often based on finding an equilibrium between throughput and keeping potential artifacts to a minimum <sup>106</sup>. For example, while pooling 100 compounds together might result in extremely fast screening capabilities, the potential problems that could arise due to solubility issues and inter-compound interaction can make such an approach prohibitive. Moreover, the effect that such a large quantity of compounds could have on the protein needs to be taken into consideration. Due to their lower affinities, fragments would often be tested in the hundreds of micromolar, or even low millimolar concentrations using more classical approaches. Therefore, the total “chemical load” can get extremely high. For example, using relatively small pool of 10 compounds which would be screened at 1 mM for each fragment, then the total amount of compounds in solution would be 10 mM. Therefore, it is difficult to predict if that chemical load could have an adverse impact on the protein being screened. On the other hand, screening of pools using ligand-observed approaches allows the use of lower ligand concentration, effectively reducing this total chemical load, further reducing risks of artifacts.

Another parameter that also comes into play is chemical compatibility since not all molecules can be pooled together and therefore, too large pools of compounds can more easily result in compounds interacting with each other <sup>107</sup>. For ligand-observed NMR, one must also consider that each molecule possesses chemical resonances that can overlap with the resonances of other compounds in the same solution. Care must be therefore be taken in order to minimize that spectral overlap in order to allow proper identification of binders within a pool.

#### 1.2.4 Fluorine NMR

While  $^1\text{H}$  NMR experiments have traditionally been used to monitor binding of compounds, fluorine NMR screening has increased in popularity in the recent decades, in parallel with the development of  $^{19}\text{F}$ -tuned cryogenic probes which lead to improved sensitivity along with the optimization of pulse sequences <sup>108</sup>. There are many advantages to the use of  $^{19}\text{F}$  NMR to detect binding. Firstly, fluorine being virtually absent from biology, this allows for a background-free spectrum where only the fluorinated molecule is visible and the use of lower screening concentrations in part due to the absence of any background from components such as the protein. Moreover, some probes allow for the decoupling of  $^1\text{H}$ , resulting in a single sharp resonance for each fluorine moiety, making data analysis even much simpler. The  $^{19}\text{F}$  nucleus is also very sensitive to changes in electronic environments (i.e. interactions). Indeed, at magnetic field traditionally used for screening, the transverse relaxation of the  $^{19}\text{F}$  spin is dominated by chemical shift anisotropy (CSA) which makes it very sensitive to changes in tumbling upon interaction with a protein <sup>109</sup>. Moreover, the large chemical shift dispersion of the fluorine nucleus also contributes to its sensitivity to binding events as this will usually result in large difference in chemical shift between the free and bound state of a molecule, consequently increasing the associated broadening effect that can be observed. This large chemical shift distribution also allows the NMR screening of cocktails containing more fragments than would usually be possible by  $^1\text{H}$  NMR. Indeed, some report the use of > 30 fragments per pool, although mixtures of 10-20 fragments are more typical <sup>108,110,111</sup>. An example of a  $^{19}\text{F}$  pool is depicted in Figure 1.17 which shows the large chemical shift dispersion of a pool containing 15 fragments.



**Figure 1.17 Screening a pool of fragment by  $^{19}\text{F}$  NMR.**

**Example of a pool screened by  $^{19}\text{F}$  NMR. The top part illustrates 2 different screening experiments performed: 1D differential line broadening experiment, T2-CPMG experiment. The bottom part represents the  $^{19}\text{F}$  1D spectrum of each 15 individual fragments for deconvolution purposes.**

Despite all the aforementioned advantages, ligand-observed experiments are often used as a qualitative “Yes/No” binding assessment, whether in an initial screen, or as an orthogonal method to validate binding. Therefore, although ligand-observed NMR is frequently used for initial screening or hit confirmation, it is seldom used as a primary method at the chemical optimization steps where various chemical analogs are tested in order to establish SAR and improve compound affinity. This optimization is usually performed using other techniques to extract binding affinity values.

Therefore, one current challenge with NMR is the affinity ranking of compounds in order to guide SAR efforts. Binding affinities usually require the use of titrations which can be challenging to do by NMR in a routine fashion due to the relatively lower throughput of the method. Another challenge that is not exclusive to NMR is that titrations of the ligand to determine  $K_D$  can be hindered by the solubility limits of the compounds, especially when dealing with relatively weak binding interaction where sufficient saturation often cannot be reached for robust  $K_D$  calculation. Therefore, as described above, ligand-detected experiments can circumvent several of these limitations.

Although binding affinities represent a useful value to possess, it is not a strictly necessary one and a project could very well move forward based on affinity ranking only (i.e. knowing which compounds are better binders and which ones are worst).

A scoring system can therefore be implemented based on the ligand fractional occupancy (fraction of bound ligand,  $p_b$ ) described above. The goal of the current publication was therefore to validate an NMR-centric platform where compounds would be scored based on their ligand-detected binding effects at a single concentration to guide medicinal chemistry efforts. To do so, the protein system HRas<sup>G12V</sup> was used as a model system. A brief description of Ras is included below.

### 1.2.5 HRas

Ras represents the most frequently mutated gene family in cancer. Ras proteins are involved in the activation of many downstream pathways including mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K) pathways that promote cell proliferation and survival<sup>112,113</sup>. As depicted in Figure 1.18, Ras proteins switch between two nucleotide states of GDP and GTP which result in inactive and active states of the protein, respectively. Ras proteins exhibit intrinsic nucleotide exchange and GTP hydrolysis, but these nucleotide states are mostly regulated via activation by guanine nucleotide exchange factors (GEFs) which catalyzes the exchange of GDP for GTP, and deactivation by GTPase-activating molecules (GAPs) which promote hydrolysis of the GTP<sup>113</sup>. Mutations of Ras result in the protein being predominantly in the active state which in turn leads to oncogenic effects.

Efforts at drugging Ras have spanned over several decades without success, which has resulted in the protein being considered as “undruggable”. This has been mostly attributed to the high affinity of the catalytic pocket for native nucleotides (GTP/GDP), as well as the high intracellular concentration of these nucleotides. Therefore, directly competing with GDP/GTP with a small-molecule is unlikely to be successful and efforts have been especially focused on allosteric sites and protein-protein interactions interfaces<sup>114</sup>.

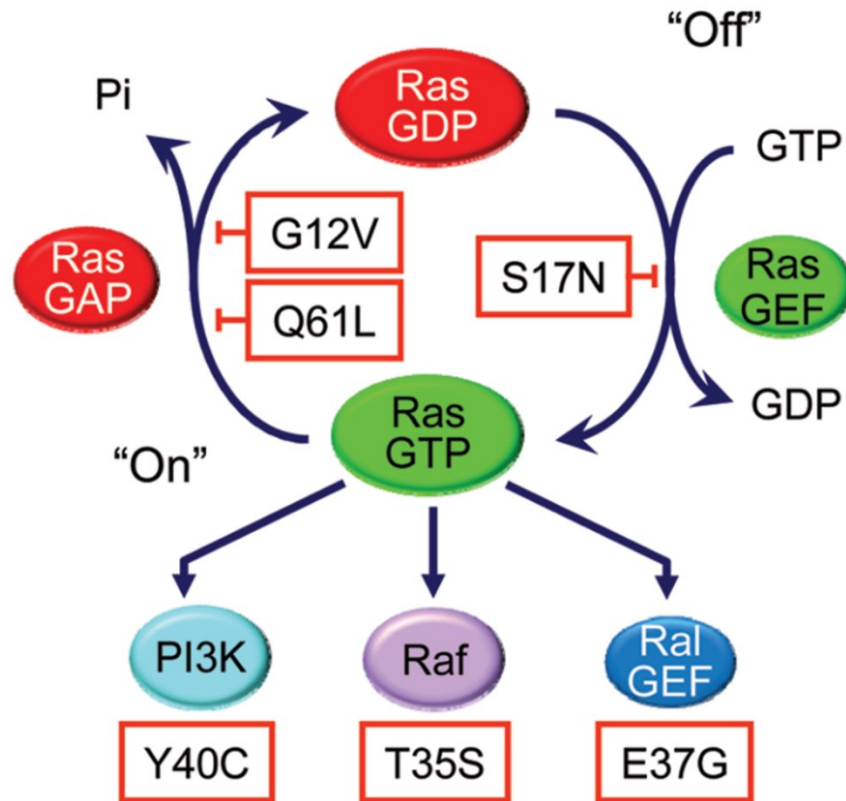
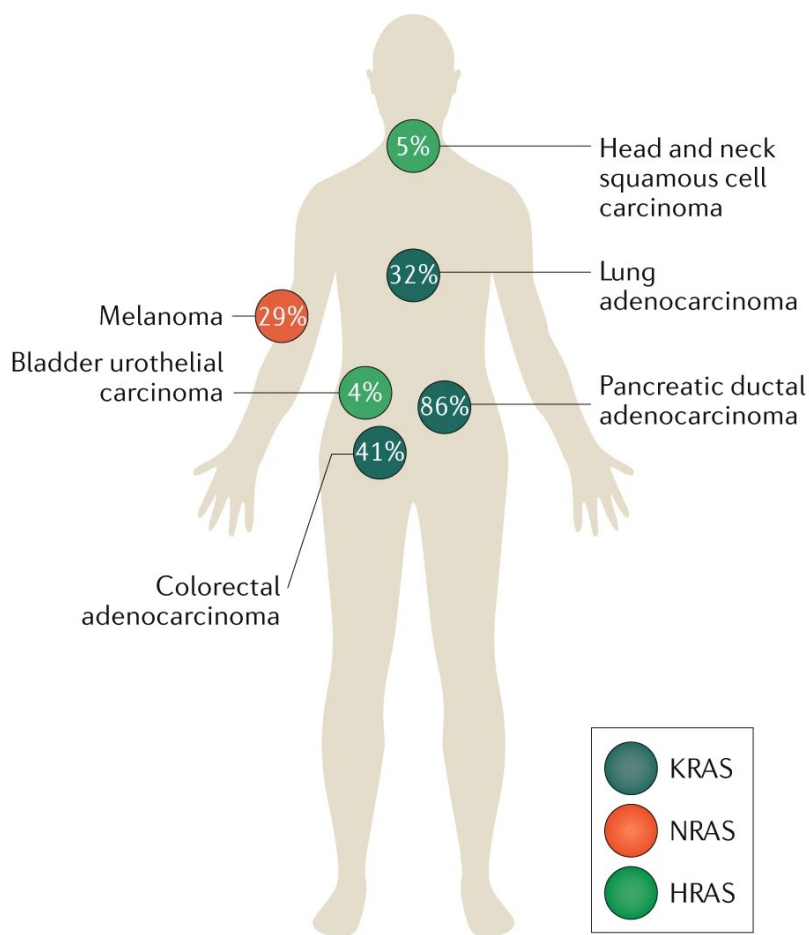


Figure 1.18 Regulation of Ras nucleotide cycle.

Guanine exchange factors (GEF) proteins catalyse exchange of GDP for GTP, therefore bringing Ras into its active state. GTPase activating proteins (GAP) on the other hand accelerate intrinsic hydrolysis of GTP into GDP to bring back Ras into its inactive state. Reproduced with permission from <https://doi.org/10.4161/sgtp.1.1.12178>.

The FDA has recently granted a Fast Track designation for AMG 510, a covalent inhibitor of KRas<sup>G12C</sup> developed by Amgen for non-small-cell lung cancer <sup>115</sup>. This has revived interest in Ras targeting and has provided a proof of concept that drugging Ras proteins is possible and can result in therapeutic benefits. However, covalent binders such as AMG 510 are specific to G12C mutant and so the question of whether other mutated forms of the protein can be successfully drugged with non-covalent molecules remains to be determined. Therefore, there is still a need to identify compounds that can modulate other Ras forms. This has proved very difficult to do in absence of covalent molecules given the shallow binding pockets available on Ras protein surface, which also can explain the lack of success in coming up with effective drugs. There have been some recent successes from fragment-based screening approaches in identifying compounds binding a pocket between the switch I and switch II (SI/II-pocket) in KRas <sup>116-118</sup>. The

molecules identified can modulate Ras function via interference of the protein-protein interaction with the guanine exchange factor SOS1 (Son of Sevenless 1) interaction and therefore, altering Ras nucleotide exchange. Although these compounds represent useful probes, they have yet to result in efficacious clinical candidates.



**Figure 1.19** Distribution of Ras isoforms across types of tumors.

Adapted from 10.1038/s41573-020-0068-6 with permission. Copyright 2020 Springer Nature.

HRas was historically the first Ras oncogene to be identified<sup>119</sup>. However, KRas is the isoform which has received the most attention due to its greater involvement in various cancers. As illustrated in Figure 1.19, HRas mutations are not as frequent as in KRas or even NRas, but they are still associated with 5% of head and neck squamous cell carcinoma and 4% of bladder cancers which are the sixth and fifth most common types of cancers, respectively<sup>120,121</sup>. Therefore, HRas represent a small but significant niche for the treatment of these cancers for



which there is still a need to come up with new treatments. We therefore decided to use the HRas<sup>G12V</sup> protein as a target to validate the aforementioned NMR platform. The G12V mutation was chosen because G12 is one of the most frequently mutated residue in Ras, and G12V is the most frequent G12 mutation in HRas <sup>122</sup>.

### 1.3 Challenges Associated with Fluorine NMR Referencing

As described in the previous section, fluorine is an excellent probe to detect binding between a compound and a protein. The most straightforward and perhaps most robust parameters that can be used to monitor binding are changes in linewidth (or associated changes in peak intensities). However, variation in chemical shift can also be expected if there is a significant difference in chemical shifts between the free and bound states of the ligand <sup>123</sup>. In some instances, especially for very weak binding events, changes in chemical shifts may be observed without any significant change in linewidth.

Unfortunately, various parameters can also cause changes in the chemical shift of a molecule, such as sample-dependent changes (e.g. pH, temperature, buffer composition, etc.) or spectrometer instabilities (e.g. drift in magnetic field over time). The high sensitivity of the fluorine nucleus to its environment exacerbates these variations and these phenomena can therefore skew interpretation of changes in chemical shifts due to binding.

There are different methods to reference NMR chemical shifts. The International Union of Pure and Applied Chemistry (IUPAC) recommends the use of either an internal reference, or a substitution method. The former involves the addition of a reference molecule within the sample in order to adjust the chemical shift value of an experimental sample relative to this reference. A substitution method involves the use of a reference compound into a separate tube and usually requires the user to avoid re-locking or re-shimming onto the experimental sample. Therefore, in a higher-throughput drug discovery context, internal referencing usually represents a more practical approach. However, contrarily to <sup>1</sup>H NMR where there are robust references available such as sodium trimethylsilylpropanesulfonate (DSS) for aqueous conditions <sup>124,125</sup>, there are no official guideline as to a recommended <sup>19</sup>F reference under such conditions. Indeed, IUPAC recommends the use of trichlorofluoromethane (CCl<sub>3</sub>F) for <sup>19</sup>F referencing <sup>124</sup>, but this molecule possesses limited aqueous solubility, is highly volatile at ambient temperature and is an ozone-depleting substance which restricts its commercial availability. In a similar fashion, reported quantitative <sup>19</sup>F NMR standard are meant to be used in organic solvents and their aqueous

solubility is very limited <sup>126</sup>, all of which makes these standards as well as CCl<sub>3</sub>F less than ideal molecules in a screening context.

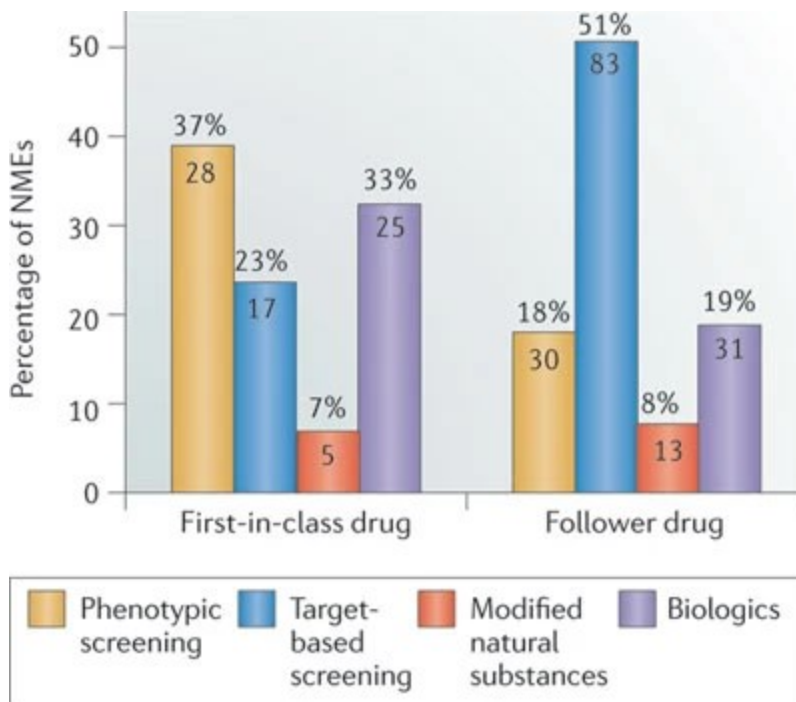
The objective of this publication was to evaluate some of the most reported reference molecules to determine if they would make good candidates for screening purposes. Due to its impractical nature, CCl<sub>3</sub>F was excluded, as well as the aforementioned quantitative (qNMR) standards due to their poor aqueous solubility <sup>126</sup>. Instead, the focus was put on molecules for which there has been reports of use in aqueous media (whether it be in screening context or not), and some other potential candidates that were used in organic solvents to see if they could be suitable in aqueous solutions. An ideal fluorine reference would possess several characteristics. The most important and obvious one being aqueous solubility to be useful in screening conditions. A reference should also be compatible with buffer components such as the buffering agents themselves, salts and other additives that can either be added to the buffer or medium, or be a residual from protein purification. In addition to being compatible with these various elements, an ideal reference should also be stable under aqueous condition throughout the duration of an entire screen. Its sensitivity to various conditions should also ideally be minimal. These conditions include variation in pH, DMSO content, or temperature, as well as having minimal interaction with other molecules in pools of compounds and limited protein binding properties. All of these conditions were therefore evaluated in a funnel-type approach where compounds that fail at one step, would be eliminated from further downstream evaluation in order to obtain a list of final candidates that could be considered suitable references.

#### **1.4 Expanding Fragment-based Lead Discovery to Phenotypic Screening Approaches**

So far, the focus of the previous sections has been on target-based approaches where a protein of interest is purified and screened in buffer conditions. This reductionist, hypothesis-based approach has several appeals, such as allowing better control over the system being investigated by reducing the number of variables present. However, before the relatively recent advancements in protein expression and purification, as well as high-throughput biochemical assay capabilities, drug discovery was based on more empirical approaches involving the use of complex systems such as entire organisms, organs, tissues or cells <sup>127,128</sup>.

One advantage of phenotypic approaches is that they do not require prior knowledge of the molecular mechanism of action. This in turn can complement target-based approach via

identification of new molecular targets <sup>129</sup>. Moreover, the effects in such phenotypic assay may translate into better biological relevance when compared to target-based screens that usually focus on a single isolated protein which does not reflect all of the molecular complexity observed in a living model. Indeed, target-based approaches do not involve the evaluation of various characteristics that are also important for the success of a new drug such as entry into cells, off-target effects, etc., whereas by default, phenotypic assays will usually filter out compounds that might not exhibit desired characteristics.



**Figure 1.20** New molecular entities and how they were identified.

Number of new molecular entities (NMEs) approved by the FDA between 1999 and 2008, and the approach used for their identification. Phenotypic screening approaches was most successful for first-in-class drugs. Adapted from permission from <https://doi.org/10.1038/nrd3480>. Copyright 2011 Springer Nature.

An analysis by Swinney and Anthony of the new molecular entities (NMEs) approved by the FDA between 1999 and 2008 (Figure 1.20) revealed that a significant portion of first-in-class small-molecules were discovered via phenotypic screening approaches. In fact, phenotypic screening was the most successful approaches for the identification of first-in-class NMEs throughout this period, while the opposite trend was observed for follower drugs where target-based approaches were most successful. Considering most of the efforts during this period was focused on target-based approaches, it is also likely that the success of phenotypic screening is actually

underestimated<sup>129</sup>. This report has significantly helped spark a renewed interest in PDD, especially for diseases that lack well-validated molecular targets.

Despite these factors, one limitation of phenotypic screening pertains to the usually lower throughput of these assays, which can make screening of large traditional HTS libraries a resource-intensive endeavor<sup>129</sup>. However, over the recent years, the trend has been toward an overall reduction in library sizes for screening to promote quality over quantity<sup>130</sup>. This change in mentality started to appear when it became apparent that the greater library sizes were not associated with better successes overall. This is highly relevant in the context of phenotypic screening approaches due to the aforementioned challenges and it has been suggested that the use of smaller libraries can yield sufficient hits in phenotypic assay settings, when compared to target-based approaches<sup>131</sup>. Since this observation was based on libraries of more traditional molecules, it can be presumed that the use of fragment libraries could require an even smaller number of molecules in order to identify interesting chemical matter.

Indeed, this mentality of using smaller-sized libraries is aligned with the core concept of fragment screening. However, FBLD has traditionally been used in *in vitro* assays, but very little has been published about the use of fragments in phenotypic settings. A report using functionalized fragments in cells showed that despite their small-size and low affinities, evidence of SAR could be found for most fragment-protein interactions identified<sup>132</sup>. However, these fragments contained photoreactive groups to induce their covalent cross-linking with interacting proteins. Another recent report screened two C-type lectin receptors in cell-based assay using a combination of flow cytometry and competition of reporter molecules<sup>133</sup>. Interestingly, even though the primary screening concentration used was relatively high (2 mM fragments), the hit rates were comparable to what is often observed in traditional *in vitro* fragment screens. They were also able to detect IC<sub>50s</sub> in the low millimolar range and the binding of several compounds could be supported by NMR. Another paper reported *in vivo* optimization of a fragment-sized hit identified from an *in vivo* screen designed to identify novel antipsychotic agents<sup>134</sup>. Although the library was not fragment-based, the hit possessed fragment-like properties. Interestingly, no efforts were made in order to understand the mechanism of action despite the fact that the series was optimized into an advanced lead.

Therefore, there is reason to believe that despite common beliefs among drug hunters, using fragment libraries in phenotypic screening context could be a viable option to identify drug seeds or chemical probes. This is further supported by the fact that although the initial approaches were not fragment-based, many existing drugs actually possess low molecular weights that would make

them fragment-like and as depicted in Figure 1.21. In fact, a large number of these low molecular weight drugs even originated from phenotypic approaches before the rise of target-based screening around the 1970s. This suggests that despite their smaller size, it is possible to observe efficacy with very small molecules.

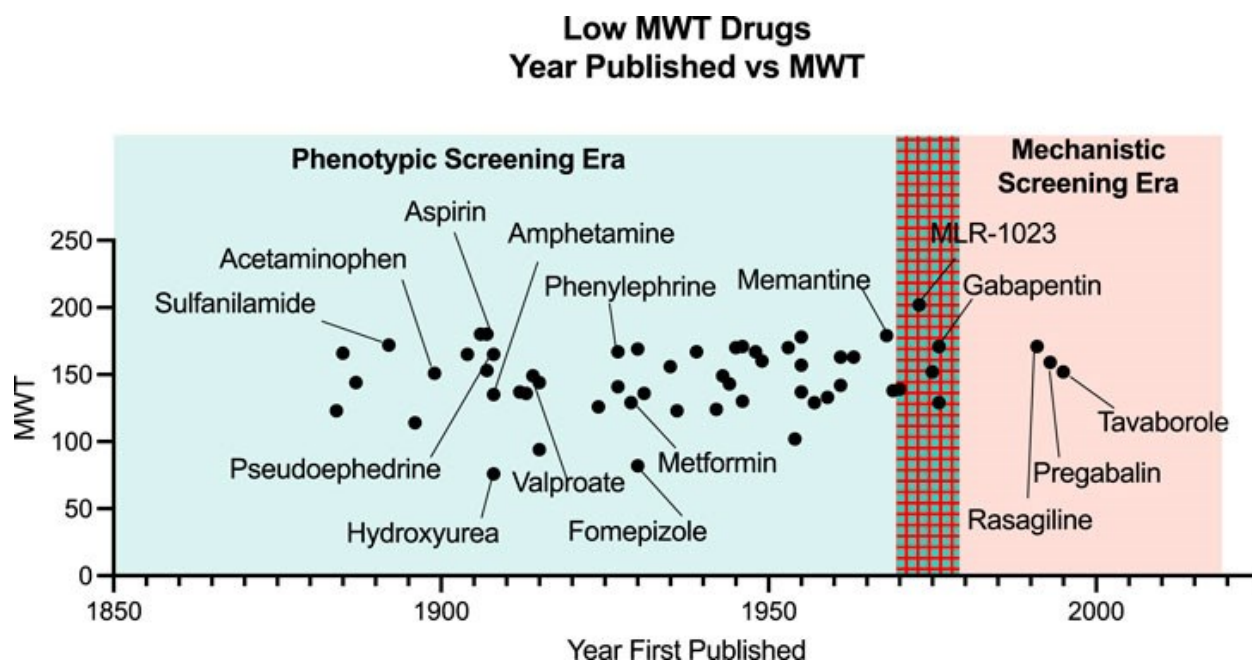


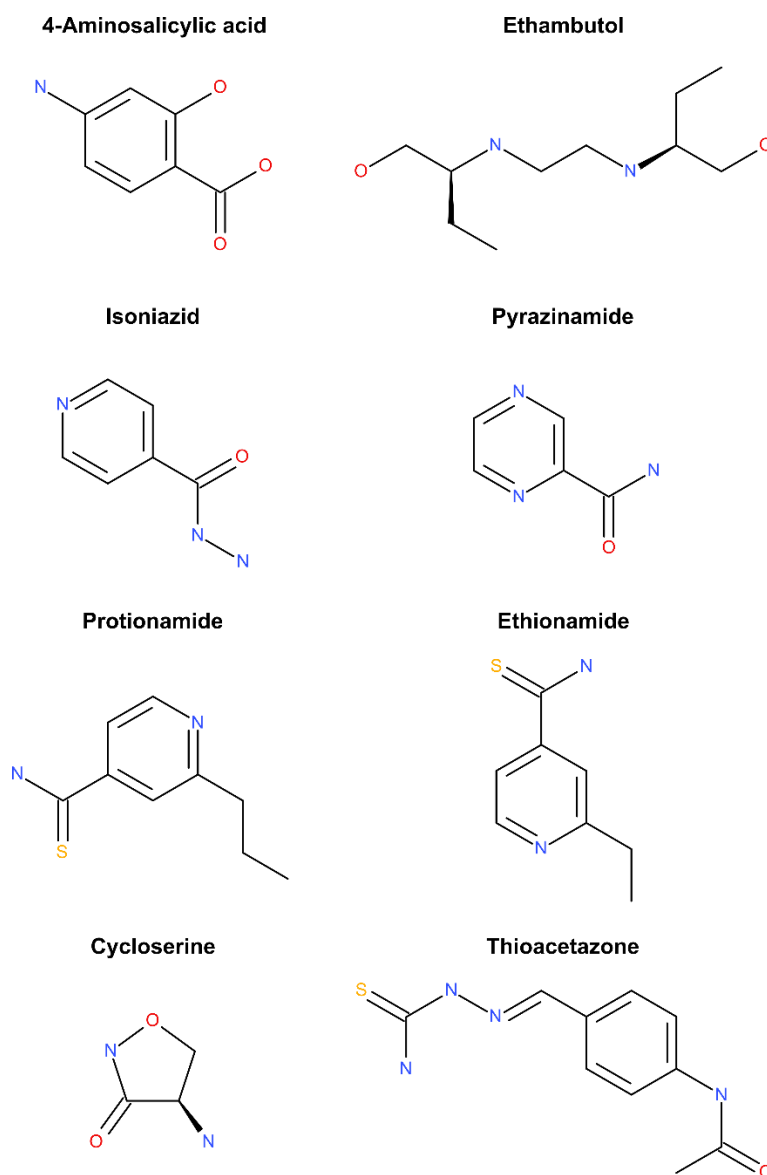
Figure 1.21 Discovery of low molecular weight drugs over the years.

Many low molecular weight drugs (MWT) were discovered through phenotypic approaches before the 1970s. Reproduced with permission from <https://doi.org/10.3389/fphar.2022.917968>.

Another potential challenge with fragment libraries is that as described in previous sections, the screening of pools of compounds is usually associated with target-based fragment screening approaches. However, there have been some examples of successful screening of mixture-based libraries in phenotypic contexts. For example, a recent study described the development of a platform allowing the testing of up to tens of millions of compounds as mixtures into animal models to identify antinociceptive molecules<sup>135</sup>. In another report, researchers selected 1000 compounds computationally from a library of 200,000 drug-like molecules which were randomly combined to generate pools of 10 compounds which were then tested intracerebroventricularly to identify molecules which would promote neuron formation in mouse hippocampus<sup>136</sup>. Therefore, the use of mixtures of compounds appears to be compatible with the phenotypic format of these assays, which should increase their throughput.

Even though, there have been various successes of target-based approaches to identify clinical candidates, there has been limited success in approving new drugs in the field of infectious diseases using these approaches. This can be attributed to various factors such as the small amount of validated drug targets for many infectious diseases, as well as the ability of many infectious organisms to establish compensatory mechanisms that can lead to drug resistance<sup>137-139</sup>. As such, phenotypic approaches tend to exhibit better translatability in the field of infectious diseases due to the usually strong correlation between efficacy in preclinical assays and efficacy in the clinic<sup>140</sup>. Indeed, most of the compounds recently in the pipelines for treatment of leishmaniasis were identified using phenotypic approaches<sup>140,141</sup>. On the other hand, even phenotypic screens have had relatively limited success against mycobacterial membrane since chemical libraries used for those screens are usually similar, if not the same ones, as the ones used for target-based screening which are often more drug-like selected using rule-of-five parameters<sup>142-144</sup>.

Despite their large incidences, there is limited investment attributed to drug discovery of infectious disease, especially tropical diseases. For example, in 2000, ~0.1% of global investment in research was allocated to malaria, leishmaniasis, trypanosomiasis and tuberculosis which combined account to ~5% of global disease burden<sup>145,146</sup>. Therefore, a significant portion of the burden of identifying new chemical matters for these afflictions falls onto the shoulder of smaller institutions and research groups. If valid an approach, the use of fragment molecules in phenotypic assays could allow these smaller entities to jump-start drug discovery initiative with fewer costs and resources. For example, FPLD could prove particularly relevant in the field of tuberculosis treatment, as fragment-sized molecules (Figure 1.22) have been used for the treatment of tuberculosis over the years<sup>147</sup>. One reason that smaller-sized molecules could prove to be a viable option is the mycobacterial cell envelope which represents a challenging permeability barrier for drugs and has resulted in the failure of many target-based approaches<sup>143,148-150</sup>. Several of the drugs illustrated in Figure 1.22 were discovered by cell or animal model screening<sup>151</sup>, and some of them are actually pro-drugs that get activated upon metabolism inside the bacteria<sup>152,153</sup>. It is therefore unlikely that these pro-drugs would have been identified in target-based screening.



**Figure 1.22** Examples of fragment-sized drugs used for the treatment of tuberculosis.

The goal of this project was to perform various proof of concept studies to establish whether FPLD appears to be a viable option in traditional, unbiased phenotypic assays. To do so, a curated fragment library was used to screen various disease models. First, it was applied to *Leishmania* parasites, and then to *Plasmodium* parasites, Zika and dengue flaviviruses, and *Neisseria* and *Mycobacterium* bacteria. Below are brief descriptions of the diseases associated with these pathogens.

### 1.4.1 Parasitic disease systems

*Leishmania* parasites are responsible for leishmaniasis, a group of diseases that affect both human and animals. The World Health Organization (WHO) estimates that there are between 700 000 and 1 million new human cases of leishmaniasis each year, with various species of *Leishmania* having the potential to cause different clinical manifestations of leishmaniasis. Indeed, this disease can be separated into three types depicted in figure 1.23: cutaneous, mucosal, and visceral leishmaniasis <sup>154</sup>. Over the years, there have been a few drugs developed against leishmaniasis, but there is still a need to identify other treatments for this disease as development of drug resistance is not uncommon and some are also associated with significant side effects. Drug accessibility is also a fundamental concern in poorer areas, as illustrated by amphotericin B which is usually administered in liposomal form which restricts its availability due to its greater cost <sup>155</sup>.

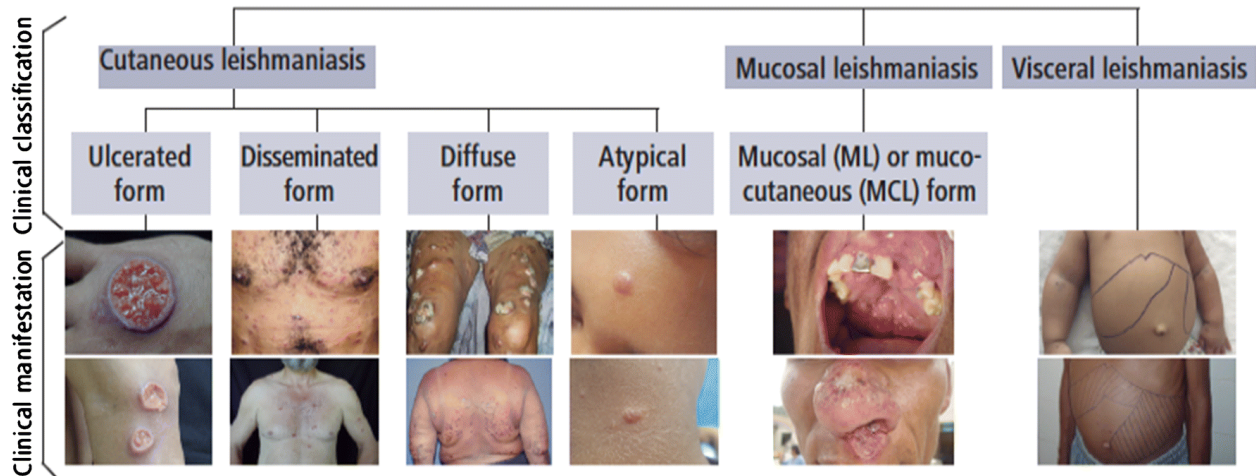


Figure 1.23 Clinical manifestations of leishmaniasis.

Leishmaniasis can be classified into three main clinical categories: cutaneous, mucosal, and visceral. Source: PAHO/WHO, 2018.

*Plasmodium* is a parasite that is responsible for malaria, with *Plasmodium falciparum* being the deadliest species causing malaria in humans <sup>156</sup>. In 2020, the WHO estimated the number of cases of malaria to be around 241 million worldwide, with a death toll above 600,000 during that same year. Despite the recent progress in malaria vaccine development, efficacy of the first RTS,S/AS01 vaccine is modest since the protective effect depends on the presence of antibodies soon after injection of the parasites by the mosquito. When the parasites reach the hepatocytes, the antibodies lose their efficacies due to the loss of expression of the targeted antigen on the



parasite surface (circumsporozoite protein) <sup>157</sup>. Current drug treatments also possess limitations in terms of efficacy due to the complex biology of the parasites and the emergence of resistance to some treatments <sup>158,159</sup>. There is therefore still a need to develop new treatments for the treatment of this disease

#### **1.4.2 Bacterial disease systems**

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis* bacteria. The WHO estimates that 10 million people contracted TB in 2020, with 1.5 million deaths for the same year, making TB the leading cause of death from a single infectious agent. It is also estimated that about one quarter of the world population is infected in a latent manner, which represents a large reservoir for potential future TB cases <sup>160</sup>. The currently available vaccine is 100-year-old (Bacille Calmette–Guérin; BCG) and possesses limited efficacy due to the decrease in protection by adolescence <sup>161</sup>. With the increasing emergence of multidrug-resistant tuberculosis, there is a constant need to come up with new drugs to combat this disease <sup>162</sup>.

Meningococcal diseases such as meningitis and sepsis can be caused by *Neisseria meningitidis* bacteria. Meningitis is fatal in about half cases if left untreated, and 1 out of 5 people surviving the disease can suffer from long-term disabilities (WHO). Although vaccine efforts have significantly helped reduced occurrence in developed countries, meningococcal diseases are still an ongoing problem in underdeveloped areas, particularly in the African meningitis belt (sub-Saharan region) <sup>163,164</sup>, partly due to the challenges in deploying vaccination campaigns. Another challenge pertaining to vaccination against *N. meningitidis* pertains to the existence of 12 meningococcal serogroups, 6 of which are responsible for most disease cases. Of these 6 most common serogroups, vaccines are available against 5 of these 6. Therefore, even fully deployed vaccination campaign does not currently protect against all meningococcal serogroups <sup>165,166</sup>, and global emergence of antibiotic-resistant bacteria highlights the need to keep developing new treatments.

#### **1.4.3 Viral disease systems**

Dengue disease is caused by viruses of the same name for which about 200-500 million infections are estimated to occur each year <sup>167,168</sup>. Although most cases of infection result in either asymptomatic or mild symptoms (dengue fever), the disease can result in serious illness (dengue shock syndrome), which is estimated to result in the death of about 500,000 to 1 million individuals

every year <sup>169</sup>. There are currently no specific treatments or for dengue infections, and the searches for effective vaccines are still ongoing.

Zika virus is another flavivirus for which there are no specific treatment available, nor any efficacious vaccine <sup>170,171</sup>. Although most people with Zika infections will be asymptomatic, it can result in neurological complications in some individuals, such as Guillain-Barré Syndrome <sup>171</sup>. It is also associated with increased congenital abnormalities for the fetus, as well as pregnancy complications for the mother. The incidence of Zika infections around the world is very difficult to estimate due to the large proportion of asymptomatic cases (about 80%), as well as cases with mild, non-specific symptoms. However, to give an idea of the magnitude of contagion, there has been between 440,000-1,300,000 suspected cases of Zika virus disease during an outbreak in Brazil in 2015 <sup>172</sup>. As is the case for dengue fever, there are still no specific treatment for Zika disease.

## **1.5 Summary of challenges, hypotheses, and research objectives**

### Validation of an NMR relaxation assay to monitor aggregation

Compound nano-entities (aggregation) can lead to artifactual results in binding and functional assays. Being a concentration-dependent phenomenon, the relatively higher compounds concentrations often required to detect weak binding from fragments can increase the risk of nano-entity formation. Detection of these nano-entities is still challenging and despite being a versatile technique to monitor these entities, currently published NMR assays are less amenable to high-throughput. Therefore, FBDD would benefit from a higher-throughput NMR technique to detect nano-entities. Our hypothesis was that transverse relaxation (T<sub>2</sub>) could be exploited to monitor the presence of nano-entities using a T<sub>2</sub>-CPMG NMR assay. We therefore validated the feasibility and performance of such an assay to detect nano-entities.

### Integrating multiple assays into a practical protocol to monitor nano-entities

As was mentioned above, detection of nano-entities is still challenging to this day. One limiting factor for many scientists is that no guidelines are available to characterize the solution behavior of compounds in aqueous solution. We therefore wanted to implement a protocol that takes advantage of various assays in order to propose a practical workflow that non-expert scientists can follow, along with possible scenarios and interpretations.

### Validating an NMR-centric platform to generate new leads

Compounds emanating from a fragment screen will typically possess weak binding affinities. NMR is the one of the most sensitive methods to weak binding events and is therefore a useful technique for assessing binding of fragments. However,  $K_D$  determination usually requires the use of titrations, which is not practical for routine testing of compounds by NMR. Moreover,  $K_D$  measurements are also often limited by the compound behavior in solution (solubility limit, aggregation). Despite being useful,  $K_D$  values are not essential for an early drug discovery project to move forward. Indeed, a relative affinity ranking of the molecule can be sufficient to guide early medicinal chemistry efforts. We therefore validated the use of an NMR-centric platform coined “NMR for SAR” which is centered around the use of relatively high-throughput ligand-observed NMR methods to rank order compounds based on their relative affinities. This ranking could be used to guide medicinal chemistry efforts, while also providing a lot of control over the system by being able to monitor both the ligand and protein in solution. HRas<sup>G12V</sup> was used as a test system for this platform. Ras proteins have long been considered as “undruggable” targets due to their few shallow binding pockets.

### Practical Considerations and Guidelines for Spectral Referencing for Fluorine NMR Ligand Screening

As previously mentioned, the fluorine (<sup>19</sup>F) nucleus is very sensitive to change in its chemical environment, which makes it a very useful tool to detect weak binding of a ligand to a protein. Changes in chemical shifts can be used to monitor binding events, especially in the absence of significant changes in linewidth. However, sample-dependent variations can also affect the <sup>19</sup>F chemical shift and skew binding interpretations. Therefore, in a drug discovery context, the use of an internal reference molecule represents one of the more practical approaches to correct for such artifactual variations in chemical shifts. However, while robust references are available for <sup>1</sup>H NMR, no guidelines exist for <sup>19</sup>F referencing in aqueous conditions. Therefore, we evaluated a subset of fluorinated molecules to determine if they represent good candidates for screen purposes. To do so, we implemented a workflow that can be adopted by others in order to expand the search for robust <sup>19</sup>F reference molecules.

### Expanding fragment-based screening to phenotypic approaches

Phenotypic screening is a complementary approach to target-based approaches where a screen is performed directly on an organism (e.g. cell, tissue, animal, etc.), often without any prior knowledge about the molecular target. While phenotypic approaches possess many advantages, one of its main limitations pertains to the often relatively lower throughput of many of these phenotypic assays, making them less amenable to the screening of large libraries. Fragment screening has historically been used in target-based contexts due to the assumptions that low affinity fragments would result in non-specific effects in phenotypic assays. However, many fragment-sized drugs have historically been discovered from phenotypic approaches so one can wonder if these examples are exceptions rather than the rule. We therefore wanted to evaluate the hypothesis that fragment libraries could represent a viable alternative to traditional, larger libraries for phenotypic screening. To do so, we performed several proof-of-concept studies and screening 6 disease models: *Leishmania* and *Plasmodium* parasites, Zika and dengue flaviviruses, and *Neisseria* and *Mycobacterium* bacteria.



## 2 ARTICLE 1 – VALIDATION OF AN NMR RELAXATION ASSAY TO MONITOR COMPOUND AGGREGATION

---

### Exposing Small-molecule Nano-entities By a Nuclear Magnetic Resonance Relaxation Assay

Yann Ayotte,<sup>1,‡</sup> Victoria M. Marando,<sup>2,‡</sup> Louis Vaillancourt,<sup>2</sup> Patricia Bouchard,<sup>2</sup> Gregory Heffron,<sup>3</sup> Paul W. Coote,<sup>2,3</sup> Sacha T. Larda,<sup>2</sup> and Steven R. LaPlante<sup>1,2,3</sup>

<sup>1</sup> INRS-Centre Armand-Frappier Santé Biotechnologie, 531 Boulevard des Prairies, Laval, Québec, H7V 1B7, CANADA

<sup>2</sup> NMX Research and Solutions, Inc., 500 Boulevard Cartier Ouest, Laval, Québec, H7V 5B7, CANADA

<sup>3</sup> Harvard Medical School, 240 Longwood Ave., Boston, MA 02115, USA

‡ These authors contributed equally to this work

Published in: Journal of Medicinal Chemistry, 2019, 62, 17, 7885-7896.

DOI : 10.1021/acs.jmedchem.9b00653

#### Authors contributions :

Yann Ayotte and Victoria Marando contributed equally to the content and successful publication of this article.

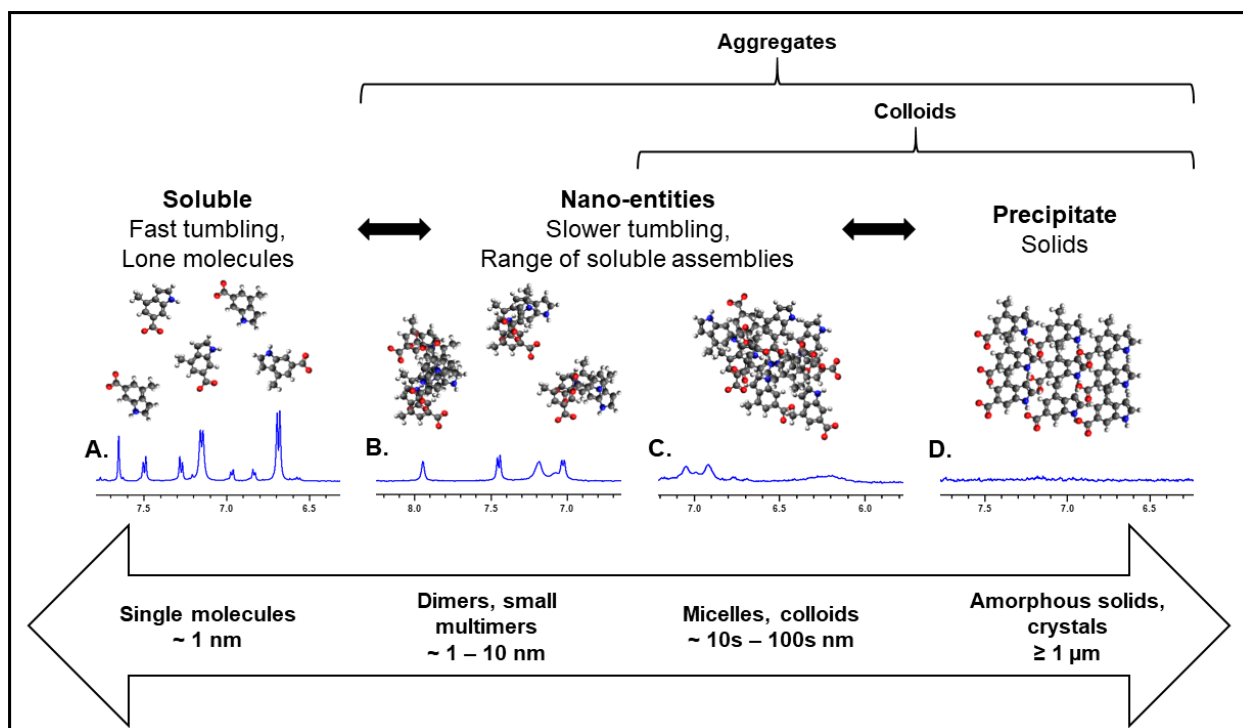
Yann Ayotte and Victoria Marando designed and performed the T2-CPMG, dilution, and dynamic light scattering experiments as well as data analysis under supervision from Sacha Larda and Steven Laplante. Yann Ayotte performed diffusions experiments and analysis under supervision from Sacha Larda. Paul Coote also contributed to data analysis. Gregory Heffron helped implement the NMR pulse sequences used in this article. Louis Vaillancourt and Patricia Bouchard contributed to the library curation which generated the data on which some figures are based on. Yann Ayotte and Victoria Marando co-wrote the manuscript and generated the figures. Sacha Larda and Steven Laplante revised the manuscript.

## ABSTRACT

Small-molecules can self-assemble in aqueous solution into a wide range of nano-entity types and sizes (dimers, n-mers, micelles, colloids, etc), each having their own unique properties. This has important consequences in the context of drug discovery including issues related to non-specific binding, off-target effects, and false-positives and -negatives. Here, we demonstrate the use of the T2-CPMG relaxation NMR experiment which is sensitive to molecular tumbling rates and can expose larger aggregate species that have slower rotational correlations. The strategy easily distinguishes lone tumbling molecules versus nano-entities of various sizes. The technique is highly sensitive to chemical exchange between single-molecule and aggregate states, and can therefore be used as a reporter when direct measurement of aggregates is not possible by NMR. Interestingly, we found differences in solution behavior for compounds within structurally related series', demonstrating structure–nano-entity-relationships. This practical experiment is a valuable tool to support drug discovery efforts.

## INTRODUCTION

A clear understanding of the behavior of compounds in aqueous solution is central to the rational design of pharmaceutical agents. However, this basic insight has historically been shaded by assumptions that neglect the full complexity of multi-phase equilibria adopted by compounds. An alternate view of compound solution behavior that is encompassing yet emphasizes pharmacologically relevant properties is needed: a general multi-phase equilibrium, with an intermediate nano-entity phase that includes a wide array of self-associated assemblies (including smaller aggregates and n-mers) (Figure 1). Within this context, each compound would possess its own fingerprint equilibrium between single molecules, an array of soluble and/or colloidal self-associated nano-entities, and insoluble precipitate; with distributions that are highly dependent on environmental conditions such as compound concentration, buffer, salt, pH, temperature, metals, etc. (*vide infra*).<sup>1-3</sup>



**Figure 1.** Compounds can adopt multi-phase equilibria when placed in aqueous media. A-D are  $^1\text{H}$  NMR spectra of compounds that are typical of the states. (A) Light green SF yellowish, (B) Evans blue, (C) acid violet 49, and (D) pranlukast. All NMR spectra were taken at nominal concentrations of 300  $\mu\text{M}$ , in 50 mM sodium phosphate, 100 mM NaCl, 10%  $\text{D}_2\text{O}$ , pH 7.4.

Given the impact of compound solution behavior in drug discovery and pharmaceutical research, focused efforts have been undertaken to more fully characterize compound solution behavior. Such initiatives are beginning to decipher some features of the intermediate self-assembly phase and several reports have demonstrated that some small-molecules can self-associate into large micelle-like entities with unexpected solution properties.<sup>4-9</sup> For example, they can bind to and adsorb protein macromolecules in a non-specific manner, leading to changes in dynamics or partial denaturation of the protein. This can lead to false-positive results in drug screens.<sup>10-12</sup> They have also been implicated in causing up to 85-95% of artifacts in early high-throughput screens (HTS),<sup>13,14</sup> and in promiscuity tendencies in *in vitro*, off-target pharmacology assays – and perhaps toxicity.<sup>5</sup>

It is important to make a distinction between nano-entities and colloidal aggregates given that the latter is often used to describe particles of sizes extending beyond the nano-scale, thereby leading to confusion when describing pharmacological properties of compounds that often arise from their behavior at the nano-scale. For example, the formation of nano-entities can also hinder cell membrane permeation leading to false-negative results in cell-based assays.<sup>15</sup> On the other hand,



some drug self-assemblies have also been associated with favorable properties such as conferring unusually high drug bioavailability.<sup>1,16</sup> Thus, it becomes clear that we are only beginning to understand the important consequences of drug nano-entities. Appropriate technologies and strategies need to be employed to comprehensively detect the array of nano-entities that exist, and only then will we be able better understand the relationships between compound solution behavior and pharmacological properties.

To date, the available technologies and strategies have been inadequate to detect the full range of self-assemblies that can exist. The most widely used methods are dynamic light scattering (DLS) and electron microscopy (EM).<sup>10,17</sup> Both have been employed for the detection of larger aggregates, and DLS is often used for determining the critical aggregation concentrations (CAC) of some compounds. However, given the nature of these techniques, both are biased toward and limited to the observation of assemblies on the order of tens of nanometers.<sup>18,19</sup> They consequently lack the capacity to reliably detect and characterize medium and small nano-entities along with mixtures of self-assemblies. Another distinct approach is aggregation-induced emission (AIE).<sup>20</sup> This method relies on the induction of chromophore fluorescence and emission upon formation of aggregates. However, this has limited applicability as it is only relevant for large, highly aromatic compounds capable of AIE. More recently, we introduced an NMR aggregation assay based on serial dilutions as a facile method to support medicinal chemistry workflows and to predict promiscuity in off-target pharmacology assays.<sup>4,5</sup>

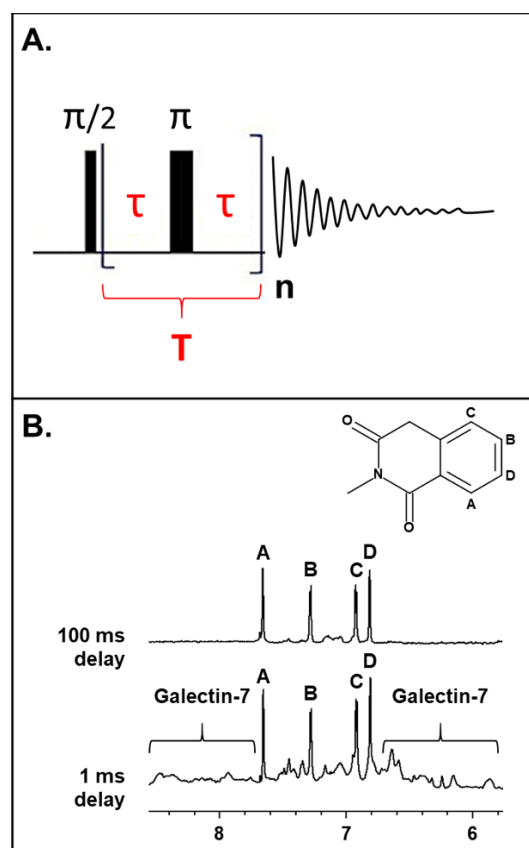
This assay is based on the notion that aggregates should be sensitive to changes in concentration from which unusual NMR spectral features can be observed. Although this technique is ideal for observing small to medium-sized nano-entities, the larger nano-entities are only visible in NMR spectra after the addition of detergents to break up the aggregates into smaller detectable species. While highly effective, the method remains burdened by the necessity of preparing multiple samples followed by longer data acquisition times, especially at the lower concentrations. In conclusion, the techniques described above have powerful utility but also have their inherent limitations. As a result, there is a high likelihood that many types of nano-entities remain undetected, and new strategies are wanting.

In this work, we introduce the use of an alternate NMR strategy that employs one-dimensional CPMG experiments (T2-CPMG). It exploits the difference in NMR relaxation properties that result from differences in sizes and tumbling rates of lone molecules versus those of self-associated nano-entities. This approach has several advantages such as the need for only a single sample and is amenable to higher throughput, with shorter data acquisition times than previously

described methods. We examine potential applications of this tool, including curating compound libraries, evaluating buffer effects, and exploring structure-nano-entity relationships (SNER).

## RESULTS AND DISCUSSION

**Designing the NMR aggregation assay based on T2-CPMG relaxation.** The spin-spin relaxation Carr-Purcell-Meiboom-Gill (T2-CPMG) experiment has long been a powerful tool for addressing a wide range of questions from monitoring protein internal dynamics to detecting small ligands binding to large proteins.<sup>21-28</sup> The experiment is relatively simple in that one applies consecutive spin-echoes ( $\tau$ - $\pi$ - $\tau$ ) after the initial excitation pulse, followed by the detection of the free-induction decay (FID) as shown in Figure 2A.

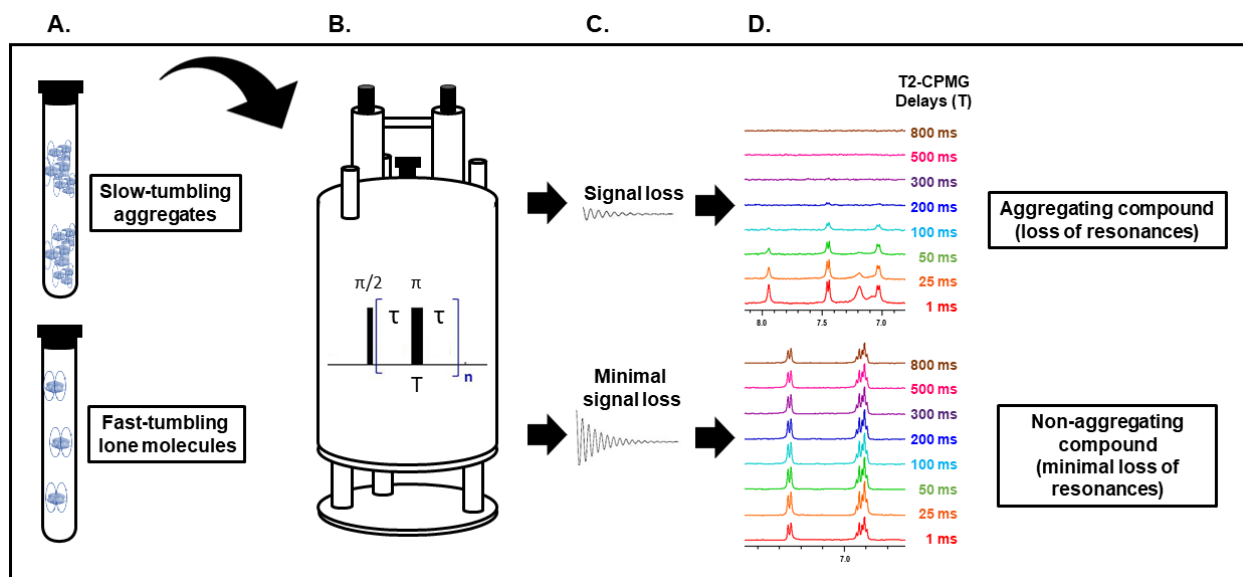


**Figure 2.** (A) In the T2-CPMG spin-echo pulse sequence, the overall length ( $T$ ) of the sequence can be varied by changing the delay time ( $\tau$ ), highlighted in red, and the number of refocusing pulses ( $n$ ). These experiments employ the spin-echo scheme which consist of a delay period ( $\tau$ ), a  $\pi$  or  $180^\circ$  radiofrequency pulse applied along the  $\pm x$  or  $\pm y$  axis, followed by a second delay period of same duration. (B) Example of the early use of the T2-CPMG NMR experiment. Both spectra are of a sample containing a mixture comprising of 100  $\mu\text{M}$  galectin-7 protein and 300  $\mu\text{M}$  of 2-methyl-1,3(2H, 4H)-isoquinolinedione small molecule in 50 mM sodium phosphate buffer at

pH 7 with 10% D<sub>2</sub>O. The bottom spectrum shows the results after applying a T2 filter with a 1 ms delay, displaying numerous broad resonances from the protein, whereas the top spectrum shows the result of applying a T2 filter with a 100 ms delay. After the 100 ms delay, most resonances from the larger species are eliminated.

The total duration (T) of the CPMG experiment can be defined by varying the number of refocusing pulses (n) and refocusing frequency, which is controlled by time  $\tau$ . One of the earliest applications was for the analysis of mixtures where one desired to observe the resonances of small molecules when in the presence of large macromolecules such as proteins.<sup>25</sup> A clear demonstration is shown in Figure 2B, where one can selectively observe the resonances of a fast tumbling small-molecule (which is a non-binder) while intentionally suppressing the undesired and slower-tumbling macromolecule, galectin-7. The bottom spectrum shows a typical <sup>1</sup>H NMR spectrum (minimal spin-echo delay of 1 millisecond) where the ligand and protein resonances are both observed. When a spin-echo delay of 100 milliseconds (ms) is introduced in the T2-CPMG experiment (top spectrum), the resonances of the larger macromolecule are effectively suppressed due to its inherent shorter transverse relaxation time (T<sub>2</sub>), whereas the resonances of the fast-tumbling small molecule are observed due to the longer T<sub>2</sub>.<sup>23</sup>

Given the above, it was rationalized that the T2-CPMG experiment could also be used for the detection of small-molecule aggregates. It is expected that non-aggregating, fast-tumbling lone molecules (with short correlation times in solution) should exhibit longer transverse relaxation (T<sub>2</sub>) and therefore should maintain signal intensity across a range of spin-echo delay times. On the other hand, aggregating molecules which tumble more slowly (longer correlation times in solution) should exhibit shorter T<sub>2</sub> due to loss of coherence and exhibit rapid resonance intensity loss across a range of delay times. This reasoning was then used to implement an initial version of our NMR T2-CPMG aggregation assay which includes the preparation of one NMR tube per compound followed by acquiring a series of eight one-dimensional T2-CPMG experiments, each having an increasing number of spin-echo pulse trains with overall CPMG times (T) ranging from 1-800 ms. This assay is illustrated in Figure 3.

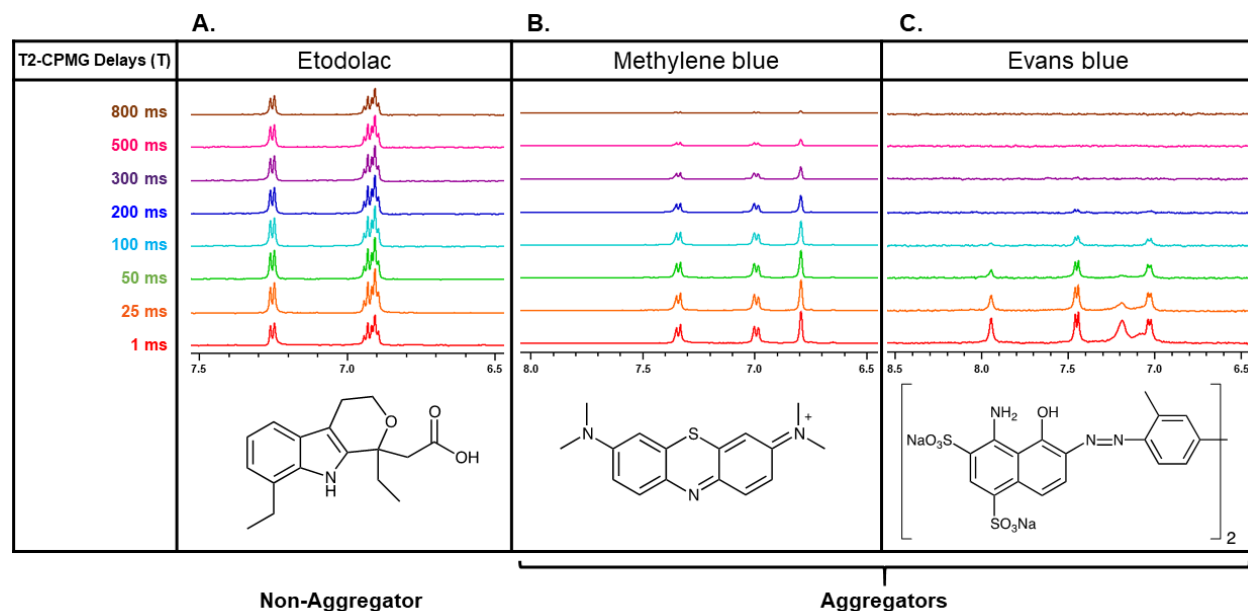


**Figure 3.** Depiction of the T2-CPMG aggregation assay. (A) Shown are representations of an aggregating compound in an NMR tube (top) and a freely soluble lone-tumbling compound (bottom). (B)  $^1\text{H}$ -NMR data are acquired with varying T2-CPMG refocusing delay times. Multiple experiments are run on a single sample. (C) The signal obtained from the NMR experiment will show significant decay in the case of slow-tumbling aggregates (top) but will be retained for a soluble, fast-tumbling molecule. (D) NMR data acquired over a range of T2-CPMG delay times. In the case of aggregation, there is a loss in resonance intensity (top) with longer delay time and in the case of lone (free) molecules, resonance intensity is maintained (bottom) across increasing delay time.

For non-aggregating molecules that tumble fast in solution (bottom, Figure 3), there is a minimal loss of resonance intensities as a function of longer T. However, for aggregating molecules that tumble slowly in solution, there is a loss of resonance signal intensity as a function of longer T (top, Figure 3). The distinct features are evident when comparing the top and bottom set of spectral overlays in Figure 3D, and thus can be used to reveal compounds that aggregate in solution.

**Application of the NMR T2-CPMG aggregation assay.** The T2-CPMG assay was applied to compounds that were reported in the literature to be known aggregators and non-aggregators (Figure 4).<sup>4,10,30-32</sup> The FDA-approved anti-inflammatory drug etodolac was used here as a non-aggregator control as its behavior in aqueous solution has been confirmed via both DLS and the NMR dilution aggregation assay reported previously.<sup>4</sup> The T2-CPMG spectra of etodolac demonstrates long T2-CPMG relaxation times for the compound in solution, indicative of a non-aggregator (Figure 4A). Taking the difference in peak area between the spectra with 1 ms and

800 ms delay times as a percentage, 60% of signal intensity is maintained for the aromatic resonances of etodolac across the range of delay times.



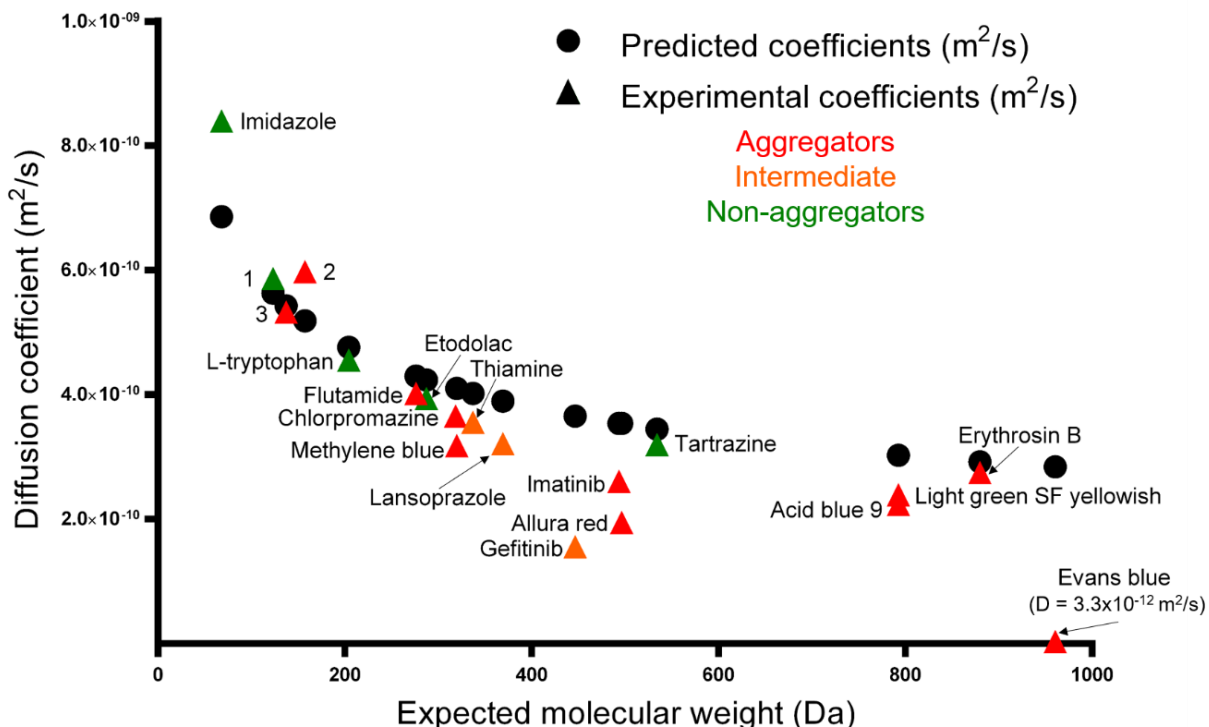
**Figure 4.** Shown are three examples of the T2-CPMG assay applied to literature compounds at 300  $\mu$ M. Panels A-C show a series the  $^1$ H-NMR spectra and the T2-CPMG delay times are shown on the left.

In contrast, the relaxation properties of known aggregator compounds are quite different. T2-CPMG spectra were acquired on solutions of methylene blue and Evans blue (Figure 4B and 4C), and unlike etodolac, resonance intensities rapidly decay with increasing spin-echo delay times. Taking the difference in peak area between the spectra with 1 ms and 800 ms delay times as a percentage, there is < 1% of the initial signal intensity remaining at 800 ms for the aromatic protons of both methylene blue and Evans blue. These same samples were also retested in an NMR dilution aggregation assay which again report that the compounds are aggregators. Data are shown In the Supporting Information.

Thus far in this report, it has been implied that the spin-spin relaxation measured in the T2-CPMG experiments directly reflect the tumbling rate of the species giving rise to the observable resonances. However, it must be kept in mind that signal intensity losses as a function of delay times can also arise from exchange phenomena. That is, a fast-tumbling molecule can also exhibit loss of resonance intensities at longer delay times due to exchange with an aggregate species.

The exchange contribution to the observed T2 for the peak of interest will be dependent on the Exchange rates between the free and self-associated states. Therefore, compounds exhibiting exchange between populations of free and aggregate states will exhibit broader resonances and a more rapid signal decay in the T2-CPMG experiment. It is because of this that the T2-CPMG experiment can reveal the presence of aggregation processes that are otherwise invisible by other methods (either because the aggregate is too small, too large, or because of weakly populated states of the aggregate in the samples). The magnitude of the effect scales according to the exchange rates and populations of free compound vs aggregate states (possibly exchange between multiple nano-entities). This technique is therefore sensitive to various conditions under which aggregation could give rise to false positives in binding studies or screening. It is also sensitive to aggregates of sizes that would normally be outside the range of detection by conventional NMR methods. For example, the aggregation behaviors of methylene blue and Evans blue have been characterized by DLS, and have been shown to form aggregates close to a hundred nanometers in diameter.<sup>10,32</sup> Large aggregates of this size would not be expected to yield sharp resonances by solution-state NMR. Given this, the NMR resonances observed here most likely arise from a fast-tumbling species and the loss of resonance intensities as a function of T2-CPMG delays reveals that the fast-tumbling species is in fast exchange with an “NMR invisible” larger nano-entity. Thus, the loss in signal intensity indirectly reveals the existence of the formation of large nano-entities.

To potentially corroborate these T2-CPMG findings, we resorted to diffusion-ordered spectroscopy (DOSY) NMR which measures translational diffusion and may be used to distinguish between lone tumbling molecules versus larger aggregate assemblies.<sup>4,33-36</sup> Diffusion coefficients ( $D$ ,  $m^2/s$ ) were measured for a series of compounds as shown in Figure 5 and plotted versus the molecular weights of the primary structures. Interestingly, the diffusion coefficients based on the primary structures (black circles) predict faster diffusion rates than the measured diffusion coefficients for several compounds deemed aggregators and intermediate aggregators (red and orange triangles) via the T2-CPMG method. On the other hand, there is agreement between predicted and measured diffusion constants for non-aggregator compounds (green triangles). It is noteworthy that all compounds with much slower diffusion coefficients than the predicted values were flagged by the T2-CPMG method. It must however be kept in mind that other species may be missed due to limitations of one or both methods (e.g, slow exchange between the fast-tumbling molecules and the larger NMR-invisible assemblies).



**Figure 5.** Shown are predicted/experimentally derived diffusion coefficients ( $D$ ) versus molecular weights of the primary structures for known aggregators, non-aggregators, and compounds with intermediate aggregation behavior. Points denoted as colored triangles are for experimentally obtained diffusion coefficients, whereas black circles represent predicted values. Experimental points were colored according to their loss in T2-CPMG signal intensity between 1 and 800 ms delays (red, more than 75% loss in signal intensity; orange, between 50% and 75%; green, less than 50%).

Given the above and taking into consideration the general multi-phase model in Figure 1, a summary of the method's blind-spots and advantages is important for clarity sake. If a compound is placed in aqueous media and no resonances are observed, then the compound can simply be insoluble. One would expect to observe precipitate in the NMR tube. However, the sample could also contain large micelles or colloids whose NMR resonances would be too broad to clearly observe in a  $^1\text{H}$  NMR spectrum (see Figure 1C). In this case, it is suggested that one centrifuge the sample and remove the supernatant for further analyses. DLS or other techniques could then be appropriate for revealing the presence of aggregates. Another option would be to add a detergent that could break up the aggregate after which a  $^1\text{H}$  NMR spectrum would be observable.<sup>4,5,31</sup> On the other hand, if sharp NMR resonances are observed when a compound is placed in aqueous media, then several scenarios can exist. One scenario is that the sample

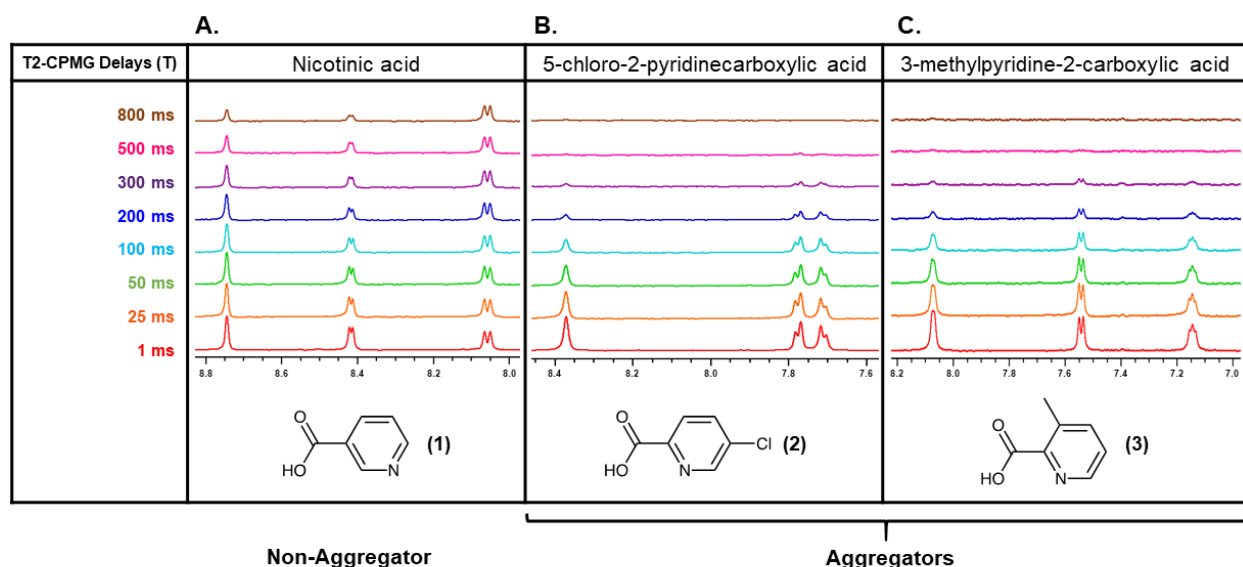
contains only fast-tumbling lone molecules. If so, then the  $^1\text{H}$  NMR spectrum would exhibit sharp resonances as shown in Figure 1A and there would be minimal loss in signal in the T2-CPMG experiment as demonstrated in Figure 3 (bottom). If the sample contains small aggregates, then the  $^1\text{H}$  NMR spectrum would appear as they do in Figure 1A or 1B, whereas the T2-CPMG data would result in loss of intensities as shown in Figure 4C.

It is also possible that a sample may contain an equilibrium mixture of free-tumbling lone-molecules, small self-assemblies and large micelle-like/colloidal entities. The  $^1\text{H}$  NMR spectrum would appear similar to that shown in Figure 1A, 1B, or 1C, and there would be losses in intensity in the T2-CPMG data which could either directly report the existence of small self-assemblies or indirectly report the fast-exchange equilibrium with a large entity. The latter can be exposed as described above either by DLS and/or by the addition of detergent followed by the observation of resonances in a  $^1\text{H}$  NMR spectrum whereas the former would be difficult to detect by DLS.

**Compound library screening and revealing structure-nano-entity-relationships.** High-throughput and virtual screens can readily screen millions of compounds from which subsets of compounds can be subjected to counter screens, etc. Unfortunately, chemical libraries are known to be contaminated by compounds that have promiscuous properties due to compound self-aggregation. NMR is exceptionally suited as a final counter screen, for fragment screening and for library curation. In addition to protein binding confirmations, NMR can also readily capture critical free-state information regarding compound solubility, aggregation, and primary structure. We therefore applied  $^1\text{H}$  NMR and our T2-CPMG strategy on over 5000 small molecules in aqueous media to better understand their solution free-state behavior and for curation purposes to optimize fragment screening campaigns.

Analyses of the  $^1\text{H}$  NMR data served as a practical method for identifying some misbehaving compounds. Compounds were removed if no NMR resonances were observed or if low solubility was exhibited (i.e.  $< 100\ \mu\text{M}$  given a nominal concentration of  $300\ \mu\text{M}$ ) as determined by employing the ERETIC (Electronic Reference To access In vivo Concentrations) method.<sup>37</sup> Of note, many of the removed compounds had reasonable clogP and topological polar surface area (TPSA) and would therefore have been expected to exhibit adequate solubility. This highlights the Importance of curating under conditions as close as possible to those in which the screening or binding studies will be conducted (a point further demonstrated in another section below).



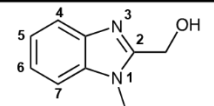
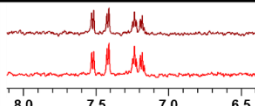
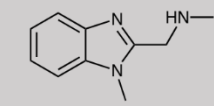
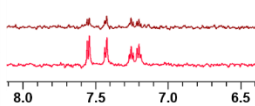
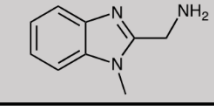
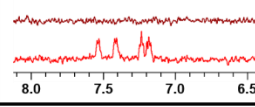
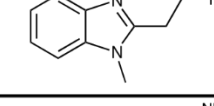
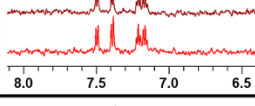
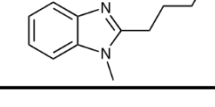
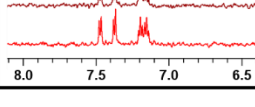


**Figure 6.** Detection of aggregates of fragment-like molecules. Panels A-C show stacked T2-CPMG spectra for **1**, **2** and **3**.

Analyses of the T2-CPMG NMR data exposed some interesting features. Figure 6 illustrates the T2-CPMG spectra for three structurally related fragment-like molecules. Nicotinic acid (**1**) (Figure 6A) maintains more than 50% of its signal intensity after 800 ms whereas 5-chloro-2-pyridinecarboxylic acid (**2**) and 3-methylpyridine-2-carboxylic acid (**3**) (Figure 6B & 6C) retain almost no signal intensity at the same delay. Interestingly, the T2-CPMG data clearly shows that **1** assumes the behavior of fast-tumbling molecules, whereas **2** and **3** show evidence of slower-tumbling molecules in solution. This observation confirms that minor differences in primary structure can result in very significant differences in self-association behavior. Further investigation into these three compounds reveals that no unusual trends (e.g. resonance shifting, broadening, new peaks, etc.) were noted in the NMR dilution aggregation assay (see Supporting Information). Characterization by dynamic light scattering (DLS) was also attempted but results were inconclusive, possibly due to sample polydispersity, absorption, small aggregates sizes or nano-entity concentrations below the limit of detection of the instrument.<sup>38-41</sup> Thus, it is evident that the T2-CPMG experiment can be more sensitive than more traditional approaches for monitoring aggregation.

An important advantage of the T2-CPMG strategy for detecting aggregators is that it requires the preparation of only a single sample for data collection. Furthermore, one could also collect fewer T2-CPMG spectra per sample making the technique more amenable for even higher throughput

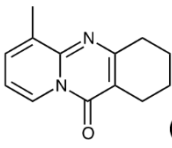
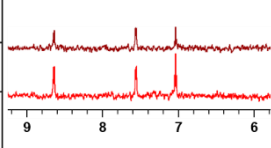
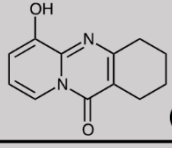
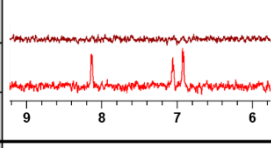
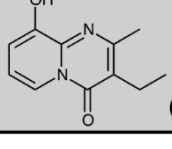
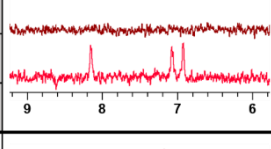
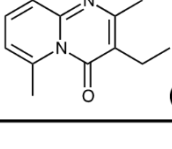
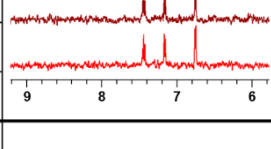
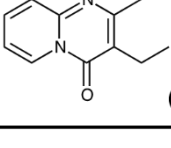
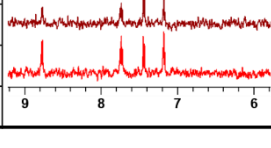
screening. For example, with our current setup each delay time requires approximately 30 seconds. Therefore, complete analysis of a 1500 compounds library would require ~100 hours of NMR time using 8 delays. In comparison, reducing the amount of T2-CPMG delays to two (e.g. 1 and 800 ms) would drop the experimental time to ~25 hours, while still obtaining pertinent relaxation information (see Supporting information). Figure 7 illustrates this approach by calculating the percentage difference in peak area of resonances derived from spectra collected at 800 versus 1 milliseconds delay (Figure 7, 4<sup>th</sup> column). Accordingly, a peak area value of 0 % represents a complete loss of signal for spectra at the 800 ms delay. Compounds can be classified as non-aggregators (e.g. values between 50 and 100 %), aggregators (e.g. peak area values between 0 and 25%), and intermediate behavior (e.g. between 25 and 50%). See column 1 in Figure 7. The percentages in column 4 or the categories in column 1 may be used during library curation to set thresholds for deciding whether or not a compound is kept, discarded or needs to be closely monitored in screens. This type of data could be taken into consideration with other parameters typically used for library curation.<sup>3,6,9,42-46</sup>

Behavior	Structure	Peak Chemical Shift (6-10 ppm)	Peak Area 800 ms/1 ms T2-CPMG	NMR (aromatic region)
Non-Aggregator	 (4)	7.15 – 7.25	80%	
		7.41 – 7.55	> 99%	
Intermediate Behavior	 (5)	7.15 – 7.25	36%	
		7.41 – 7.55	34%	
Aggregator	 (6)	7.15 – 7.25	0%	
		7.41 – 7.55	0%	
Non-Aggregator	 (7)	7.15 – 7.25	73%	
		7.37 – 7.51	79%	
Non-Aggregator	 (8)	7.15 – 7.25	74%	
		7.37 - 7.48	77%	

**Figure 7.** Series of structurally related N-methyl benzimidazole core compounds. For each compound, peaks in the aromatic region of the spectra (6-10 ppm) are identified and the peak area was calculated by integration. The peak areas between 1 ms and 800 ms delay times were compared and the percentage difference was calculated. These percentages can be compared to the NMR spectra in the far right column. The bottom, red spectrum has a 1 ms delay time and the top, brown spectrum has an 800 ms delay time.

Analyses of the T2-CPMG NMR data for larger sets of compounds also exposed some interesting features such as structure-nano-entity relationships (SNER). This can be illustrated by the sets of compounds in Figures 7 and 8. Note in Figure 7 that minor changes to the substituent at position 2 of the benzimidazole core results in differences in T2-CPMG trends. It is noteworthy that a primary amine (**6**) promotes aggregation, a secondary methyl amine (**5**) results in intermediate aggregation tendencies, whereas an alcohol (**4**) results in longer-lived signal intensities. Furthermore, changes to the alkyl chain length also has a dramatic impact on the aggregate behavior (**6**, **7** & **8**). Figure 8 displays the data found for compounds in a pyrido-pyrimidinone series. Again, structurally-related compounds exhibit distinct T2-CPMG values indicating that they have different behaviors in solution. The Supporting Information shows additional interesting trends for another benzimidazole series.

It is noteworthy that the observed SNERs are surprising and not yet understood. More specifically, it is unclear which substituents or cores induce aggregation. Although highly hydrophobic compounds (high clogP) likely have a higher propensity to self-assemble, the fine underlying mechanistic details have yet to be uncovered.<sup>5</sup> Attempts to establish predictive quantitative nano-entity relationships (QSNER) have failed to date. Perhaps a larger ensemble of data from T2-CPMG library screens would be needed for further evaluating QSNERs.

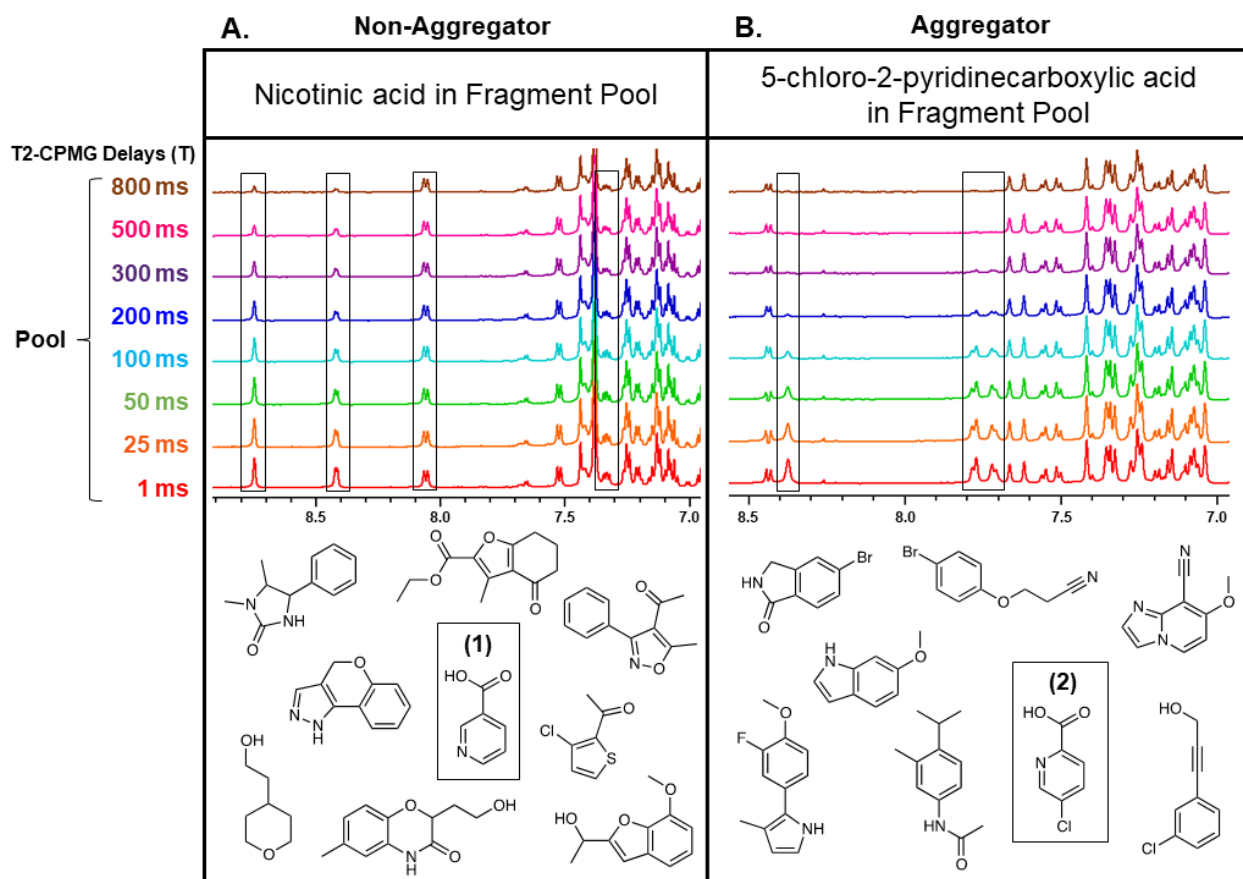
Behavior	Structure	Peak Chemical Shift (6-10 ppm)	Peak Area 800 ms/1 ms T2-CPMG	NMR (aromatic region)
Non-Aggregator		7.03	60%	
		7.55	96%	
		8.63	77%	
Aggregator		6.92	9%	
		7.05	15%	
		8.12	13%	
Aggregator		6.92	0%	
		7.06	12%	
		8.14	10%	
Non-Aggregator		6.75	70%	
		7.16	90%	
		7.44	> 99%	
Non-Aggregator		7.17	66%	
		7.44	> 99%	
		7.73	61%	
		8.77	64%	

**Figure 8.** Series of structurally related pyrido-pyrimidinone core compounds. For each compound, peaks in the aromatic region of the spectra (6-10 ppm) are identified and the peak area was calculated by integration. The peak areas between 1 ms and 800 ms delay times were compared and the percentage difference was calculated. These percentages can be compared to the NMR spectra in the far right column. The bottom, red spectrum has a 1 ms delay time and the top, brown spectrum has an 800 ms delay time.

In summary, it is clear that the T2-CPMG experiment can provide rich and detailed information on the behavior of compounds in aqueous media. Interestingly, there appears to be SNER that potentially affect multiple properties of compounds (e.g. promiscuity, affinity, etc.) which will become clearer as this T2-CPMG strategy gains more wide-spread use by medicinal chemists. An important take-home message from the above data is clear: Medicinal chemists may find more clarity in interpreting structure-activity-relationships (SAR) given complementary T2-CPMG data.

**Monitoring solution behavior of compounds in mixtures.** A frequent practice in NMR drug screening is the pooling of compounds for screening against a target.<sup>22,47-52</sup> This technique allows for more efficient screening of a larger chemical library by testing for binding of several compounds to a target within a single NMR sample. This reduces the cost related to materials, sample preparation and data acquisition time. This is also highly pertinent in the context of

combinatorial libraries<sup>53,54</sup> and natural products<sup>55,56</sup> where mixtures of molecules are used against targets, followed by isolation or deconvolution of the active molecules. To evaluate if the T2-CPMG is able to identify aggregating molecules in mixtures of compounds, **1** and **2** were assessed amongst pools of small molecules as depicted in Figure 9. The maintenance of the resonance intensities for the observed spectra of **1** in the mixture (Figure 9A) is similar and consistent as that observed when **1** was a singleton in solution and behaved as a non-aggregating compound (Figure 6A). Also, the resonances of compound **2** diminish as a function of delay time as a singleton (Figure 6B) and within a mixture (Figure 9) demonstrating that the T2-CPMG experiment clearly reports that **2** behaves as an aggregator in both conditions.



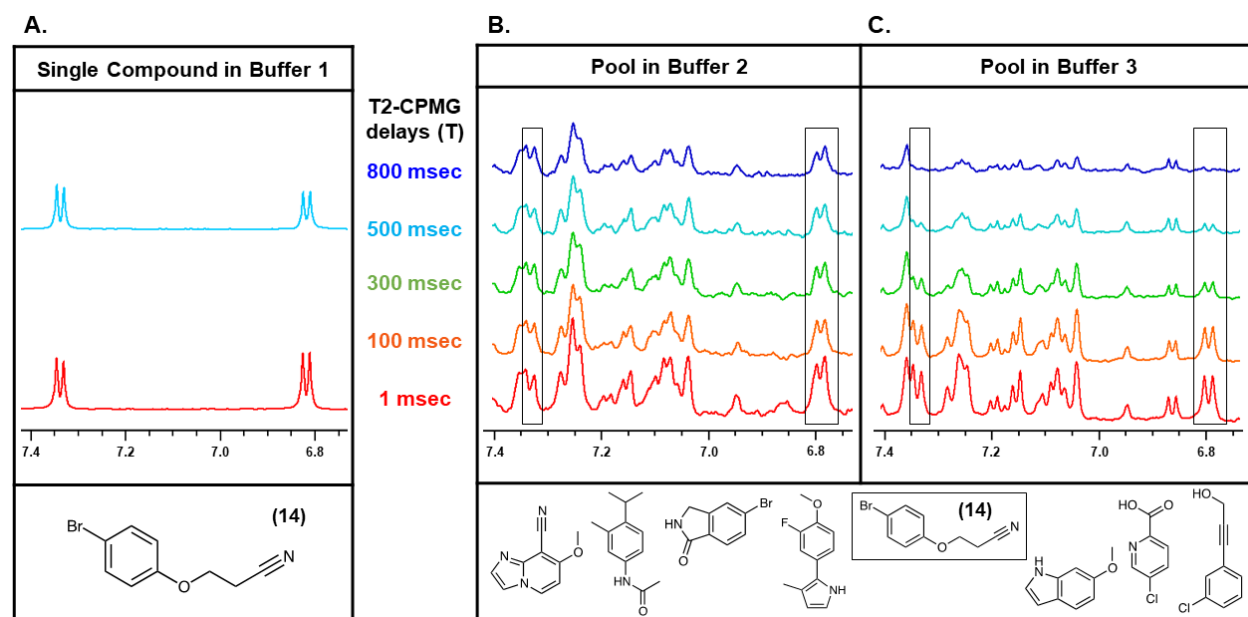
**Figure 9.** Detection of aggregates amongst pools of compounds. T2-CPMG spectra overlays of (A) **1** and (B) **2**. Both profiles are consistent with their isolated form in Figure 6.

The data in Figure 9B also shows that the aggregation behavior of **2** does not adversely affect the solution behavior of the remaining compounds in the mixture. In other examples (e.g. Figure 10), we observed that some aggregating compounds induced aggregation of other compounds in the mixture or the other compounds were adsorbed into the aggregate – we refer to this as the

“bad apple effect”. These types of T2-CPMG observations are valuable and always used for our NMR fragment screening campaigns. Pools would be eliminated from further consideration if a “bad apple effect” was noted. Also, fragment compounds would only be prioritized if no aggregation evidence was observed.

**T2-CPMG can monitor aggregation under various buffer conditions.** It is well known that buffer conditions can seriously affect assay results – this being mostly attributed to target protein behavior. However, we have seen over many screening campaigns that compound self-aggregation is also highly influenced by buffer conditions. Therefore, one must keep in consideration the buffer conditions and pH that are optimal for biochemical assays and how these might affect compound behavior in solution.<sup>9,57</sup>

Fortunately, the T2-CPMG can be employed on a variety of buffer conditions of choice allowing one to match, as close as possible, the NMR fragment screening with biochemical assay conditions. An example of the distinct behavior of compounds in various buffer conditions is shown in Figure 10.



**Figure 10.** Buffer dependency of aggregation. T2-CPMG NMR spectra of compound 3-(4-bromophenoxy)propanenitrile (**14**) was taken in various conditions. (A) On its own for the purpose of library curation in Buffer 1. (B) In Buffer 2, **14** retains its non-aggregating profile. (C) Conversely, in Buffer 3, **14** exhibits a significantly different behavior. Buffer 1: 50 mM sodium phosphate, 100 mM NaCl, 10% D<sub>2</sub>O, pH 7.4; Buffer 2: 10 mM sodium phosphate, 1.8 mM potassium phosphate, 137 mM NaCl, 2.7 mM KCl, 10 % D<sub>2</sub>O, pH 7.4; Buffer 3: 50 mM HEPES, 50 mM NaCl, 2 mM MgCl<sub>2</sub>, 2 mM TCEP, 0.02% NaN<sub>3</sub>, 10% D<sub>2</sub>O, pH 7.4.

Note that the T2-CPMG resonance intensities of singleton compound 3-(4-bromophenoxy)propanenitrile (**14**) exhibit little change in buffer 1 from 1 ms (bottom of Figure 10A) as compared to the 500 ms (top Figure 10A) used for library curation. Thus, compound **14**, as a singleton, does not aggregate in buffer 1. Similarly, the T2-CPMG changes are minor in buffer 2 as shown in Figure 10B. However, changes are dramatically different for buffer 3 where the resonances dramatically diminish at the longer delay times (Figure 10C). Thus, illustrating the important effects of buffers on compound solution behavior. Another feature of interest is that many of the compounds in the mixture in buffer 3 (Figure 10C) also appear to be aggregating as their resonance intensities also are reduced as a function of delay time. This is a clear example of the “bad apple effect” mentioned above. Note that it has been reported that multiple aggregators can form co-aggregates, even when each compound are below their respective CAC.<sup>58-60</sup> Therefore, it is conceivable that some compounds may exert a detrimental effect on otherwise well-behaved compounds.

## CONCLUSION

**Advantages and limitations.** This method is exceptionally suited for the elucidation and study of a wide variety of nano-entity types and sizes – especially small to medium-sized aggregates, but also large aggregates indirectly via exchange phenomena. One important advantage of this method is that experiments can be run under a large variety of conditions, which includes a wide range of compound concentrations, pH, buffer components and temperatures. Running the T2-CPMG experiment at a range of concentrations will allow one to explore critical aggregate concentration (CAC) of some types of nano-entities, if the CAC is above the limit of detection for acquiring a 1D NMR spectrum. The limit of detection can be as low as single-digit micromolar compound concentration, whereas from a practical viewpoint for screening purposes samples made to double- or triple-digit micromolar range are recommended. Nonetheless, CAC determination can be a significant limitation for the methodology presented here. If possible within a project workflow, it is recommended to combine the T2-CPMG assay and serial dilution NMR assay to be able to extract the maximal amount of information about nano-entities.

It should also be noted that deuterated versions of the buffer may be required depending on resonance overlap with the compound signal of interest, but in many cases, even non-deuterated buffers can be used at lower concentration. Inorganic buffers are ideally suited (phosphate, borate) as they exhibit no proton resonances in aqueous solution.

A considerable advantage of this method over the previously reported NMR dilution aggregation assay is that it is much more amenable to higher throughput screening. Only a single sample is required and acquisition time is relatively short. Curating and screening fragment compound libraries become more feasible and can ultimately save crucial resources further down the drug discovery pipeline.

Despite the significant advancements this method offers, there are several shortcomings that must be acknowledged. Given that this method monitors the relaxation properties of compounds of interest, any process which leads to enhanced chemical exchange can result in faster resonance decay. Tautomerization, exchangeable protons, and other intramolecular conformational exchange processes can lead to enhanced relaxation, and thereby to a false aggregation flag. In practice, however, most compounds (especially fragments) can be screened reliably using this method even when rotamers are known to exist. In such cases, comparing the T2-CPMG decay across multiple protons in the structure can give insight into whether rotamers exists and can therefore aid eliminate false aggregation flags. It should also be noted that there may also be limitations with the characterization of paramagnetic compounds, due to their short transverse relaxation times.<sup>61,62</sup>

**Perspective of the aggregation phenomena.** From a drug discovery perspective, it is critical to have an understanding of small molecule aggregation and behavior of nano-entities in solution. Given that aggregates are one of the primary sources of off-target pharmacological effects and can give rise to false positives (via aggregation-induced mechanisms) in screening and binding studies, it is clear that having a powerful tool such as the T2-CPMG method in hand, along with other complementary technologies, will aid in ensuring productive drug-discovery initiatives. A better understanding of SNER along with SAR can certainly help.<sup>63</sup> One also needs to consider at what stage the phenomenon of aggregation is being observed. For example, it is prudent to avoid hits ( $\mu\text{M}$  affinity range) that aggregate from an HTS screen. Likewise, for fragment screens, one should avoid binders that aggregate given that binding is usually detected in the high  $\mu\text{M}$  and mM range of affinity – these are compound concentrations where aggregation and promiscuity are prevalent and give rise to artifacts.

During lead optimization campaigns the phenomenon of compound aggregation should be carefully considered. A potent compound can have both target-specific binding along with self-association attributes. On one hand, compounds that aggregate also have a tendency to be promiscuous in *in vitro*, off-target pharmacology screens. The NMR dilution aggregation assay



has previously been used as a key criterion to select non-aggregating compounds for preclinical studies and promotion for the clinic.<sup>5</sup>

On the other hand, other groups have employed the phenomenon of aggregation for desirable features. For example, aggregation can promote increased bioavailability<sup>1,64</sup> and some efforts are being devoted in order to harness the intrinsic properties of these entities to improve stability and delivery.<sup>2,16,65-69</sup> Some reports even describe inhibition of protein-protein interaction and amyloid polymerization by exploiting small-molecule aggregation.<sup>70,71</sup>

**Perspective of wider applicability.** The T2-CPMG NMR aggregation assay presented in this work can have wide-ranging utility in many chemical fields. It is a simple and practical experiment that provides a wealth of information on compound self-assemblies and interactions. It is also amenable to higher throughput screening given that only a single sample per compound or mixture is needed. We have employed this assay for high-throughput analysis on thousands of compounds across a wide range of solution conditions. Furthermore, to increase throughput, data subsets can be acquired to reduce screening time and simplify analyses. We anticipate that this assay will begin to expose the wide range of nano-entities that can exist and help keep drug discovery initiatives on target.

## EXPERIMENTAL SECTION

**Compounds and libraries.** All compounds investigated in this work were ordered from external vendors. The suppliers and catalog numbers are provided in the Supporting Information. All fragment compounds used in this work were identified through the curation of the ChemBridge ([https://www.chembridge.com/screening\\_libraries/fragment\\_library/](https://www.chembridge.com/screening_libraries/fragment_library/)) and NMX Research and Solutions Inc. (<https://www.nmxresearch.com/>) fragment libraries. A subset of the NMX library is available as Key Organics' BIONET library (<https://www.keyorganics.net/services/bionet-products/fragment-libraries/>).

**NMR sample preparation.** Each compound was prepared as a 10 mM stock solution in DMSO-d<sub>6</sub> from the purchased powder or oil. This solution was then diluted in the test buffer to give the desired final compound concentrations. The buffer used, unless stated otherwise, consisted of 50 mM NaPi, 100 mM NaCl, pH 7.4 with 10% D<sub>2</sub>O. Water-suppression using excitation sculpting with gradients was therefore necessary.<sup>72</sup> This allows the experiment to be more generally applicable to a range of buffer conditions and handles any water impurities. Samples were stored at 277 K and data was acquired at 298 K.

**T2-CPMG.** The experiments shown here were run on a 600 MHz Bruker Avance III Spectrometer equipped with a helium cryoprobe. The T2-CPMG experiment employed is a modified version of the standard Bruker 1D  $^1\text{H}$  experiment with excitation sculpting (zgesgp) with the addition of a CPMG pulse train after the initial 90-deg excitation pulse. The total duration of each spin echo was fixed to 1 msec ( $\tau=500 \mu\text{sec}$ ) whereas the number of echoes in the pulse train was varied according to the total time (T). The number of scans for all spectra was 64, although sufficient signal-to-noise could be achieved using only 4 scans at 300  $\mu\text{M}$  of compound. It should also be noted that in this work we report  $T2^*$  (effective T2) given the following:  $T2^* = T2$  (natural) +  $T2_{\text{inhom}}$  (field inhomogeneity) +  $T2_{\text{exch}}$  (exchange contribution).

**1D  $^1\text{H}$ .** The standard Bruker 1D  $^1\text{H}$  sequence with excitation sculpting (zgesgp) was employed. A relaxation delay of 10 sec was employed to ensure near complete relaxation of all aromatic spins prior to subsequent experiments. Spectra were acquired with 64 scans.

**Diffusion.** DOSY experiments were performed using the standard Bruker ledbpgp2s pulse sequence.<sup>73</sup> 128 scans were employed for each acquisition. To minimize water signal, the buffer used was prepared in 100 %  $\text{D}_2\text{O}$ , with the same components as mentioned above.

**Data interpretations.** Data visualization was done in Bruker's TopSpin software. Quantification of T2-CPMG peak area was calculated using Matlab (R2016b. The MathWorks, Inc. Natick, Massachusetts, United States) (Supporting Information). DOSY data analysis was conducted using Bruker's Dynamics Center. Prediction of diffusion coefficients was done using the SEGWE calculator with the Stokes-Einstein method (<https://nmr.chemistry.manchester.ac.uk/?q=node/432>).<sup>74,75</sup>

## AUTHOR INFORMATION

### Corresponding Authors

[\\*steven.laplante@iaf.inrs.ca](mailto:steven.laplante@iaf.inrs.ca)

[\\*sacha.larda@nmxresearch.com](mailto:sacha.larda@nmxresearch.com)

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

### Funding Sources

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2016-06747), Quebec Consortium for Drug Discovery (CQDM),

Mitacs, Canadian Foundation for Innovation (CFI) (35334) and by NMX Research and Solutions Inc. Y.A. was supported in part by a Graduate Student Scholarship from the Quebec Network for Research on Protein Function, Engineering, and Applications (PROTEO) (FT109664).

## ACKNOWLEDGMENT

The authors thank F. Bilodeau, M. Serrano-Wu, A. Lowerson and P. McCarren for their help in preparing the fragment library.

## ABBREVIATIONS

AIE, aggregation-induced emission; CAC, critical aggregation concentration; DLS, dynamic light scattering; FID, free-induction decay; HTS, high-throughput screening; ms, millisecond(s); QSNER, quantitative structure-nano-entity-relationship; SAR, structure-activity-relationship; SNER, structure-nano-entity-relationship; T, T<sub>2</sub>-CPMG duration.

## ASSOCIATED CONTENT

### Supporting Information.

The Supporting Information is available free of charge on the ACS Publications website at DOI: <https://doi.org/10.1021/acs.jmedchem.9b00653>.

Methods for the quantification of peak area and relaxation decay rates; NMR dilution assay of control compounds; additional evidence of the T<sub>2</sub>-CPMG sensitivity for some aggregates of small molecules; additional series of structurally related compounds with distinct T<sub>2</sub>-CPMG profiles; use of T<sub>2</sub>-CPMG at lower compounds concentrations; evidence for the detection of chemical exchange; consistency between relaxation T<sub>2</sub> and peak area using only two T<sub>2</sub>-CPMG delays; compound information.

## REFERENCES

- (1) Frenkel, Y. V.; Clark, A. D.; Das, K.; Wang, Y.-H.; Lewi, P. J.; Janssen, P. A. J.; Arnold, E., Concentration and pH Dependent Aggregation of Hydrophobic Drug Molecules and Relevance to Oral Bioavailability. *J. Med. Chem.* **2005**, *48* (6), 1974-1983.
- (2) Ma, W.; Cheetham, A. G.; Cui, H., Building Nanostructures with Drugs. *Nano Today* **2016**, *11* (1), 13-30.

- (3) Rao, H.; Li, Z.; Li, X.; Ma, X.; Ung, C.; Li, H.; Liu, X.; Chen, Y., Identification of Small Molecule Aggregators From Large Compound Libraries by Support Vector Machines. *J. Comput. Chem.* **2010**, *31* (4), 752-763.
- (4) LaPlante, S. R.; Carson, R.; Gillard, J.; Aubry, N.; Coulombe, R.; Bordeleau, S.; Bonneau, P.; Little, M.; O'Meara, J.; Beaulieu, P. L., Compound Aggregation in Drug Discovery: Implementing a Practical NMR Assay for Medicinal Chemists. *J. Med. Chem.* **2013**, *56* (12), 5142-5150.
- (5) LaPlante, S. R.; Aubry, N.; Bolger, G.; Bonneau, P.; Carson, R.; Coulombe, R.; Sturino, C.; Beaulieu, P. L., Monitoring Drug Self-Aggregation and Potential for Promiscuity in Off-Target In Vitro Pharmacology Screens by a Practical NMR Strategy. *J. Med. Chem.* **2013**, *56* (17), 7073-7083.
- (6) Seidler, J.; McGovern, S. L.; Doman, T. N.; Shoichet, B. K., Identification and Prediction of Promiscuous Aggregating Inhibitors among Known Drugs. *J. Med. Chem.* **2003**, *46* (21), 4477-4486.
- (7) Schreier, S.; Malheiros, S. V. P.; de Paula, E., Surface Active Drugs: Self-association and Interaction With Membranes and Surfactants. Physicochemical and Biological Aspects. *Biochim. Biophys. Acta* **2000**, *1508* (1), 210-234.
- (8) McGovern, S. L.; Helfand, B. T.; Feng, B.; Shoichet, B. K., A Specific Mechanism of Nonspecific Inhibition. *J. Med. Chem.* **2003**, *46* (20), 4265-4272.
- (9) McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K., A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *J. Med. Chem.* **2002**, *45* (8), 1712-1722.
- (10) Owen, S. C.; Doak, A. K.; Wassam, P.; Shoichet, M. S.; Shoichet, B. K., Colloidal Aggregation Affects the Efficacy of Anticancer Drugs in Cell Culture. *ACS Chem. Biol.* **2012**, *7* (8), 1429-1435.
- (11) Coan, K. E. D.; Maltby, D. A.; Burlingame, A. L.; Shoichet, B. K., Promiscuous Aggregate-Based Inhibitors Promote Enzyme Unfolding. *J. Med. Chem.* **2009**, *52* (7), 2067-2075.
- (12) Auld, D. S.; Inglese, J.; Dahlin, J. L., Assay Interference by Aggregation. In *Assay Guidance Manual*, Sittampalam, G. S.; Coussens, N. P.; Brimacombe, K.; Grossman, A.; Arkin, M.; Auld, D.; Austin, C.; Baell, J.; Bejcek, B.; Caaveiro, J. M. M.; Chung, T. D. Y.; Dahlin, J. L.; Devanaryan, V.; Foley, T. L.; Glicksman, M.; Hall, M. D.; Haas, J. V.; Inglese, J.; Iversen, P. W.; Kahl, S. D.; Kales, S. C.; Lal-Nag, M.; Li, Z.; McGee, J.; McManus, O.; Riss, T.; Trask, O. J., Jr.; Weidner, J. R.; Wildey, M. J.; Xia, M.; Xu, X., Eds. Bethesda MD, **2004**.

- (13) Duan, D.; Torosyan, H.; Elnatan, D.; McLaughlin, C. K.; Logie, J.; Shoichet, M. S.; Agard, D. A.; Shoichet, B. K., Internal Structure and Preferential Protein Binding of Colloidal Aggregates. *ACS Chem. Biol.* **2017**, *12* (1), 282-290.
- (14) Feng, B. Y.; Simeonov, A.; Jadhav, A.; Babaoglu, K.; Inglese, J.; Shoichet, B. K.; Austin, C. P., A High-throughput Screen for Aggregation-based Inhibition in a Large Compound Library. *J. Med. Chem.* **2007**, *50* (10), 2385-2390.
- (15) Owen, S. C.; Doak, A. K.; Ganesh, A. N.; Nedyalkova, L.; McLaughlin, C. K.; Shoichet, B. K.; Shoichet, M. S., Colloidal Drug Formulations Can Explain “Bell-Shaped” Concentration–Response Curves. *ACS Chem. Biol.* **2014**, *9* (3), 777-784.
- (16) Ganesh, A. N.; McLaughlin, C. K.; Duan, D.; Shoichet, B. K.; Shoichet, M. S., A New Spin on Antibody–Drug Conjugates: Trastuzumab-Fulvestrant Colloidal Drug Aggregates Target HER2-Positive Cells. *ACS Appl. Mater. Interfaces* **2017**, *9* (14), 12195-12202.
- (17) Ganesh, A. N.; Donders, E. N.; Shoichet, B. K.; Shoichet, M. S., Colloidal Aggregation: From Screening Nuisance to Formulation Nuance. *Nano Today* **2018**, *19*, 188-200.
- (18) Tomaszewska, E.; Soliwoda, K.; Kadziola, K.; Tkacz Szczesna, B.; Celichowski, G.; Cichomski, M.; Szmaja, W.; Grobelny, J., *Detection Limits of DLS and UV-Vis Spectroscopy in Characterization of Polydisperse Nanoparticles Colloids*. **2013**; Vol. 2013.
- (19) Hoo, C. M.; Starostin, N.; West, P.; Mecartney, M. L., A Comparison of Atomic Force Microscopy (AFM) and Dynamic Light Scattering (DLS) Methods to Characterize Nanoparticle Size Distributions. *J. Nanopart. Res.* **2008**, *10* (1), 89-96.
- (20) Hong, Y.; Lam, J. W.; Tang, B. Z., Aggregation-induced Emission: Phenomenon, Mechanism and Applications. *Chem. Commun. (Camb.)* **2009**, (29), 4332-4353.
- (21) Meiboom, S.; Gill, D., Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **1958**, *29* (8), 688-691.
- (22) Pellecchia, M.; Sem, D. S.; Wüthrich, K., NMR in Drug Discovery. *Nat. Rev. Drug Discov.* **2002**, *1*, 211-219.
- (23) Kleckner, I. R.; Foster, M. P., An Introduction to NMR-based Approaches For Measuring Protein Dynamics. *Biochim. Biophys. Acta* **2011**, *1814* (8), 942-968.
- (24) Boehr, D. D.; Dyson, H. J.; Wright, P. E., An NMR Perspective on Enzyme Dynamics. *Chem. Rev.* **2006**, *106* (8), 3055-3079.

- (25) Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W., One-Dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules. *J. Am. Chem. Soc.* **1997**, *119* (50), 12257-12261.
- (26) Wider, G.; Wüthrich, K., NMR Spectroscopy of Large Molecules and Multimolecular Assemblies in Solution. *Curr. Opin. Struct. Biol.* **1999**, *9* (5), 594-601.
- (27) Korzhnev, D. M.; Kay, L. E., Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding. *Acc. Chem. Res.* **2008**, *41* (3), 442-451.
- (28) Palmer, A. G., 3rd, Chemical Exchange in Biomacromolecules: Past, Present, and Future. *J. Magn. Reson.* **2014**, *241*, 3-17.
- (29) Dalvit, C.; Caronni, D.; Mongelli, N.; Veronesi, M.; Vulpetti, A., NMR-based Quality Control Approach for the Identification of False Positives and False Negatives in High Throughput Screening. *Curr. Drug Discov. Technol.* **2006**, *3* (2), 115-124.
- (30) Mukerjee, P.; Ghosh, A. K., Thermodynamic Aspects of the Self-association and Hydrophobic Bonding of Methylene Blue. Model System for Stacking Interactions. *J. Am. Chem. Soc.* **1970**, *92* (22), 6419-6424.
- (31) Murugesan, J. R.; Shahout, F.; Dlim, M.; Langella, M. M.; Cuadra-Foy, E.; Forgione, P.; LaPlante, S. R., Revealing Dye and Dye-drug Aggregation Into Nano-entities Using NMR. *Dyes Pigm.* **2018**, *153*, 300-306.
- (32) Doak, A. K.; Wille, H.; Prusiner, S. B.; Shoichet, B. K., Colloid Formation by Drugs in Simulated Intestinal Fluid. *J. Med. Chem.* **2010**, *53* (10), 4259-4265.
- (33) Šmejkalová, D.; Piccolo, A., Aggregation and Disaggregation of Humic Supramolecular Assemblies by NMR Diffusion Ordered Spectroscopy (DOSY-NMR). *Environ. Sci. Technol.* **2008**, *42* (3), 699-706.
- (34) Neufeld, R.; Stalke, D., Accurate Molecular Weight Determination of Small Molecules Via DOSY-NMR by Using External Calibration Curves with Normalized Diffusion Coefficients. *Chem. Sci.* **2015**, *6* (6), 3354-3364.
- (35) Simpson, A. J., Determining the Molecular Weight, Aggregation, Structures and Interactions of Natural Organic Matter Using Diffusion Ordered Spectroscopy. *Magn. Reson. Chem.* **2002**, *40* (13), S72-S82.

- (36) Chen, A.; Johnson, C. S.; Lin, M.; Shapiro, M. J., Chemical Exchange in Diffusion NMR Experiments. *J. Am. Chem. Soc.* **1998**, *120* (35), 9094-9095.
- (37) Akoka, S.; Barantin, L.; Trierweiler, M., Concentration Measurement by Proton NMR Using the ERETIC Method. *Anal. Chem.* **1999**, *71* (13), 2554-2557.
- (38) Bhattacharjee, S., DLS and Zeta Potential – What They Are And What They Are Not? *J. Controlled Release* **2016**, *235*, 337-351.
- (39) Stetefeld, J.; McKenna, S. A.; Patel, T. R., Dynamic Light Scattering: A Practical Guide and Applications in Biomedical Sciences. *Biophys. Rev.* **2016**, *8* (4), 409-427.
- (40) Hassan, P. A.; Rana, S.; Verma, G., Making Sense of Brownian Motion: Colloid Characterization by Dynamic Light Scattering. *Langmuir* **2015**, *31* (1), 3-12.
- (41) Wang, J.; Matayoshi, E., Solubility at the Molecular Level: Development of a Critical Aggregation Concentration (CAC) Assay for Estimating Compound Monomer Solubility. *Pharm. Res.* **2012**, *29* (7), 1745-1754.
- (42) Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E.-M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjögren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G., Development of a Virtual Screening Method for Identification of “Frequent Hitters” in Compound Libraries. *J. Med. Chem.* **2002**, *45* (1), 137-142.
- (43) Walters, W. P.; Murcko, A. A.; Murcko, M. A., Recognizing Molecules With Drug-like Properties. *Curr. Opin. Chem. Biol.* **1999**, *3* (4), 384-387.
- (44) Muegge, I.; Heald, S. L.; Brittelli, D., Simple Selection Criteria for Drug-like Chemical Matter. *J. Med. Chem.* **2001**, *44* (12), 1841-1846.
- (45) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings 1PII of Original Article: S0169-409X(96)00423-1. The article was originally published in *Advanced Drug Delivery Reviews* *23* (1997) 3–25.1. *Adv. Drug Del. Rev.* **2001**, *46* (1), 3-26.
- (46) Irwin, J. J.; Duan, D.; Torosyan, H.; Doak, A. K.; Ziebart, K. T.; Sterling, T.; Tumanian, G.; Shoichet, B. K., An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **2015**, *58* (17), 7076-7087.
- (47) Harner, M. J.; Frank, A. O.; Fesik, S. W., Fragment-based Drug Discovery Using NMR Spectroscopy. *J. Biomol. NMR* **2013**, *56* (2), 65-75.

- (48) Dalvit, C.; Ardini, E.; Flocco, M.; Fogliatto, G. P.; Mongelli, N.; Veronesi, M., A General NMR Method for Rapid, Efficient, and Reliable Biochemical Screening. *J. Am. Chem. Soc.* **2003**, *125* (47), 14620-14625.
- (49) Harner, M. J.; Mueller, L.; Robbins, K. J.; Reily, M. D., NMR in Drug Design. *Arch. Biochem. Biophys.* **2017**, *628*, 132-147.
- (50) Lepre, C. A.; Moore, J. M.; Peng, J. W., Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *Chem. Rev.* **2004**, *104* (8), 3641-3676.
- (51) Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* **1999**, *38* (12), 1784-1788.
- (52) Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G., Perspectives on NMR in Drug Discovery: A Technique Comes of Age. *Nat. Rev. Drug Discov.* **2008**, *7*, 738-745.
- (53) Houghten, R. A.; Pinilla, C.; Appel, J. R.; Blondelle, S. E.; Dooley, C. T.; Eichler, J.; Nefzi, A.; Ostresh, J. M., Mixture-Based Synthetic Combinatorial Libraries. *J. Med. Chem.* **1999**, *42* (19), 3743-3778.
- (54) Pinilla, C.; Appel, J. R.; Borrás, E.; Houghten, R. A., Advances in the Use of Synthetic Combinatorial Chemistry: Mixture-based Libraries. *Nat. Med.* **2003**, *9* (1), 118-122.
- (55) Koehn, F. E., High Impact Technologies for Natural Products Screening. In *Natural Compounds as Drugs Volume I*, Petersen, F.; Amstutz, R., Eds. Birkhäuser Basel: Basel, **2008**; 175-210.
- (56) Koehn, F. E.; Carter, G. T., The Evolving Role of Natural Products in Drug Discovery. *Nat. Rev. Drug Discov.* **2005**, *4*, 206-220.
- (57) Shoichet, B. K., Screening in a Spirit Haunted World. *Drug Discov. Today* **2006**, *11* (13-14), 607-615.
- (58) Trasi, N. S.; Taylor, L. S., Dissolution Performance of Binary Amorphous Drug Combinations—Impact of a Second Drug on the Maximum Achievable Supersaturation. *Int. J. Pharm.* **2015**, *496* (2), 282-290.
- (59) Feng, B. Y.; Shoichet, B. K., Synergy and Antagonism of Promiscuous Inhibition in Multiple-Compound Mixtures. *J. Med. Chem.* **2006**, *49* (7), 2151-2154.



- (60) Alhalaweh, A.; Bergström, C. A. S.; Taylor, L. S., Compromised In Vitro Dissolution and Membrane Transport of Multidrug Amorphous Formulations. *J. Controlled Release* **2016**, *229*, 172-182.
- (61) Simpson, A. J., Determining the Molecular Weight, Aggregation, Structures and Interactions of Natural Organic Matter Using Diffusion Ordered Spectroscopy. *Magn. Reson. Chem.* **2002**, *40* (13), S72-S82.
- (62) Brath, U.; Swamy, S. I.; Veiga, A. X.; Tung, C.-C.; Van Petegem, F.; Erdélyi, M., Paramagnetic Ligand Tagging To Identify Protein Binding Sites. *J. Am. Chem. Soc.* **2015**, *137* (35), 11391-11398.
- (63) Ferreira, R. S.; Bryant, C.; Ang, K. K. H.; McKerrow, J. H.; Shoichet, B. K.; Renslo, A. R., Divergent Modes of Enzyme Inhibition in a Homologous Structure–Activity Series. *J. Med. Chem.* **2009**, *52* (16), 5005-5008.
- (64) Volovik Frenkel, Y.; Gallicchio, E.; Das, K.; Levy, R. M.; Arnold, E., Molecular Dynamics Study of Non-nucleoside Reverse Transcriptase Inhibitor 4-[[4-[[4-[(E)-2-Cyanoethenyl]-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino]benzotrile (TMC278/Rilpivirine) Aggregates: Correlation between Amphiphilic Properties of the Drug and Oral Bioavailability. *J. Med. Chem.* **2009**, *52* (19), 5896-5905.
- (65) Ganesh, A. N.; Logie, J.; McLaughlin, C. K.; Barthel, B. L.; Koch, T. H.; Shoichet, B. K.; Shoichet, M. S., Leveraging Colloidal Aggregation for Drug-Rich Nanoparticle Formulations. *Mol. Pharm.* **2017**, *14* (6), 1852-1860.
- (66) McLaughlin, C. K.; Duan, D.; Ganesh, A. N.; Torosyan, H.; Shoichet, B. K.; Shoichet, M. S., Stable Colloidal Drug Aggregates Catch and Release Active Enzymes. *ACS Chem. Biol.* **2016**, *11* (4), 992-1000.
- (67) Qin, S.-Y.; Zhang, A.-Q.; Cheng, S.-X.; Rong, L.; Zhang, X.-Z., Drug Self-Delivery Systems for Cancer Therapy. *Biomaterials* **2017**, *112*, 234-247.
- (68) Zhou, J.; Xu, B., Enzyme-Instructed Self-Assembly: A Multistep Process for Potential Cancer Therapy. *Bioconjugate Chem.* **2015**, *26* (6), 987-999.
- (69) Ganesh, A. N.; Aman, A.; Logie, J.; Barthel, B. L.; Cogan, P.; Al-awar, R.; Koch, T. H.; Shoichet, B. K.; Shoichet, M. S., Colloidal Drug Aggregate Stability in High Serum Conditions and Pharmacokinetic Consequence. *ACS Chem. Biol.* **2019**, *14* (4), 751-757.

(70) Blevitt, J. M.; Hack, M. D.; Herman, K. L.; Jackson, P. F.; Krawczuk, P. J.; Lebsack, A. D.; Liu, A. X.; Mirzadegan, T.; Nelen, M. I.; Patrick, A. N.; Steinbacher, S.; Milla, M. E.; Lumb, K. J., Structural Basis of Small-Molecule Aggregate Induced Inhibition of a Protein–Protein Interaction. *J. Med. Chem.* **2017**, *60* (8), 3511-3517.

(71) Feng, B. Y.; Toyama, B. H.; Wille, H.; Colby, D. W.; Collins, S. R.; May, B. C. H.; Prusiner, S. B.; Weissman, J.; Shoichet, B. K., Small-Molecule Aggregates Inhibit Amyloid Polymerization. *Nat. Chem. Biol.* **2008**, *4*, 197-199.

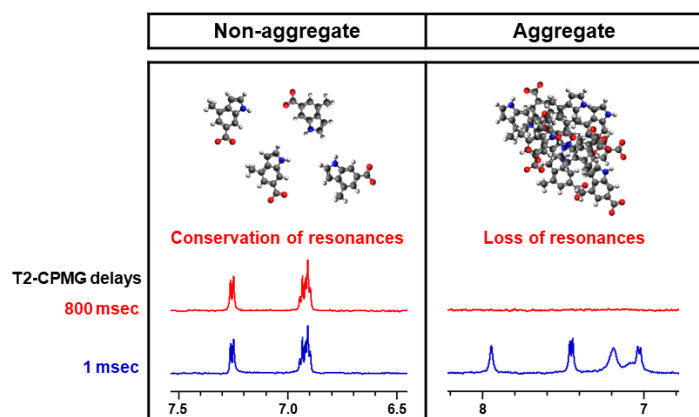
(72) Hwang, T. L.; Shaka, A. J., Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients *J. Magn. Res., Series A* **1995**, *112* (2), 275-279.

(73) Wu, D. H.; Chen, A. D.; Johnson, C. S., An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *J. Magn. Res., Series A* **1995**, *115* (2), 260-264.

(74) Evans, R.; Dal Poggetto, G.; Nilsson, M.; Morris, G. A. Improving the Interpretation of Small Molecule Diffusion Coefficients. *Anal. Chem.* **2018**, *90* (6), 3987-3994.

(75) Evans, R.; Deng, Z.; Rogerson, A.K.; MacLachlan, A.S.; Richards, J.J.; Morris, G.A. Quantitative Interpretation of Diffusion-Ordered NMR Spectra: Can We Rationalize Small Molecule Diffusion Coefficients? *Angew. Chem.* **2013**, *125* (11), 3281-3284.

### Table of content graphic



### 3 SUPPORTING INFORMATION - ARTICLE 1

---

## Exposing Small-molecule Nano-entities By a Nuclear Magnetic Resonance Relaxation Assay

Yann Ayotte,<sup>1‡</sup> Victoria M. Marando,<sup>2‡</sup> Louis Vaillancourt,<sup>2</sup> Patricia Bouchard,<sup>2</sup> Gregory Heffron,<sup>3</sup> Paul W. Coote,<sup>2,3</sup> Sacha T. Larda,<sup>\*,2</sup> and Steven R. LaPlante<sup>\*,1,2,3</sup>

<sup>1</sup> INRS-Institut Armand-Frappier Research Centre, 531 Boulevard des Prairies, Laval, Québec, H7V 1B7, CANADA

<sup>2</sup> NMX Research and Solutions, Inc., 500 Boulevard Cartier Ouest, Laval, Québec, H7V 5B7, CANADA

<sup>3</sup> Harvard Medical School, 240 Longwood Ave., Boston, MA 02115, USA

‡ These authors contributed equally to this work

\* Corresponding authors

Published in: Journal of Medicinal Chemistry, 2019, 62, 17, 7885-7896.

DOI : 10.1021/acs.jmedchem.9b00653

#### Table of Contents:

- 1) Quantification of peak area and relaxation decay rates
- 2) Control non-aggregator and known aggregators in the NMR dilution assay
- 3) Sensitivity for aggregates of small molecules
- 4) Additional series of structurally related compounds with distinct T2-CPMG profiles
- 5) Consistency of T2-CPMG at other compound concentrations
- 6) Evidence for the detection of chemical exchange
- 7) Consistency between the use of T2 values and peak area percentages
- 8) Compound information

## 1) Quantification of peak area and relaxation decay rates

CPMG spectra for all compounds were read into Matlab (R2016b, The MathWorks, Inc. Natick, Massachusetts, United States) using the toolbox `matNMR` (<http://matnmr.sourceforge.net/>). Artificial line broadening of 3 Hz was added to each spectrum by convolving with a Lorentzian line shape kernel using the Matlab function 'conv'. The kernel was defined explicitly using the equation of an absorption-mode Lorentzian line shape.<sup>1</sup> Line broadening was added to improve signal-to-noise ahead of numerical calculations.

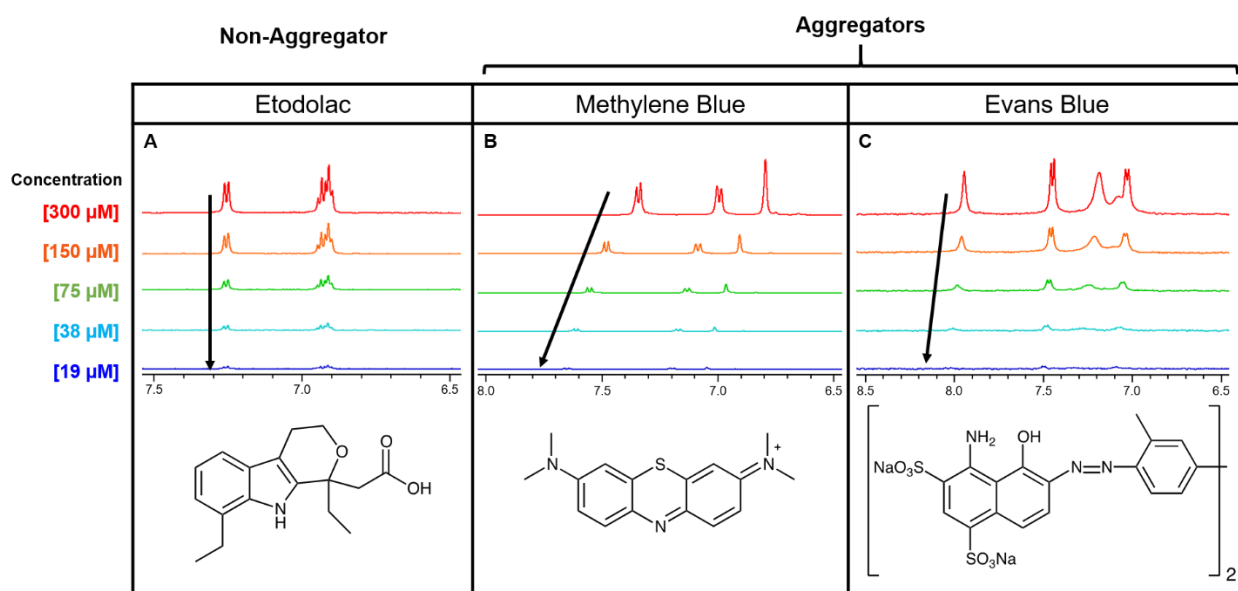
Regions containing peaks were detected in each 1 ms delay CPMG spectrum by thresholding above the 96th percentile of spectrum amplitude (excluding a 0.3 ppm window around the strong DMSO reference peak). Regions narrower than 0.007 ppm were discarded as being due to spikes in the noise. Each contiguous region was then dilated by 0.05 ppm to ensure that the tails of the peak were included. This used the 'imdilate' function from the Matlab Image Processing Toolbox with a unit-kernel of width  $0.05/(\text{sweep width in ppm}) \times (\text{number of points})$ , rounded to the nearest integer. In some cases, a cluster of nearby peaks (either a multiplet or closely spaced peaks) end up in the same region, and are therefore treated together in subsequent calculations. Due to the 0.05 ppm dilation, the minimum space between peaks in separate regions is 0.1 ppm, and peaks closer together than that are treated as a single cluster or multiplet.

Amplitude in the first CPMG spectrum (1 ms delay) was calculated for each region by integrating the spectrum in that region using the 'sum' function in Matlab. For each region, the decrease in amplitude due to T2 decay in all CPMG spectra was found by projecting the 1 ms CPMG onto each subsequent CPMG spectrum. For this projection we used the '\' (backslash or left matrix divide) function in Matlab. This is the solution  $a(\Delta)$  in the least squares sense of the equation  $CPMG(\Delta) = a(\Delta) * CPMG(1)$ , where  $\Delta$  is the relaxation delay period,  $CPMG(1)$  is the spectrum with 1 ms relaxation delay,  $CPMG(\Delta)$  is a CPMG spectrum with another delay, and  $a(\Delta)$  is scalar showing the decrease in intensity between the two spectra. This calculation is done separately for each region containing a peak.

The scalars  $a(\Delta)$  for  $\Delta = 1, 25, 50, 100, 200, 300, 500,$  and 800 ms were fit to an exponential decay using the 'fit' function with option 'exp1' in Matlab, to estimate T2 decay rates for each region/peak in each spectrum. In high throughput applications, we instead only use spectra for  $\Delta = 1$  and  $\Delta = 800$  ms. Scalars  $a(\Delta)$  for  $\Delta = 800$  ms were reported in the third column in Figures 6, 7, and S3.

## 2) Control non-aggregator and known aggregators in the NMR dilution assay

The results obtained from the T2-CPMG assay for etodolac, methylene blue and Evans blue correlate with results obtained from a previously published NMR dilution-based aggregation assay. The NMR dilution-based assay flags aggregates based on characteristic changes in ligand spectra across a range of concentrations.<sup>2-4</sup> In this assay, etodolac exhibits the behavior expected for a non-aggregator: as compound concentration decreases, peak intensity decreases with no changes observed in peak shape, chemical shift or number of resonances (Figure S1A). In contrast, two previously reported aggregators, methylene blue and Evans blue, show the expected behavior of aggregating compounds (Figures S1B,C). As the concentration of both methylene blue and Evans blue decreases, there are systematic changes in chemical shift, number of peaks, or line broadening. These behaviors indicate that these compounds are self-associating and forming aggregates at high concentrations, hindering their ability to freely tumble thus changing the molecular environment of the observed nuclei.

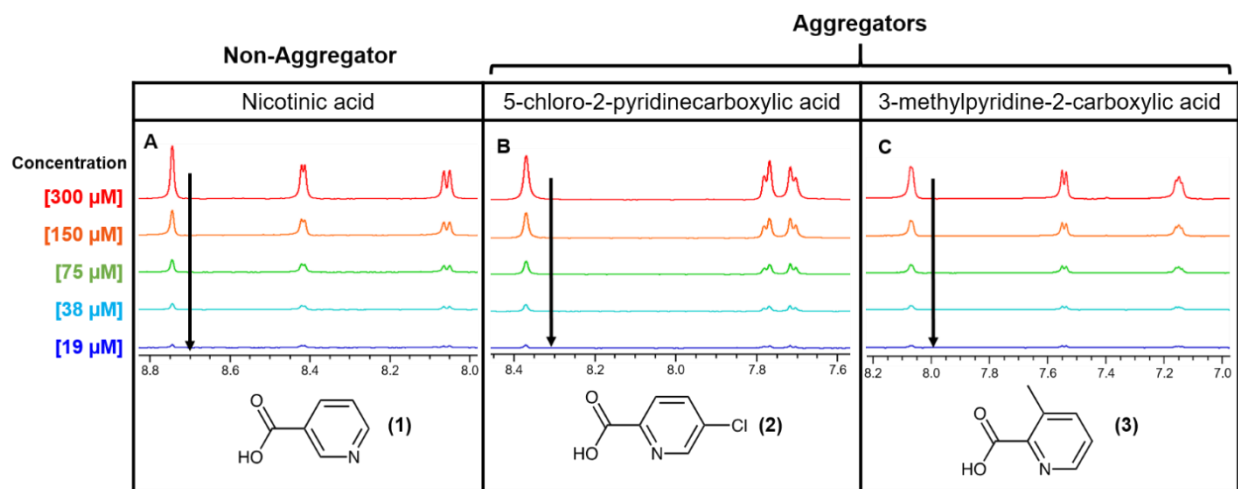


**Figure S1.** Dilution assay NMR spectra for three compounds illustrated in Figure 4. Panels A-C show a series of <sup>1</sup>H-NMR spectra for concentrations ranging from 300 to 19 μM. The dilution profile supports the data obtained by the T2-CPMG aggregation assay.

## 3) Sensitivity for aggregates of small molecules

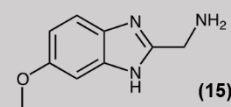
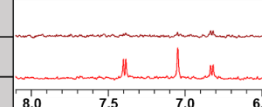
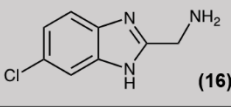
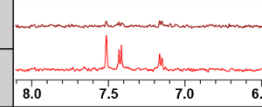
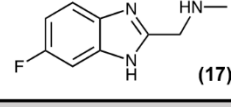
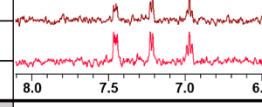
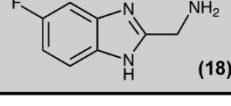
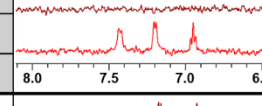
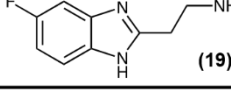
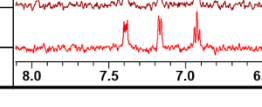
In the NMR dilution method, the control fragment (**1**), displays no significant changes in resonance number, shape and chemical shift (Figure S2A), consistent with the behavior of a non-aggregator.

The T2-CPMG profile is also consistent with that of a non-aggregator. However, in the case of aggregating fragments **2** and **3**, the T2-CPMG results from Figure 5 do not correlate with the observations made with the NMR dilution assay (Figure S2B,C). As compound concentration increases, there are no telltale signs of aggregation in the dilution assay and based solely on this assay, these compounds would have been cleared as non-aggregators. The T2-CPMG assay clearly demonstrates that this is not the case.



**Figure S2.** Dilution assay NMR spectra for three fragments illustrated in Figure 5. Panels A-C show a series of  $^1\text{H}$ -NMR spectra for concentrations ranging from 300 to 19  $\mu\text{M}$ . While the dilution assay suggest that these compounds do not aggregate, the T2-CPMG assay indicates that these compounds undergo self-association.

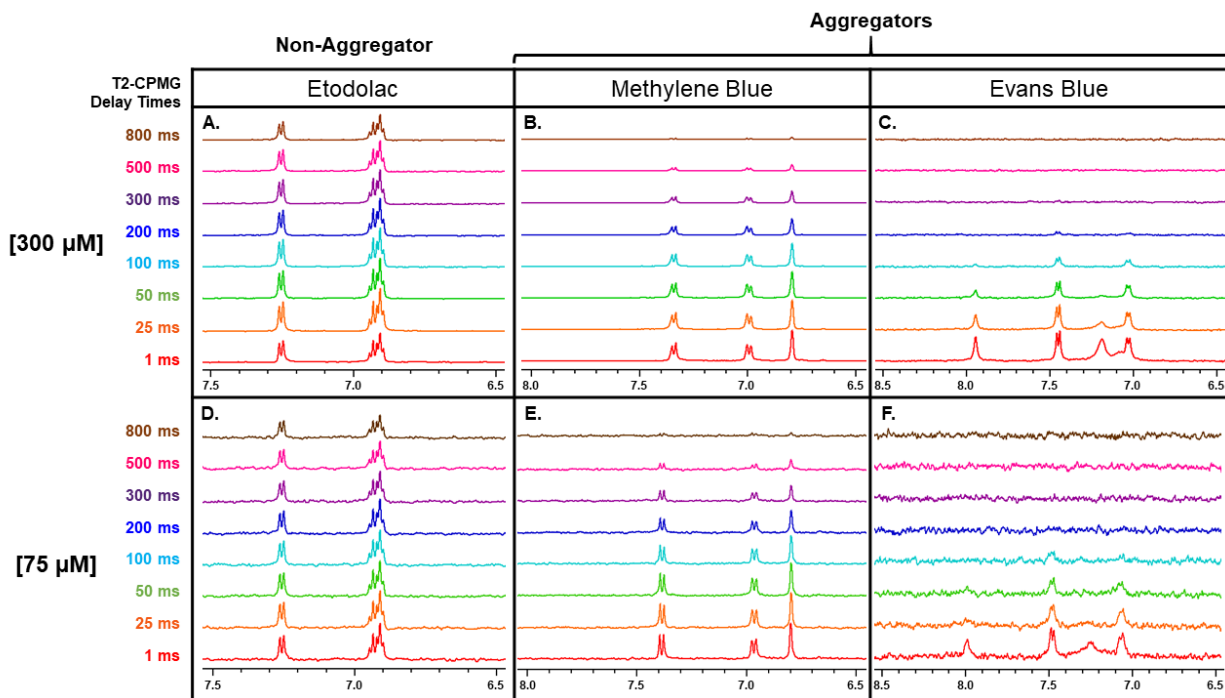
#### 4) Additional series of structurally related compounds with distinct T2-CPMG profiles

Behavior	Structure	Peak Chemical Shift (6-10 ppm)	Peak Area 800 ms/1 ms T2-CPMG	NMR (aromatic region)
Aggregator	 (15)	6.83	41%	
		7.05	12%	
		7.40	15%	
Aggregator	 (16)	7.16	34%	
		7.47 - 7.52	16%	
Non-Aggregator	 (17)	6.98	71%	
		7.25	71%	
		7.46	52%	
Aggregator	 (18)	6.96	6%	
		7.20	0%	
		7.43	4%	
Non-Aggregator	 (19)	6.93	73%	
		7.17	85%	
		7.40	80%	

**Figure S3.** Quantification of a series of structurally related benzimidazole core compounds. For each compound, peaks in the aromatic region of the spectra (6-10 ppm) are identified and the peak area was determined by integration. The peak areas between 1 ms and 800 ms delay times were compared and the percentage difference was calculated. These percentages can be compared to the NMR spectra in the far right column. The bottom, red spectrum has a 1 ms delay time and the top, brown spectrum has an 800 ms delay time.

#### 5) Consistency of T2-CPMG at other compound concentrations

Although the compound concentrations used for T2-CPMG experiments throughout this manuscript have been 300  $\mu\text{M}$ , the T2-CPMG assay can be used at various concentrations so long as there is sufficient NMR signal. Figure S4 compares the T2-CPMG profiles illustrated in Figure 4 at 300  $\mu\text{M}$  (Figure S4A-C) and the same experiments performed at 75  $\mu\text{M}$  compound concentration (Figure S4D-F). There is excellent agreement between the spectra observed between 300 and 75  $\mu\text{M}$ , although it is highly conceivable that different results could be observed as a compound may aggregate at a higher concentration but not at a lower one.



**Figure S4.** Shown are three examples demonstrating that the T2-CPMG assay can be run at different compound concentrations. Panels A-C show a stack of the  $^1\text{H}$ -NMR spectra from various T2-CPMG delay times for a constant compound concentration of  $300\ \mu\text{M}$ . Panels D-F show the stack of  $^1\text{H}$ -NMR spectra superimposed from the same T2-CPMG delay times with a constant compound concentration of  $75\ \mu\text{M}$ .

## 6) Evidence for the detection of chemical exchange

A range of compounds were further evaluated along with **1**, **2**, **3**, etodolac, methylene blue and Evans blue (Table 1). Observed concentration for each compound was estimated by NMR both in buffer and solvent using an external standard reference.<sup>5</sup>



**Table 1. Comparison of solution properties for a range of small molecules.**

Compounds	Observed concentration ( $\mu\text{M}$ )		Peak area (800 ms/1 ms) (%)	T2 (ms)	D ( $\text{m}^2/\text{s}$ )	Descriptor
	Buffer	Solvent				
L-tryptophan	406	531	71	2405	4.56E-10	Non-aggregators
Etodolac	428	375	60	1446	3.94E-10	
Tartrazine	295	284	52	1267	3.2E-10	
1	305	379	52	1225	5.87E-10	
Imidazole	483	369	51	1109	8.4E-10	
Thiamine	373	308*	44	1008	3.56E-10	Intermediate behaviors
Lansoprazole	390	393	41	1070	3.21E-10	
Gefitinib	267	381	31	780	1.56E-10	
Erythrosin B	330	329	24	551	2.75E-10	Aggregators
Chlorpromazine	345	391	11	362	3.66E-10	
Light green SF yellowish	230	165	10	356	2.39E-10	
Acid blue 9	275	316	9	357	2.24E-10	
Flutamide	378	436	7	305	4.02E-10	
3	347	399	6	61	5.33E-10	
2	380	265	0	116	5.98E-10	
Evans blue	188	179	0	34	3.27E-12	
Imatinib	191	333	11	334	2.61E-10	
Methylene blue	274	395*	1	130	3.18E-10	
Allura red	284	361	0	121	1.95E-10	

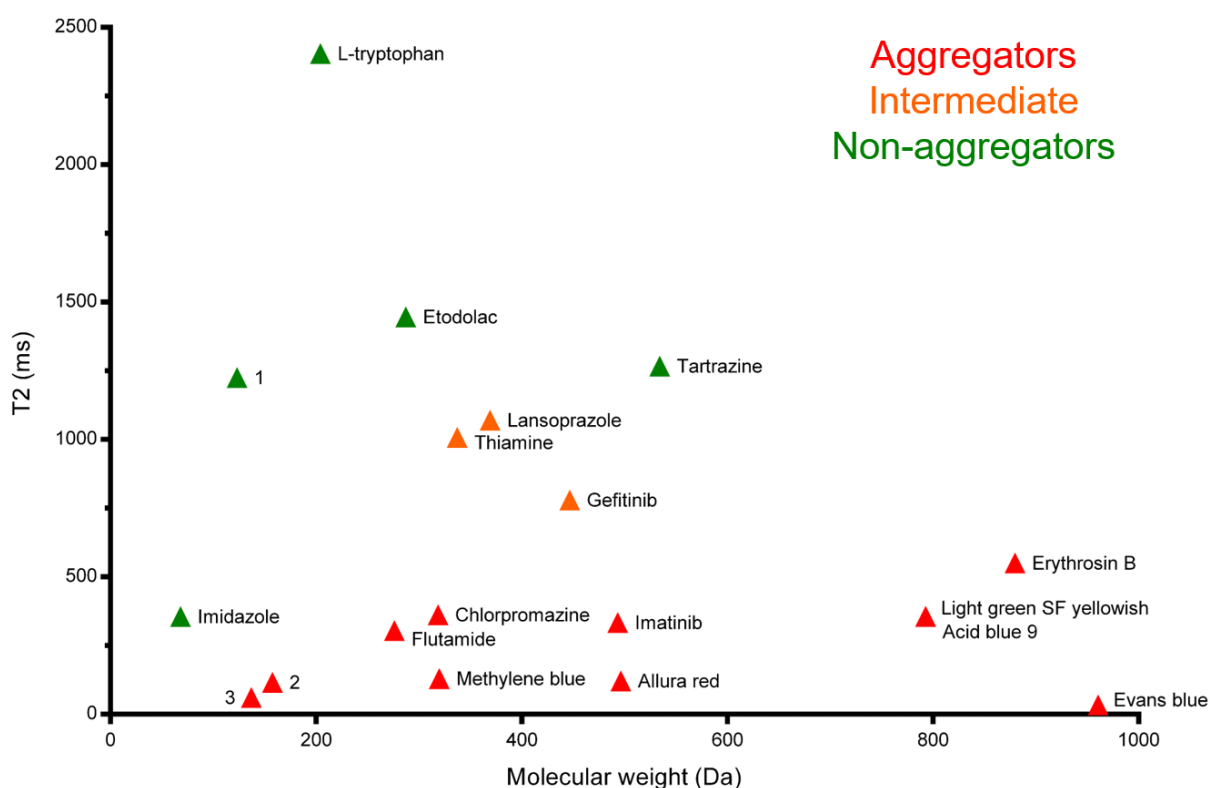
Nominal concentration was 300  $\mu\text{M}$ . Compounds with remaining T2-CPMG signal intensity below 25%, between 25% and 50% and above 50% were classified as aggregators, intermediate behaviors and non-aggregators, respectively.

Buffer: 50 mM NaPi, 100 mM NaCl, 100 %  $\text{D}_2\text{O}$ , pH 7.4. Solvent: 1:1 DMSO- $\text{d}_6$  and ACN- $\text{d}_3$ .

\*Due to poor solubility in DMSO-ACN, the solvent used for these samples was 100 % DMSO- $\text{d}_6$ .

Measured peak area percentages span a wide range of values and when combined with solubility data, allows for a separation of these molecules into three categories: 1) non-aggregators, 2) intermediate behaviors, and 3) aggregators. Interestingly, the majority of compounds in the last two categories have previously been reported as aggregators or as having promiscuous behaviors.<sup>4,6-9</sup>

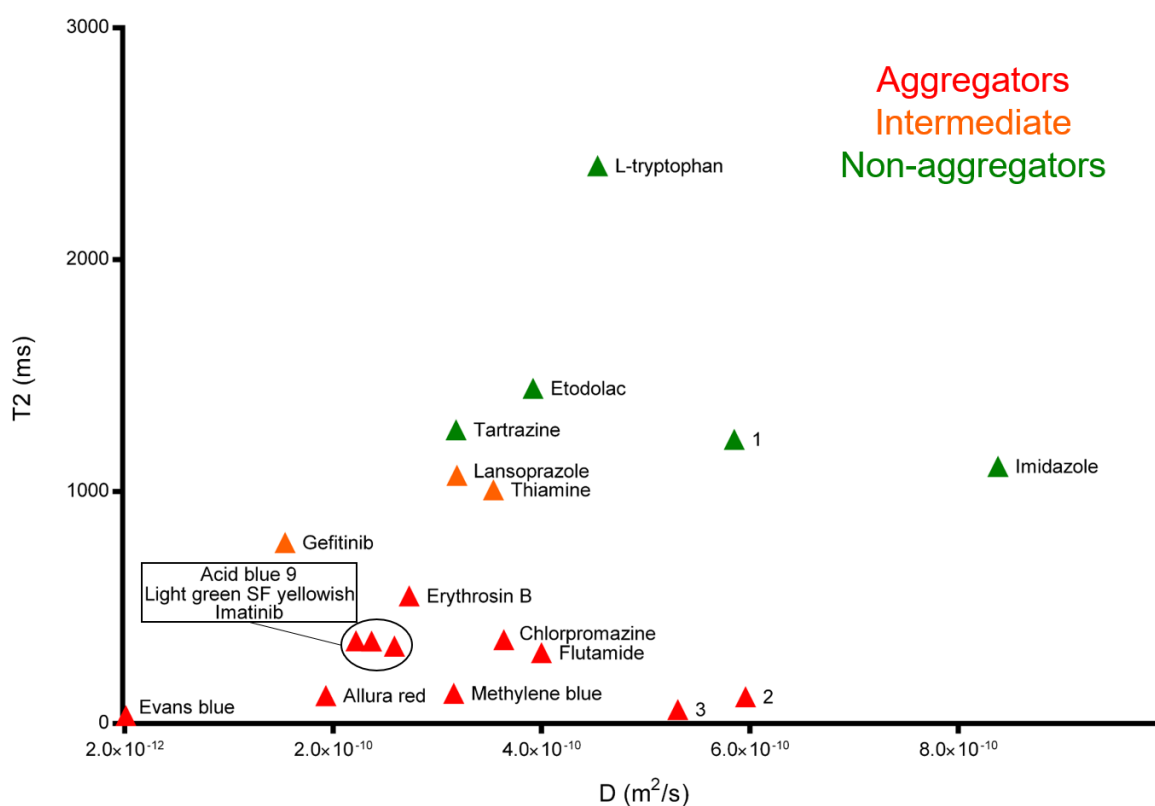
T2 relaxation times and diffusion coefficients were determined for all samples using T2-CPMG and diffusion-ordered spectroscopy (DOSY) NMR experiments, respectively. A comparison was done to determine if diffusion values corroborate the measurements of T2. DOSY provides a measure of translational diffusion and can therefore provide information on molecular size and shape, as well as dynamics.<sup>10-13</sup> Comparisons between T2-CPMG and diffusion coefficients versus molecular weight of the compounds primary structures are shown in Figures S5 and S6, respectively. Overall correlations can be noted, especially when considering theoretical versus measured diffusion data as shown in Figure 5.



**Figure S5.** Correlation between T2 relaxation and molecular weight. While there is an expected correlation between T2 and molecular weight, chemical exchange processes also contribute to transverse relaxation rates. Therefore, compounds undergoing chemical exchange between states would exhibit shorter T2 relaxation times than compounds not undergoing exchange processes, as shown. Furthermore, compound aggregates are of larger size and would also exhibit shorter transverse relaxation times than freely soluble species. It is important to note that the molecular weights shown above are those expected for the free compounds. Aggregators therefore have T2 relaxation times that are much shorter than would be expected for the free

compounds. Experimental points were colored according to their loss in T2-CPMG signal intensity between 1 and 800 ms delays (red, more than 75% loss in signal intensity; orange, between 50% and 75%; green, less than 50%).

In contrast, the T2-CPMG experiment is able to indirectly detect the presence of some very large “NMR-invisible” aggregates. The chemical exchange contribution to T2 resulting from exchange between single molecules and aggregated species of various sizes is therefore a more appropriate reporter of self-association and aggregation.

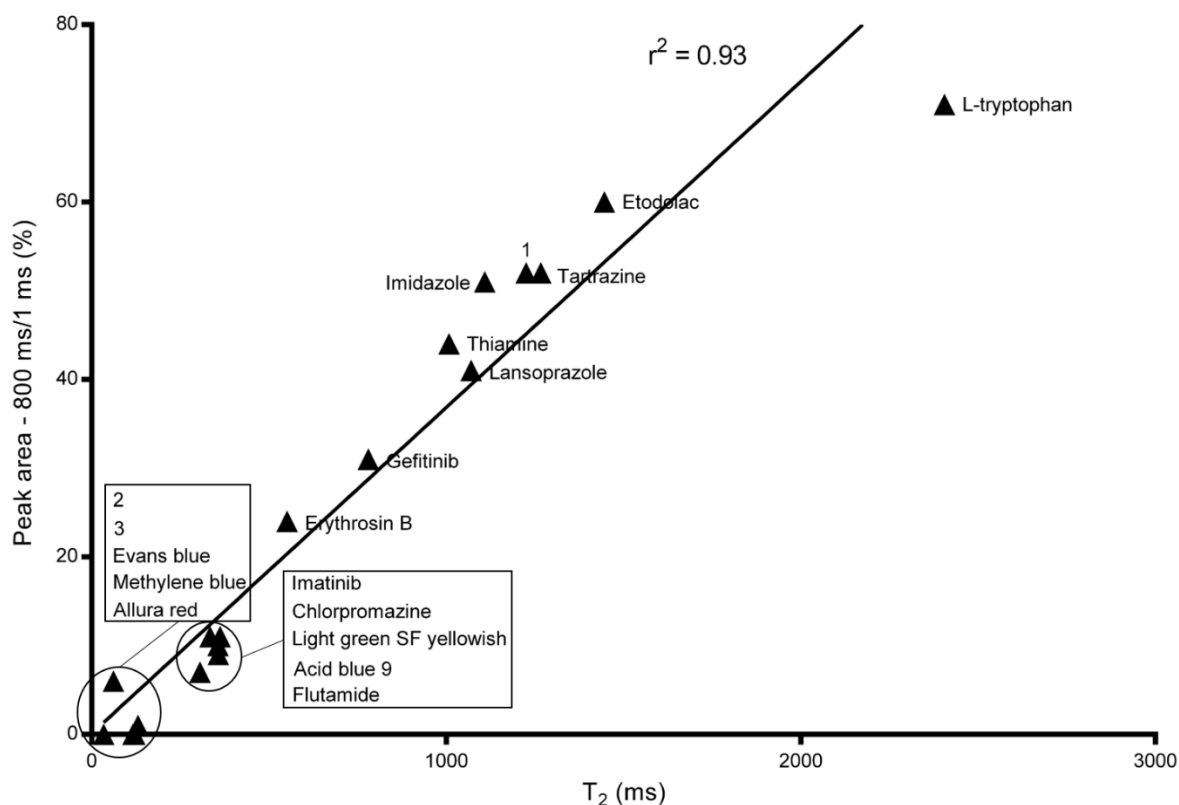


**Figure S6.** There is only modest correlation between T2 and diffusion coefficient. As mentioned previously, this can be explained by chemical exchange contributions to T2. While diffusion is sensitive primarily to molecular size and has a detection range limited to smaller species that are detectable by NMR, T2 can elucidate very large “NMR invisible” entities indirectly via chemical exchange. It is possible that the diffusion coefficients measured for aggregators only represents the “NMR visible” fraction, thereby misrepresenting the true size distribution of aggregate species in the sample. Experimental points were colored according to their loss in T2-CPMG signal

intensity between 1 and 800 ms delays (red, more than 75% loss in signal intensity; orange, between 50% and 75%; green, less than 50%).

### 7) Consistency between the use of T2 and peak area percentages

Figures 6, 7 and S3 illustrate that only two T2-CPMG delays can be used to rapidly assess compound solution behaviors. This can prove very useful in reducing NMR acquisition time when curating large compound libraries. Figure S7 illustrates the correlation between the T2 and peak area calculation for all compounds present in Table 1. Peak areas were measured using two delay times as described above, while the T2 relaxation times were obtained by fitting across eight T2-CPMG delays. Therefore, peak areas can serve as reliable and fast filters in order to quickly curate large numbers of compounds.



**Figure S7.** Percentages of peak area for 800 versus 1 milliseconds delays correlate well with the T2 relaxation times obtained by fitting several delay points.

## 8) Compound information

Compound Name	SMILES	Supplier	Catalog No.	CAS	In this manuscript
Light green SF yellowish	<chem>CCN(CC1=CC(=CC=C1)S(=O)(=O)[O-])C2=CC=C(C=C2)C(=C3C=CC(=[N+](CC)CC4=CC(=CC=C4)S(=O)(=O)[O-])C=C3)C5=CC=C(C=C5)S(=O)(=O)[O-].[Na+].[Na+]</chem>	Alfa Aesar	B23330	5141-20-8	Figures 1, 5, S5-S7, Table 1
Evans Blue	<chem>CC1=C(C=CC(=C1)C2=CC(=C(C=C2)N=NC3=C(C4=C(C=C3)C(=CC(=C4N)S(=O)(=O)[O-])S(=O)(=O)[O-])O)C)N=NC5=C(C6=C(C=C5)C(=CC(=C6N)S(=O)(=O)[O-])S(=O)(=O)[O-])O.[Na+].[Na+].[Na+].[Na+]</chem>	Sigma	E2129	314-13-6	Figures 1, 4, 5, S1, S4-S7, Table 1
Acid violet 49	<chem>CCN(CC1=CC(=CC=C1)S(=O)(=O)O)C2=CC=C(C=C2)C=C3C=CC(=[N+](C)C)C=C3)C4=CC=C(C=C4)N(CC)CC5=CC(=CC=C5)S(=O)(=O)[O-].[Na+]</chem>	TCI	42640	1694-09-3	Figure 1

Pranlukast	<chem>C1=CC=C(C=C1)CC CCOC2=CC=C(C=C 2)C(=O)NC3=CC=C C4=C3OC(=CC4=O) C5=NNN=N5</chem>	Beta Pharma	56-05418	103177-37-3	Figure 1
2-methyl-1,3(2H,4H)- isoquinolinedione	<chem>CN1C(=O)CC2=CC= CC=C2C1=O</chem>	Key Organics	10L-733	4494-53-5	Figure 2
Etodolac	<chem>CCC1=CC=CC2=C1 NC3=C2CCOC3(CC) CC(=O)O</chem>	Sigma	E0516	41340-25-4	Figures 4, 5, S1, S4-S7, Table 1
Methylene Blue	<chem>CN(C)C1=CC2=C(C =C1)N=C3C=CC(=[N +](C)C)C=C3S2.[Cl-]</chem>	Sigma	M9140	122965-43-9	Figures 4, 5, S1, S4-S7, Table 1
Nicotinic acid (1)	<chem>C1=CC(=CN=C1)C(= O)O</chem>	Key Organics	PS-4255	59-67-6	Figures 5, 6, 9, S2, S5-S7, Table 1
5-chloro-2- pyridinecarboxylic acid (2)	<chem>C1=CC(=NC=C1Cl)C (=O)O</chem>	Key Organics	BB-0607	86873-60-1	Figures 5, 6, 9, S2, S5-S7, Table 1
3-methylpyridine-2- carboxylic acid (3)	<chem>CC1=C(N=CC=C1)C (=O)O</chem>	Key Organics	PS-3239	4021-07-2	Figures 5, 6, S2, S5-S7, Table 1

(1-methyl-1H-benzimidazol-2-yl)methanol (4)	<chem>CN1C2=CC=CC=C2 N=C1CO</chem>	Chembridge	4900605	7467-35-8	Figure 7
N-methyl-1-(1-methyl-1H-benzimidazol-2-yl)methanamine dihydrochloride (5)	<chem>CNCC1=NC2=CC=C C=C2N1C.Cl.Cl</chem>	Chembridge	4015772	N/A	Figure 7
[(1-methyl-1H-benzimidazol-2-yl)methyl]amine hydrochloride (6)	<chem>CN1C2=CC=CC=C2 N=C1CN</chem>	Chembridge	4100112	N/A	Figure 7
[2-(1-methyl-1H-benzimidazol-2-yl)ethyl]amine dihydrochloride (7)	<chem>CN1C2=CC=CC=C2 N=C1CCN.Cl.Cl</chem>	Chembridge	4100113	N/A	Figure 7
[3-(1-methyl-1H-benzimidazol-2-yl)propyl]amine dihydrochloride hydrate (8)	<chem>CN1C2=CC=CC=C2 N=C1CCCN.O.Cl.Cl</chem>	Chembridge	4033392	N/A	Figure 7
6-methyl-1,2,3,4-tetrahydro-11H-pyrido[2,1-b]quinazolin-11-one (9)	<chem>CC1=CC=CN2C(=O) C3=C(CCCC3)N=C1 2</chem>	Chembridge	9221258	N/A	Figure 8
6-hydroxy-1,2,3,4-tetrahydro-11H-pyrido[2,1-b]quinazolin-11-one (10)	<chem>OC1=CC=CN2C(=O) C3=C(CCCC3)N=C1 2</chem>	Chembridge	9200099	N/A	Figure 8

3-ethyl-9-hydroxy-2-methyl-4H-pyrido[1,2-a]pyrimidin-4-one (11)	<chem>CCC1=C(N=C2C(=CC=CN2C1=O)O)C</chem>	Chembridge	9226831	N/A	Figure 8
3-ethyl-2,6-dimethyl-4H-pyrido[1,2-a]pyrimidin-4-one (12)	<chem>CCC1=C(N=C2C(=CC=C(N2C1=O)C)C)C</chem>	Chembridge	9205478	N/A	Figure 8
3-ethyl-2-methyl-4H-pyrido[1,2-a]pyrimidin-4-one (13)	<chem>CCC1=C(N=C2C(=CC=CN2C1=O)C)C</chem>	Chembridge	9281741	N/A	Figure 8
3-(4-bromophenoxy)propanenitrile (14)	<chem>C1=CC(=CC=C1OCC#N)Br</chem>	Key Organics	PS-3130	118449-57-3	Figure 10
[(6-methoxy-1H-benzimidazol-2-yl)methyl]amine dihydrochloride (15)	<chem>Cl.Cl.COC1ccc2nc(CN)[nH]c2c1</chem>	Chembridge	4101295	N/A	Figure S3
[(6-chloro-1H-benzimidazol-2-yl)methyl]amine dihydrochloride (16)	<chem>Cl.Cl.NCc1nc2ccc(Cl)cc2[nH]1</chem>	Chembridge	4101293	N/A	Figure S3
[(6-fluoro-1H-benzimidazol-2-yl)methyl]methylamine dihydrochloride (17)	<chem>Cl.Cl.CNCc1nc2ccc(F)cc2[nH]1</chem>	Chembridge	4004431	N/A	Figure S3



[(5-fluoro-1H-benzimidazol-2-yl)methyl]amine dihydrochloride (18)	<chem>C1=CC2=C(C=C1F)NC(=N2)CN.Cl.Cl</chem>	Chembridge	4002683	N/A	Figure S3
[2-(5-fluoro-1H-benzimidazol-2-yl)ethyl]methylamine dihydrochloride (19)	<chem>CNCCC1=NC2=C(N1)C=C(C=C2)F.Cl.Cl</chem>	Chembridge	4009689	N/A	Figure S3
Imidazole	<chem>C1=CN=CN1</chem>	Sigma	I2399	288-32-4	Figures 5, S5-S7, Table 1
Lansoprazole	<chem>CC1=C(C=CN=C1C S(=O)C2=NC3=CC=CC=C3N2)OCC(F)(F)F</chem>	Sigma	L8533	103577-45-3	Figures 5, S5-S7, Table 1
Tartrazine	<chem>C1=CC(=CC=C1N=NC2C(=NN(C2=O)C3=CC=C(C=C3)S(=O)(=O)[O-])C(=O)[O-])S(=O)(=O)[O-].[Na+].[Na+].[Na+]</chem>	Alfa Aesar	A17682	1934-21-0	Figures 5, S5-S7, Table 1
Thiamine	<chem>CC1=C(SC=[N+]1CC2=CN=C(N=C2N)C)CCO</chem>	Sigma	T1270	67-03-8	Figures 5, S5-S7, Table 1
L-tryptophan	<chem>C1=CC=C2C(=C1)C(=CN2)CC(C(=O)O)N</chem>	RPI	T60080	73-22-3	Figures 5, S5-S7, Table 1

Erythrosin B	<chem>C1=CC=C2C(=C1)C(=O)OC23C4=CC(=C(C(=C4OC5=C(C(=C(C=C35)I)O)I)O)I</chem>	TCI	T0557	16423-68-0	Figures 5, S5-S7, Table 1
Gefitinib	<chem>COC1=C(C=C2C(=C1)N=CN=C2NC3=CC(=C(C=C3)F)CI)OCCCN4CCOCC4</chem>	Beta Pharma	86-33451	184475-35-2	Figures 5, S5-S7, Table 1
Acid blue 9	<chem>CCN(CC1=CC(=CC=C1)S(=O)(=O)[O-])C2=CC=C(C=C2)C(=C3C=CC(=[N+](CC)CC4=CC(=CC=C4)S(=O)(=O)[O-])C=C3)C5=CC=CC=C5S(=O)(=O)[O-].[NH4+].[NH4+]</chem>	TCI	B0790	3844-45-9	Figures 5, S5-S7, Table 1
Chlorpromazine	<chem>CN(C)CCCN1C2=CC=CC=C2SC3=C1C=C(C=C3)Cl</chem>	AK Scientific	M176	50-53-3	Figures 5, S5-S7, Table 1
Flutamide	<chem>CC(C)C(=O)NC1=CC(=C(C=C1)[N+](=O)[O-])C(F)(F)F</chem>	Sigma	F9397	13311-84-7	Figures 5, S5-S7, Table 1
Allura red	<chem>CC1=CC(=C(C=C1S(=O)(=O)[O-])OC)N=NC2=C(C=C3C3=C2C=CC(=C3)S(=O)(=O)[O-])O.[Na+].[Na+]</chem>	TCI	A0943	25956-17-6	Figures 5, S5-S7, Table 1

Imatinib	<chem>CC1=C(C=C(C=C1)NC(=O)C2=CC=C(C=C2)CN3CCN(CC3)C)NC4=NC=CC(=N4)C5=CN=CC=C5</chem>	Beta Pharma	86-33437	152459-95-5	Figures 5, S5-S7, Table 1
----------	---	----------------	----------	-------------	---------------------------

## References

1. Brown, K. C., *Essential Mathematics for NMR and MRI Spectroscopists*. Royal Society of Chemistry: **2016**.
2. LaPlante, S. R.; Carson, R.; Gillard, J.; Aubry, N.; Coulombe, R.; Bordeleau, S.; Bonneau, P.; Little, M.; O'Meara, J.; Beaulieu, P. L., Compound Aggregation in Drug Discovery: Implementing a Practical NMR Assay for Medicinal Chemists. *J. Med. Chem.* **2013**, *56* (12), 5142-5150.
3. LaPlante, S. R.; Aubry, N.; Bolger, G.; Bonneau, P.; Carson, R.; Coulombe, R.; Sturino, C.; Beaulieu, P. L., Monitoring Drug Self-Aggregation and Potential for Promiscuity in Off-Target In Vitro Pharmacology Screens by a Practical NMR Strategy. *J. Med. Chem.* **2013**, *56* (17), 7073-7083.
4. Murugesan, J. R.; Shahout, F.; Dlim, M.; Langella, M. M.; Cuadra-Foy, E.; Forgione, P.; LaPlante, S. R., Revealing Dye and Dye-drug Aggregation Into Nano-entities using NMR. *Dyes Pigm.* **2018**, *153*, 300-306.
5. Akoka, S.; Barantin, L.; Trierweiler, M., Concentration Measurement by Proton NMR Using the ERETIC Method. *Anal. Chem.* **1999**, *71* (13), 2554-2557.
6. Doak, A. K.; Wille, H.; Prusiner, S. B.; Shoichet, B. K., Colloid Formation by Drugs in Simulated Intestinal Fluid. *J. Med. Chem.* **2010**, *53* (10), 4259-4265.
7. Schreier, S.; Malheiros, S. V. P.; de Paula, E., Surface Active Drugs: Self-association and Interaction with Membranes and Surfactants. Physicochemical and Biological Aspects. *Biochim. Biophys. Acta* **2000**, *1508* (1), 210-234.
8. Ganesan, L.; Buchwald, P., The Promiscuous Protein Binding Ability of Erythrosine B Studied by Metachromasy (metachromasia). *J. Mol. Recognit.* **2013**, *26* (4), 181-9.

9. Owen, S. C.; Doak, A. K.; Wassam, P.; Shoichet, M. S.; Shoichet, B. K., Colloidal Aggregation Affects the Efficacy of Anticancer Drugs in Cell Culture. *ACS Chem. Biol.* **2012**, 7 (8), 1429-1435.
10. Šmejkalová, D.; Piccolo, A., Aggregation and Disaggregation of Humic Supramolecular Assemblies by NMR Diffusion Ordered Spectroscopy (DOSY-NMR). *Environ. Sci. Technol.* **2008**, 42 (3), 699-706.
11. Neufeld, R.; Stalke, D., Accurate Molecular Weight Determination of Small Molecules Via DOSY-NMR by Using External Calibration Curves with Normalized Diffusion Coefficients. *Chem. Sci.* **2015**, 6 (6), 3354-3364.
12. Simpson, A. J., Determining the Molecular Weight, Aggregation, Structures and Interactions of Natural Organic Matter Using Diffusion Ordered Spectroscopy. *Magn. Reson. Chem.* **2002**, 40 (13), S72-S82.
13. Chen, A.; Johnson, C. S.; Lin, M.; Shapiro, M. J., Chemical Exchange in Diffusion NMR Experiments. *J. Am. Chem. Soc.* **1998**, 120 (35), 9094-9095.



## 4 ARTICLE 2 – INTEGRATING MULTIPLE ASSAYS INTO A PRACTICAL PROTOCOL TO MONITOR NANO-ENTITIES

---

### **Probing the free-state solution behavior of drugs and their tendencies to self-aggregate into nano-entities**

Steven R. LaPlante<sup>1,2\*</sup>, Valérie Roux<sup>1</sup>, Fatma Shahout<sup>1</sup>, Gabriela LaPlante<sup>2</sup>, Simon Woo<sup>1</sup>, Maria M. Denk<sup>2</sup>, Sacha T. Larda<sup>2</sup>, Yann Ayotte<sup>1</sup>

<sup>1</sup> Université du Québec, INRS-Centre Armand-Frappier Santé Biotechnologie, 531 Boulevard des Prairies, Laval, Québec, H7V 1B7, CANADA

<sup>2</sup> NMX Research and Solutions, Inc., 500 Boulevard Cartier Ouest, Laval, Québec, H7V 5B7, CANADA

\* Corresponding author: [steven.laplante@inrs.ca](mailto:steven.laplante@inrs.ca)

Published in : Nature Protocols, 2021, 16, 5250-5273.

DOI : 10.1038/s41596-021-00612-3.

#### **Authors contributions :**

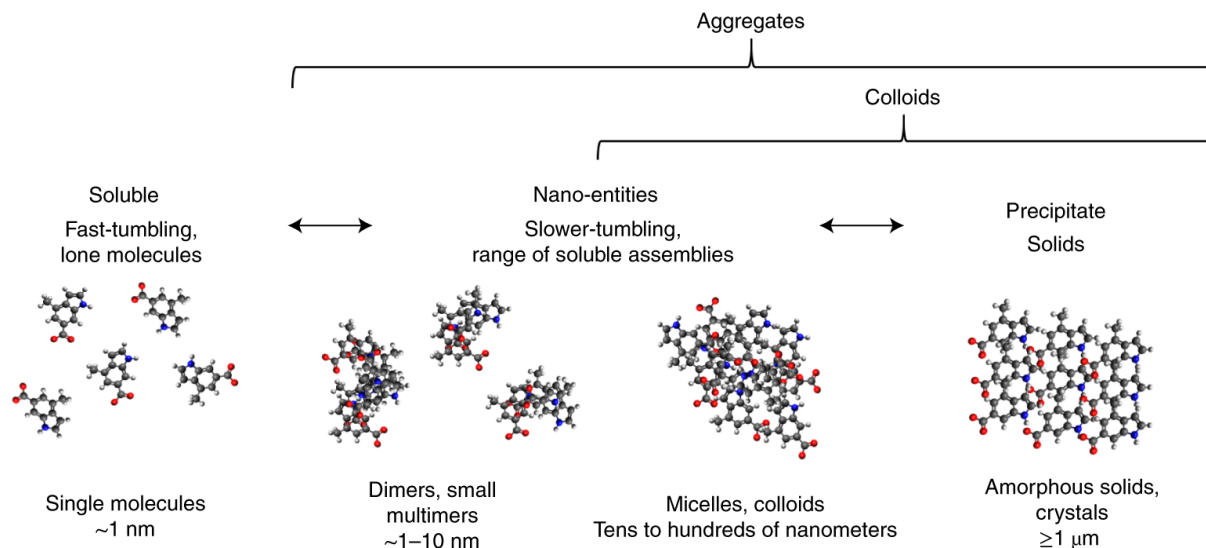
Steven Laplante conceived the concepts described in this report. Steven Laplante and Valérie Roux wrote the paper. Yann Ayotte critiqued and revised the manuscript and figures, before and following submission. Valérie Roux, Fatma Shahout, Gabriella Laplante and Maria Denk performed the experiments under supervision from Yann Ayotte, Steven Laplante and Sacha Larda. Steven Laplante, Sacha Larda, Yann Ayotte, and Simon Woo helped with interpretations of the results. Yann Ayotte and Sacha Larda implemented some of the experiments used herein. Yann Ayotte mined the data and stripped it from any proprietary information, packaged it and deposited it in a public depository.

## Abstract

The free-state solution behaviors of drugs profoundly affect their properties. Therefore, it is critical to properly evaluate a drug's unique multiphase equilibrium when in an aqueous environment, which can comprise lone molecules, self-associating aggregate states and solid phases. To date, the full range of nano-entities that drugs can adopt has been a largely unexplored phenomenon. This protocol describes how to monitor the solution behavior of drugs, revealing the nano-entities formed as a result of self-associations. The procedure begins with a simple NMR  $^1\text{H}$  assay, and depending on the observations, subsequent NMR dilution, NMR T2-CPMG (spin-spin relaxation Carr-Purcell-Meiboom-Gill) and NMR detergent assays are used to distinguish between the existence of fast-tumbling lone drug molecules, small drug aggregates and slow-tumbling colloids. Three orthogonal techniques (dynamic light scattering, transmission electron microscopy and confocal laser scanning microscopy) are also described that can be used to further characterize any large colloids. The protocol can take a non-specialist between minutes to a few hours; thus, libraries of compounds can be evaluated within days.

## Introduction

A clear understanding of the behavior of compounds in aqueous solution is central to the rational design and development of pharmaceutical agents<sup>1,2,3,4</sup>. The behavior of compound solutions can be complex and involve mixtures of soluble, aggregate and precipitate forms (Fig. 1). Each compound adopts its own fingerprint multi-phase equilibrium in solution that is highly dependent on environmental conditions, such as concentration, buffer, salt, pH, temperature, metals, proteins and the presence of other molecular entities<sup>1,2,3,4,5</sup>. Knowledge of the relative solubility of a compound is therefore important because the relative solubility of a compound can affect its activity in chemical and biological assays as well as in vivo. Unfortunately, a lack of appropriate detection technologies has hindered acquisition of information about a compound's multi-phase equilibrium. As a consequence, the solution behavior of compounds under aqueous conditions remains poorly understood and is largely undetected.



**Fig. 1: Drugs exist in unique multi-phase equilibria in solution. They can range from lone molecules (left) to a solid (right), with intermediate states (center) of nano-entities that can exhibit many shapes and sizes.**

The aim of this protocol is to enable expert and non-expert researchers to monitor the solution behaviors of their compounds, with an emphasis on monitoring the aggregate phases. It includes detailed procedures for sample preparation, data acquisition, interpretation and how to use the information obtained in decision-making during the drug-discovery pipeline. The workflow is based on years of experience of the corresponding author in the pharmaceutical industry and involves the judicious use of various equipment and assays that, as an ensemble, present the best potential means for exposing the full range of drug nano-entity types and sizes that can exist. For example, we have previously used the components of this protocol in references <sup>1,6,7,8,9</sup>.

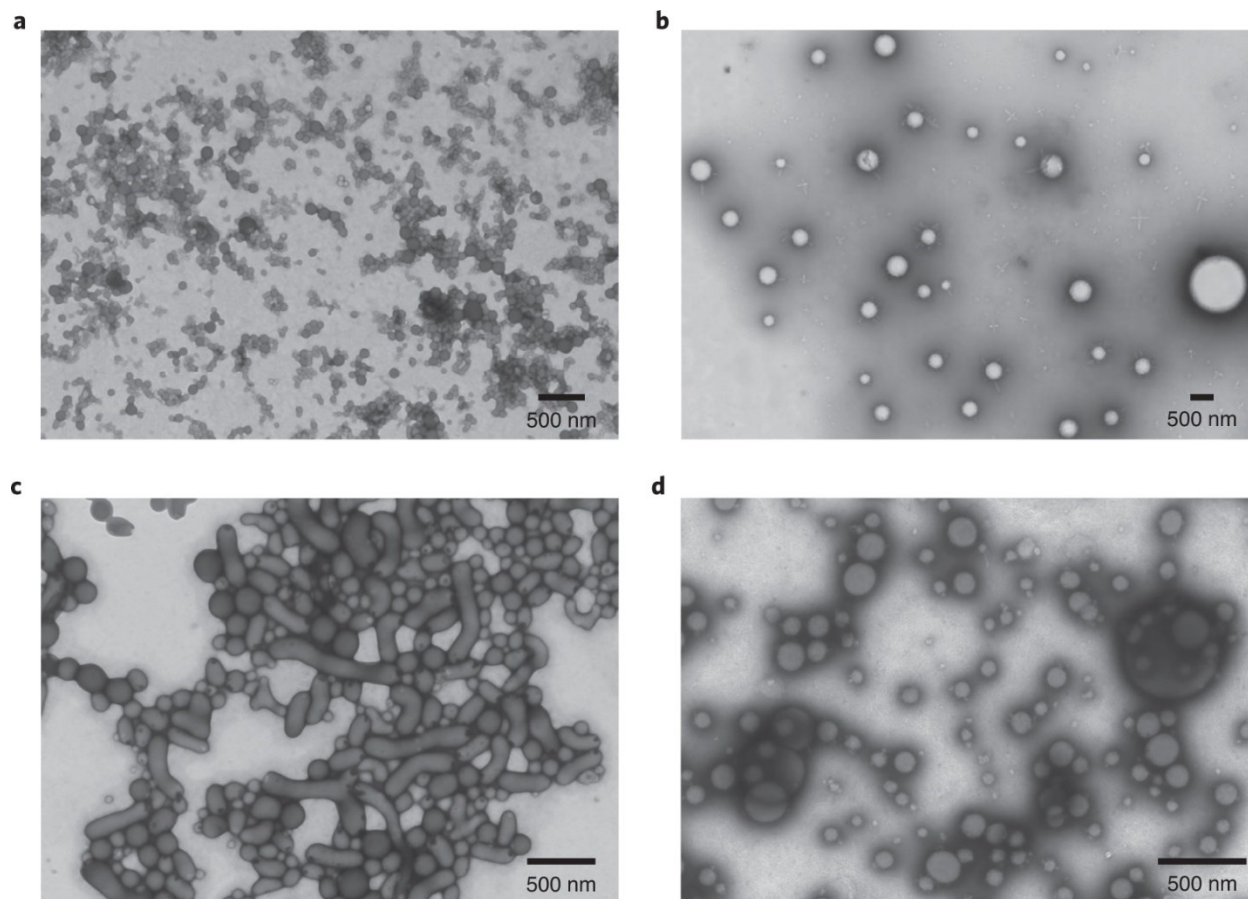
## Rationale for the development of the protocol

### Existence of compound aggregates (nano-entities)

All compounds naturally adopt a multi-phase equilibrium in solution. From a practical point of view, this property can be categorized as a three-phase equilibrium, ranging from single lone-tumbling molecules to insoluble solid precipitate with an intermediate array of soluble self-associated nano-entities or colloids. This is illustrated in Fig. 1, which also shows that distinctions can be made on the basis of molecular sizes. The term colloidal aggregates is often used to describe particles with



sizes extending beyond the nanoscale, whereas nano-entities refer to soluble aggregates having sizes between 1 and 1,000 nm. In general, compounds that behave as single lone-tumbling drug molecules fall within a 1-nm range, whereas amorphous solids can be  $\geq 1 \mu\text{m}$ . Drugs can also adopt self-assembled intermediate nano-entities such as dimers and small multimers sized  $\sim 1\text{--}10 \text{ nm}$  (Fig. 1). NMR spectroscopy is ideally suited for detection of 1–10-nm entities. In contrast, some nano-entities are much larger and thus visible via transmission electron microscopy (TEM) as shown in Fig. 2. For example, the images in Fig. 2 show that an anti-leprosy drug (clofazimine), two anti-cancer drugs (lapatinib and sorafenib) and curcumin form self-assemblies ranging from hundreds of nanometers to several micrometers in size. Although the exact molecular architecture and environmental parameters that dictate these equilibria have yet to be deciphered, it is clear that environmental conditions have an impact. For example, changes in colloid features can be observed when different media are used, and smaller aggregates can experience profound equilibrium shifts upon exposure to different buffer conditions<sup>1,6</sup>.



**Fig. 2: The presence of large drug colloidal aggregates can be visualized by TEM.**

**A–d, TEM images show aggregates of clofazimine (a), curcumin (b), sorafenib (c) and lapatinib (d). Compounds were incubated at 100  $\mu$ M (except for lapatinib, which was tested at 50  $\mu$ M) in DMEM in the presence of 5% (vol/vol) FBS.**

### **Importance of the characterization of nano-entities for decision making during drug discovery and development**

To date, the pharmaceutical industry has undertaken little characterization of multi-phase characteristics of compounds or correlation of these characteristics with function. The initiatives that have been undertaken have demonstrated that these properties have considerable impact at various stages of drug discovery and development. For example, compound aggregates can act as pan-assay interference-like entities and have been implicated in up to 85–95% of artifacts in early high-throughput screens (HTSs)<sup>2,4</sup>. These hits appear because of the formation of large drug colloids that can bind to and adsorb protein macromolecules in a non-specific manner, leading to changes in dynamics or partial denaturation of the protein. During unpublished industrial research projects, we frequently observed nano-entities when evaluating hits from HTSs and virtual and biophysics screens. These types of hits are often non-stoichiometric and/or non-specific binders.

The impact on decision making at the early lead identification stages of a project when searching for new starting-chemical matter can be profound. When evaluating hits from an HTS, counter-screens can be implemented by adding detergents that break up compound colloids<sup>1,3</sup>. Significant changes in activity can then be construed as undesirable activity, induced by hits that form colloidal aggregates. Deprioritizing such hits could avoid issues and downstream waste in productivity. Implementing well-designed counter-screens also enables detection of promiscuous hits.

The impact of nano-entities on fragment-based lead discovery can also be profound. Fragment screens usually require high compound concentrations and often involve pools of multiple compounds all at high concentrations. These conditions tend to shift equilibria toward the aggregation and solid phases. Thus, it would be prudent to implement an aggregation assay to first curate libraries of free fragments in buffer<sup>7</sup>. Furthermore, given findings that compound aggregation is also buffer dependent, one should consider running screens and aggregation tests on the free state of compounds in the same buffer used for biological assays. We suggest deprioritizing hits that exhibit aggregation at this stage, given the difficulties in distinguishing between stoichiometric-specific hits and aggregate-binding hits.

The impacts at the lead optimization stage can also be significant. Colloids have been found to result in false-negative activity in cell-culture assays<sup>5,10</sup>. In addition, inaccurate results from other biological assays and biophysics affinity measurements are likely, given that it is hard to determine a compound's solubility. Structure-activity relationship conclusions can also be skewed if the influence of aggregation is not taken into account. Distinct variations in aggregation can occur even within a series of closely related compounds<sup>7</sup>. Thus, it is prudent to continuously monitor and evaluate compounds' free-state solution behaviors throughout the lead optimization stages. When aggregates are observed at such a stage, it would probably be worthwhile to determine if stoichiometric single-molecule binding also exists. Aggregation tendencies can then be monitored as the series progresses, and the information can be incorporated in decision making. For example, we have shown that aggregation behavior can be minimized via minor chemical modifications<sup>1</sup>.

Monitoring aggregation tendencies is also recommended for compound series that are approaching selection for clinical studies. Compound aggregation has been implicated in undesirable off-target and promiscuous inhibition<sup>9</sup>, toxicity<sup>1,6</sup>, altered pharmacokinetics<sup>11,12</sup> and immune responses (unpublished data). To minimize these undesirable properties, we have previously used NMR aggregation assays to prioritize non-aggregating compounds for pre-clinical testing<sup>1</sup>. Nevertheless, some drug self-assemblies have also been associated with favorable properties such as conferring unusually high drug bioavailability<sup>13,14</sup>. Thus, the decision to prioritize or deprioritize aggregation will depend on the desired properties required for the drug-discovery program. However, it is likely that the examples of aggregators exhibiting desirable properties might be the exception rather than the rule. Either way, it is becoming clear that the solution behavior of compounds is important; thus, detection technologies are needed to better understand the relationships between these multi-phase solution behaviors and pharmacological properties.

### **Limitations in free-state detection and monitoring methods**

The most widely used methods to detect aggregates in solution are dynamic light scattering (DLS) and electron microscopy<sup>5,15</sup>. However, both are limited to the detection of large (>10 nm) nano-entities and colloids<sup>16,17</sup> and are thus unable to detect the full range of self-assemblies that can exist in solution. DLS can also be inadequate in the case of inhomogeneous (polydisperse) samples, whereas electron microscopy usually has relatively limited throughput. Although there is no single scientific instrument that can directly detect the full range of nano-entity sizes and types that exist, NMR is probably the most encompassing option. Fortunately, spectrometers are

widely available at most research institutes; thus, NMR is usually the most practical technology available that is also sensitive to the widest range of behaviors of compounds in solution. We introduced several NMR aggregation assays<sup>1,6,7,8</sup>, and these are described within this protocol. However, for optimal characterization, interested scientists must first learn and understand the pros and cons of various instruments and techniques, which can be overwhelming because of the challenges in acquiring and interpreting datasets within the context of the peculiarities of nano-entities and colloids.

### Comparison of methods: advantages and limitations

Table 1 provides a comparison of the advantages and limitations of the methods used in this protocol. Given this, the protocol described here attempts to establish a pragmatic approach that capitalizes on the advantages and mitigates the limitations of each technique.

**Table 1: Comparison of techniques for detecting drug aggregates**

Methods	Advantages	Limitations
NMR assays	<ul style="list-style-type: none"> <li>Provides atomic-level details of compounds</li> <li>Best methods for direct detection of small- to medium-sized aggregates</li> <li>Simple methods to implement</li> <li>Can be used with mixtures of aggregates</li> <li>Both qualitative and quantitative methods</li> <li>Amenable to a wide variety of buffers/media</li> <li>Amenable to HTSs</li> <li>Instruments widely available at institutions</li> <li>Can detect CACs</li> </ul>	<ul style="list-style-type: none"> <li>Large aggregates are detected indirectly via changes in relaxation properties or by changes in measured concentrations via addition of detergents or other additives</li> <li>Multiple samples required for NMR dilution assay</li> <li>Aggregate sizes can usually only be determined qualitatively</li> <li>Relatively insensitive technique and thus requires relatively high concentrations</li> </ul>
DLS assay	<ul style="list-style-type: none"> <li>Can detect large aggregates and determine sizes</li> <li>Can detect CACs</li> <li>Amenable to high throughput</li> <li>Amenable to a wide variety of buffers</li> </ul>	<ul style="list-style-type: none"> <li>Cannot detect small-to-medium-sized aggregates</li> <li>Not compatible with mixtures of aggregates (polydisperse samples)</li> <li>Many potential artifacts make detection and analyses difficult or confusing</li> </ul>

Methods	Advantages	Limitations
TEM assay	Can visualize large aggregates and determine sizes Can detect mixtures of large aggregates Amenable to a wide variety of buffers/media	Cannot detect small-to-medium-sized aggregates Can require high concentrations Visualization usually requires the addition of negative stains Not amenable for high throughput Size of aggregates may be underestimated compared to other techniques because of the need to dehydrate
CLSM assay	Can visualize large aggregates in cells Possible to observe distribution of aggregates in cells	Preparation is time consuming Limited to fluorescent compounds Cannot detect small-to-medium-sized aggregates If aggregates/compounds do not cross cell membranes, they will get washed away during the fixation steps Addition of FBS may sequester some/all aggregates, leading to no fluorescence being observed inside the cells (false negatives)

**CAC, critical aggregate concentration; CLSM, confocal laser scanning microscopy.**

Perhaps the most versatile and widely encompassing technique described in Table 1 is NMR spectroscopy. The NMR assays best allow one to monitor the solution behavior of compounds and their multi-phase equilibria. Lone molecules, as well as small- and medium-sized aggregates, can be directly monitored, whereas large colloids can be exposed only upon breakup via the addition of detergents. The presence of the latter can also be indirectly inferred through quantification (e.g., by using the electronic reference to access in vivo concentrations (ERETIC) method)<sup>18</sup>, the acquisition of spectra acquired in dimethyl sulfoxide (DMSO)-d<sub>6</sub> solvent or by difference in spin-spin relaxation Carr-Purcell-Meiboom-Gill (T2-CPMG).

Other traditional techniques used in this protocol are limited to the detection of medium to large aggregates and are insensitive to small aggregates and lone molecules. For example, DLS has been used extensively to study colloids<sup>17,19,20,21</sup>; thus, it is a great method for detecting large aggregates and for determining sizes and critical aggregation concentrations (CACs). Another added benefit is that DLS is also amenable to high-throughput screening. However, DLS cannot detect small-to-medium-range aggregates and has trouble resolving mixtures of aggregates of roughly equivalent size. Furthermore, it is prone to potential artifacts, which can make detection and analysis difficult or confusing.

TEM can be useful for visualizing large aggregates and for determining sizes<sup>6,22,23</sup>. It can detect mixtures of large aggregates and could be used to determine CACs. In some cases, it is also

amenable to a wide variety of buffers and media. However, it cannot detect small-to-medium-range aggregates, may require high concentrations and is not amenable to high-throughput screening. It is also noteworthy that visualization usually requires the addition of negative stains, which may affect compound multi-phase equilibria. In addition, the size of aggregates can be lower as a result of dehydration.

Confocal laser scanning microscopy (CLSM) is an excellent method to visualize large aggregates in cells along with their effects on cells<sup>6,22</sup>. However, sample preparations are lengthy, analyses are limited to fluorescent compounds and it is less sensitive to small-to-medium-sized aggregates. Moreover, only compounds that are able to cross the cell membrane, either as monomers or as aggregates, will be observed by using this method.

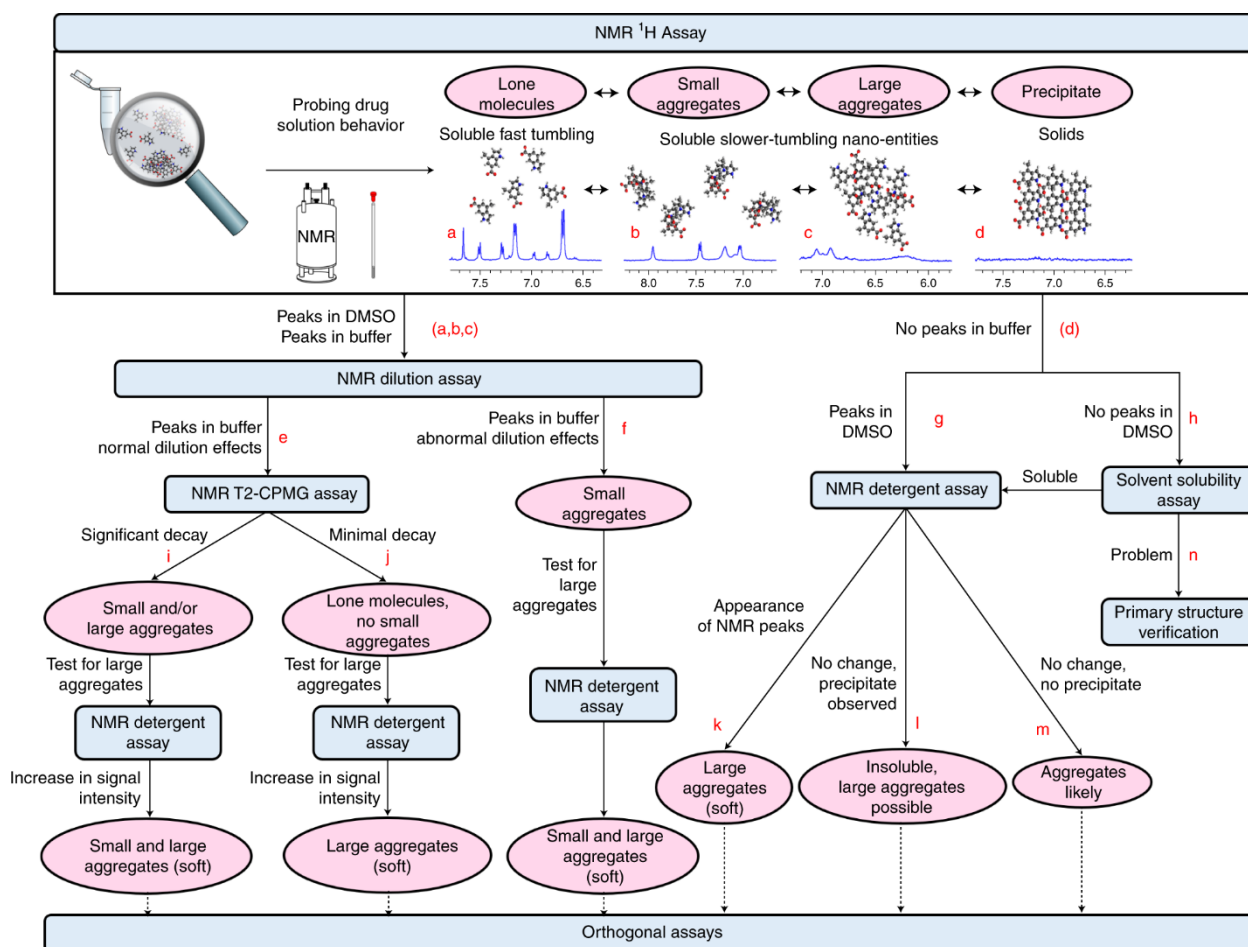
Other techniques that are not detailed in this protocol, such as aggregation-induced emission<sup>24,25</sup>, nanoparticle tracking analysis (NTA)<sup>26</sup>, nephelometry<sup>27</sup>, ultracentrifugation<sup>11</sup>, small-angle X-ray scattering<sup>2</sup> and other NMR methods (saturation-transfer difference<sup>28,29</sup>, diffusion ordered spectroscopy<sup>7,8</sup> and nuclear Overhauser effect spectroscopy<sup>1</sup>) can potentially be used to study solution behaviors of compounds. We chose not to describe the use of these assays because the current scope of these techniques is limited. Aggregation-induced emission relies on the induction of chromophore fluorescence and emission upon the formation of aggregates, which limits its applicability to only large, highly aromatic compounds<sup>24,25</sup>. NTA has been used in the tracking of colloids<sup>26</sup> and to track aggregates; however, we observed that the aggregates seemed to have a tendency to stick to the capillary. Similar to DLS, NTA is sensitive to contaminants<sup>26</sup>. Nephelometry has been previously used to determine compound solubility<sup>27</sup>; however, there is no mention in the literature on use to detect drug aggregation. Ultracentrifugation and small-angle X-ray scattering represent interesting approaches, but instrumentation is not always readily available. Many other NMR methods exist, but they are not as straightforward to set up and interpret, can be more time consuming and may not be as sensitive as the methods suggested here. Finally, one can also refer to databases that predict whether compounds have a tendency to aggregate<sup>30,31</sup>. However, these databases were constructed by using mainly DLS data and thus have inherent limitations.

Besides the technological limitations, there are multiple 'gaps' that continue to impede pharmaceutical scientists from effectively pursuing and detecting the existence of drug aggregates. Although some impacts are described here, we have not covered all issues and are therefore currently preparing a more encompassing evaluation in a review article. Perhaps the most daunting gap, and the easiest to overcome, is the lack of appreciation that the phenomenon

of drug aggregation can have serious impacts at all stages of drug discovery and development. Another important gap consists of the disconnect in pharmaceutical workflows and the expertise of personnel. For example, medicinal chemists synthesize new compounds and characterize them (at the atomic level by NMR) in organic solvents before delivering the powder form of these compounds (or stocks in organic solvent) to biochemists, who dissolve them in aqueous buffer and perform assays at the macroscopic level. Because no one is responsible for characterizing the compounds in water at the atomic level, important aspects of their solution behaviors can go undetected<sup>8</sup>.

## Overview of the procedure

Figure 3 provides an overview of the whole protocol. The protocol was designed to allow a scientist to first easily comprehend the solution behavior of a compound (or library of compounds) by using the quick NMR <sup>1</sup>H assay and NMR dilution assay. Depending on the results, other optional assays can then be undertaken to further characterize the multi-phase equilibrium and nano-entities present.



**Fig. 3: Overview of the protocol to probe drug solution behavior.**

**The protocol consists of the following parts to be used depending on each result: NMR  $^1\text{H}$  assay, NMR dilution assay, NMR T2-CPMG assay, NMR detergent assay and orthogonal assays. The red lettering indicates possible scenarios along the various stages of the protocol, and further details regarding the interpretation are given in the main text.**

The protocol begins with an NMR  $^1\text{H}$  assay in which a simple  $^1\text{H}$  NMR spectrum of the compound in aqueous buffer at 200  $\mu\text{M}$  nominal concentration is acquired. In parallel, the sample should be visually inspected to assess solubility (a clear sample suggests adequate solubility, whereas a turbid sample and/or formation of precipitate suggests limited solubility). The top of Fig. 3 provides examples of typical results seen for solutions behaving as lone molecules (a), small aggregates (b), large aggregates (c) and precipitate (d). The shape of the  $^1\text{H}$  NMR resonances of a compound depends on the tumbling rate of the species present in solution and on exchange between multiple states/entities. In scenario a in Fig. 3, visibly sharp resonances are seen as a result of lone molecules that tumble rapidly in solution. In contrast, larger aggregated entities tumble more slowly, resulting in broader resonances (scenario b in Fig. 3), or have resonances that are so broad that they appear to not exist (scenarios c and d in Fig. 3). If the observed NMR resonances are similar to scenarios a or b, then the concentration in solution can be determined (see Procedure below).

As a control experiment, the acquisition of a  $^1\text{H}$  NMR spectrum of the compound at the same 200  $\mu\text{M}$  nominal concentration in  $\text{DMSO-d}_6$  solvent is also recommended. This has several purposes. Assuming that most drug-like compounds are more soluble and are expected to aggregate less in DMSO solvent than in buffer, one can compare the observed concentration of a compound between the two solvents by using an external reference<sup>18</sup>. This allows the quantitative determination of the total concentration of lone molecules and/or small aggregates. Note that appreciable differences are often observed between nominal and observed concentrations of a compound in spectra from  $\text{DMSO-d}_6$  solvent because of powder-weighing issues and solid sample adducts (e.g., salt forms and hydrated powder).

If NMR resonances are observed for the compound in both buffer and DMSO, then the NMR dilution assay<sup>1,8</sup> can be undertaken along pathways e or f (Fig. 3). For this, the acquisition of a  $^1\text{H}$  NMR spectrum of the compound at lower concentration (50  $\mu\text{M}$  in buffer) is compared with the 200  $\mu\text{M}$  spectrum in buffer. If abnormal dilution effects are observed, this confirms that small aggregates are present (Fig. 3, scenario f), whereas if only normal dilution effects are detected, then it is likely that only lone molecules are present, with no small aggregates (Fig. 3, scenario



e). However, it is prudent to also run the NMR T2-CPMG assay<sup>7,32</sup> on the 200  $\mu\text{M}$  spectrum in buffer, given that some nano-entities have been reported to be insensitive to the dilution assay but are sensitive to the T2-CPMG assay. Thus, if significant resonance decay is observed in the NMR T2-CPMG assay, then small and/or large aggregates exist (Fig. 3, scenario i), whereas if only minimal decay is noted, then the sample has lone molecules and no small aggregates (Fig. 3, scenario j).

If the actual measured concentration of the compound in buffer is significantly less than that in DMSO- $\text{d}_6$  solvent, then it is prudent to test if large NMR-invisible colloidal aggregates also exist. In general, aggregates are sensitive to detergents, such that they show detergent reversibility<sup>19,33</sup>. There have been a few assays developed for testing for aggregates on the basis of this property<sup>3,34</sup>. Running the NMR detergent assay tests for the presence of aggregates. This assay comprises the addition of a series of detergents (examples are given in Supplementary Table 1) to the samples containing the compound. The rationale is that because NMR resonances of large colloids are often very broad or too broad to be visible by NMR, adding detergent can effectively dissociate the colloids into smaller, faster-tumbling pieces. The subsequent observation of any increases in intensity of compound resonances thus confirms the pre-existence of large aggregates, which can be quantified by using the ensemble of NMR data. These colloids are referred to here as 'soft aggregates' (e.g., candesartan; vide infra), whereas large colloids that are insensitive to detergents are referred to here as 'hard aggregates' (e.g., lapatinib; vide infra).

The NMR detergent assay should be run if no peaks are observed after acquisition of the simple  $^1\text{H}$  NMR spectrum for the compound in buffer but resonances are observed for the compound in DMSO- $\text{d}_6$  (Fig. 3, scenario G). This assay should also be run if compound resonances begin to appear at the 50- $\mu\text{M}$  concentration in buffer. The appearance of compound resonances after the addition of detergent suggests the existence of large soft aggregates (Fig. 3, scenario K). If compound resonances do not appear (and there was precipitate), it is likely that the compound is insoluble in buffer, and large hard aggregates could be present (Fig. 3, scenario L). In contrast, if compound resonances do not appear (and there was no precipitate), then it is possible that the compound exists as a large hard aggregate that is insensitive to the detergents used in the assay (Fig. 3, scenario M). Such results should be verified by an orthogonal method such as electron microscopy or DLS.

If no resonances were observed when the compound was placed in DMSO or buffer, this could indicate that the compound is insoluble (Fig. 3, scenario H). In this scenario, other solvents should be tested by using the solvent solubility assay. Once a solvent is identified for which the compound

is soluble, then the NMR detergent assay should be run for the compound in buffer. However, if the compound is not soluble in any of the solvents tested, then we recommend performing a primary structure verification or abandoning the compound. It is difficult to obtain reproducible results from biological assays that attempt to evaluate insoluble compounds.

The final part of the procedure uses orthogonal assays to confirm whether large colloids are present. We describe how to use DLS, TEM and CLSM. However, alternative assays can be used. By combining all the assays that we describe here, a comprehensive analysis of compounds and their behaviors in aqueous solution can be obtained.

### **Development of the protocol**

The initial development of the protocol began in 1996 as a result of a simple observation. The corresponding author had dissolved a drug-like compound in PBS buffer at 200  $\mu$ M and noted that the sample was clear (no precipitate); he then proceeded to take a  $^1\text{H}$  NMR spectrum. To his surprise, there were no observable NMR resonances for the compound as it resembled the spectrum shown in Fig. 3d. Given that the subsequent  $^1\text{H}$  NMR spectrum of the stock in DMSO- $d_6$  solvent confirmed the presence of the compound, the only reasonable conclusion was that the compound formed soluble self-associating aggregates that were very large in size in the buffer. Because there were no reports of this phenomenon at that time, further investigations were undertaken. Subsequent  $^1\text{H}$  NMR spectra on this series of hepatitis C virus polymerase inhibitors showed diverse spectral observations that resembled the spectra A, B, C and D in Fig. 3, demonstrating that this series exhibited a wide range of aggregate types that varied depending on minor differences in primary structure (structure nano-entity relationships)<sup>1</sup>. Thus, in the process of undertaking this study, the first components of this protocol were developed.

Further additions to the procedure were introduced subsequently, and the whole procedure has been successfully applied to industry drug-discovery projects. We explored orthogonal methods routinely used in drug discovery such as DLS, diffusion ordered spectroscopy, nuclear Overhauser effect spectroscopy, etc.<sup>35</sup>, testing their ability to enable characterization of aggregates. Only an ensemble of strategies was found to reliably detect the full range of nano-entity sizes and types<sup>7,8</sup>. Concurrent to these efforts, literature reports described large drug colloidal aggregates visualized by using DLS as the main technology<sup>19,20</sup>. As noted below, DLS has proven to be a valuable technique and is applied here as a useful orthogonal method that uses its advantages while addressing its limitations (*vide infra*). Over time, we developed and published the practical NMR-based methods (NMR  $^1\text{H}$  assay, NMR dilution assay and NMR detergent assay)<sup>1,6,8</sup>, which are amenable for routine use by non-specialists and medicinal

chemists. These NMR-based methods also showed the presence of small aggregates for the first time in the context of drug discovery. More recently, we implemented a new aggregation assay based on the observation of relaxation in T2-CPMG experiments<sup>7</sup>, and we have incrementally improved parts of this NMR T2-CPMG assay (e.g., quantification) and exploited orthogonal assays (e.g., TEM, CLSM and DLS)<sup>3,6,8</sup>.

The ensemble of tools mentioned above is already having an impact and revealing drug aggregates, but until the publication of this protocol, no clear 'best practices protocol' was available to guide scientists monitoring the solution behavior of their compounds. Thus, this protocol attempts to combine the details we and others have learned regarding nano-entities, to provide a protocol for best practice.

### **Application of the protocol**

The procedure described here was designed to provide new users with a series of practical and efficient strategies to address their experimental problems. Users can use the protocol as is, in which case the NMR <sup>1</sup>H assay should be run first to qualitatively assess the predominant state adopted by the compound of interest. On the basis of results obtained, specific subsequent assays and pathways are suggested in Fig. 3 that enable more precise conclusions and quantifications to be obtained.

The general nature of the protocol renders it amenable to many different applications along the drug-discovery pipeline. The protocol can easily be further adapted to address specific needs for various applications. For example, alternative buffers can be used in place of the phosphate buffer used in the current procedure if these are more relevant to any subsequent biological assays. However, a deviation from phosphate buffer may require the use of deuterated buffer components for NMR assays. Individual assays can be run in place of the ensemble of assays shown in Fig. 3.

As another example, if a higher-throughput assay is required, for library curation or triaging of many hits from high-throughput or virtual screens, each assay can be modified. For the NMR <sup>1</sup>H assay, compounds can be screened in buffer only. Spectra of compounds in DMSO-d<sub>6</sub> can be run later if needed. For the NMR dilution assay, data can be acquired for samples only at 200- and 50- $\mu$ M concentrations. Follow-up studies can then use the full set of dilution experiments for selected compounds. For the T2-CPMG assay, one could run only the 1- and 800-ms experiments. If relaxation rates are desired for selected compounds, follow-up experiments can use the full set of delays. For the NMR detergent assay, one could use only one detergent as an

initial screen (e.g., Tween). Supplementary Table 2 gives guidance on procedure adaptations and timing for sample sets ranging from 1 to >50 compounds.

Other modifications could be made to better address the question at hand. For example, the NMR assays can be run at lower concentrations (e.g., compounds at 10  $\mu\text{M}$ ) if such concentrations are more relevant to other experiments (such as cell culture). Moreover, researchers who are working with optimized compounds that have more limited solubility may want to decrease the concentrations if the dilution range yields no signal. For this, an increase in the number of NMR scans would be recommended. Alternatively, higher millimolar concentrations may be used to mimic conditions in the stomach for orally administered drugs or as needed for X-ray crystallography studies. One could also run experiments in other pharmacologically relevant buffers<sup>1</sup>, formulations and conditions (e.g., temperatures). One may also want to solely use orthogonal methods. For example, if screening only for the presence of large colloids, a DLS screen may be most appropriate.

### **Expertise needed to implement the protocol**

The protocol described here was designed to enable non-experts to prepare samples, to acquire data and to make their own interpretations. The main focus is on the NMR <sup>1</sup>H assay, the NMR dilution assay, the NMR T2-CPMG assay and the NMR detergent assay, all of which can be easily performed by graduate students and even undergraduate students. NMR spectrometers are readily available in most graduate and undergraduate institutions. No special experimental NMR setup is required to acquire a water-suppressed <sup>1</sup>H spectrum; however, it should be possible to modify some parameters as described in Equipment setup. The T2-CPMG experimental parameters are often configured on most modern NMR spectrometers. If not, they can be easily implemented by an in-house NMR technician or the instrument manufacturer. The ERETIC method can also be implemented for concentration determination. NMR data processing can be done directly on the spectrometer or through software that can be downloaded from the Bruker BioSpin website (<https://www.bruker.com/en/products-and-solutions/mr/nmr-software/topspin.html>).

Instruments required for the orthogonal assays (DLS, TEM and CLSM) can often be found at most graduate institutions. As with NMR, institutions that have these instruments often also have in-house technicians or users who can help with experimental setup and use.

## Materials

### Reagents

- *Compounds of interest.* We show results from valsartan (Sigma-Aldrich, cat. No. SML0142), methylene blue (Sigma-Aldrich, cat. No. 66720), candesartan cilexetil (Sigma-Aldrich, cat. No. SML0245) and lapatinib (Sigma-Aldrich, cat. No. SML2259) as example data.

### Required for the NMR assays only

- Deuterated NMR solvents, including DMSO-d<sub>6</sub> (Cambridge Isotope Laboratories, cat. No. DLM-10-10×1) and acetonitrile-d<sub>3</sub> (€-d<sub>3</sub>) (CDN Isotopes, cat. No. 2206-26-0)

## CRITICAL

Solvents should be handled under a fume hood.

- Deuterium oxide (D<sub>2</sub>O) (Cambridge Isotope Laboratories, cat. No. DLM-4-100)
- Tween 80 (Fisher Scientific, cat. No. BP338-500)
- Tween 20 (Fisher Scientific, cat. No. BP337-100)
- Nonidet P-40 (Roche, cat. No. 11754599001), Nonidet P-40 Substitute solution (Sigma-Aldrich, cat. No. 98379) or IGEPAL CA-630 (Sigma-Aldrich, cat. No. I3021)
- Triton X-100 (Sigma-Aldrich, cat. No. X100)
- SDS (Sigma-Aldrich, cat. No. 8.22050)
- 3-((3-Cholamidopropyl) dimethylammonium)-1 propane sulfonate detergent (Fisher Scientific, cat. No. PI28299)
- Milli-Q or HPLC-grade water
- Sodium phosphate dibasic anhydrous (Fisher Scientific, cat. No. S374-500)
- Sodium phosphate monobasic anhydrous (Fisher Scientific, cat. No. BP329-500)
- HCl (Sigma-Aldrich, cat. No. 7647-01-0)

## CRITICAL

Concentrated HCl should be handled under a fume hood.

- NaOH (Supelco, cat. No. 1310-73-2)

- NaCl (Wisent Bioproducts, cat. No. 600-082-WG)

Required for the TEM (transmission electron microscopy) assay only

- Phosphotungstic acid (PTA) (MECALAB, cat. No. 4098)
- Milli-Q or HPLC-grade water
- NaOH solution, 1 N (Fisher Scientific, cat. No. SS2661)

Required for the CLSM assay only

- DRAQ-5 (Biostatus, cat. No. DR50200)
- Wheat germ agglutinin conjugate (WGA-555) (Invitrogen, cat. No. W32464)
- Glycerol (Wisent Bioproducts, cat. No. 800-040-LL)
- Paraformaldehyde (PFA) at 16% (vol/vol) (Fisher Scientific, cat. No. AA433689L)

### **CRITICAL**

PFA should be handled under a fume hood or in a well-ventilated area.

- Media required to culture cells (we use DMEM to culture RAW 264.7) (Gibco, cat. No. 11039-021)

### **CRITICAL**

The presence of serum in the medium may sequester aggregates in the medium, resulting in less fluorescence observed inside the cells.

- *Cells of interest.* We have used RAW 264.7 cells (macrophages) (Sigma, cat. No. 91062702; European Collection of Authenticated Cell Cultures, cat no. 91062702; RRID: CVCL\_0493), but other cells could be used

### **CAUTION**

The cell lines used in your research should be regularly checked to ensure that they are authentic and are not infected with mycoplasma.

- PBS (HyClone; GE Healthcare, cat. No. SH30256.01)
- Trypsin (Gibco, cat. No. 25200-056)

### **Equipment**

#### **Standard equipment required for all assays**

- Recommended protective equipment: laboratory coat, gloves and safety glasses

#### **Required for the NMR assays only**

- 400-MHz or higher-field NMR spectrometer (e.g., 600-MHz Bruker AV III NMR)
- Magnetic stir bar
- Beaker (500 ml or larger)
- Microcentrifuge tubes
- Laboratory microcentrifuge
- NMR tubes (3- or 5-mm sizes)
- Pasteur pipettes
- Bulbs
- pH meter
- Magnetic stir plate

#### **Required for the TEM assay only**

- Beaker (250 ml)
- Pasteur pipette
- Bulbs
- Magnetic stir plate
- TEM (a Hitachi H-7100 was used here)
- Parafilm
- Copper grid (200 mesh)

#### **Required for the CLSM assay only**

- Confocal microscope (a Zeiss CLSM-780 confocal microscope was used here)
- 250-ml beaker
- Glass stirring rod
- Biological safety cabinet
- Cell culture flasks (e.g., 25 cm<sup>2</sup>)

- Cell culture plate (24 wells)
- Microscope coverslips (Fisher Scientific, cat. No. 12-565-88)
- Microscope slide (Fisher Scientific, cat. No. 12-552)
- Vacuum
- 50-ml conical centrifuge tubes (Fisher Scientific, cat. No. 14-432-22)
- Task wipers
- Laboratory curved tweezers
- CO<sub>2</sub> incubator (Sanyo)

#### **Required for the DLS assay only**

- Zetasizer Nano (or another instrument in the Zetasizer range) (Malvern Panalytical)
- Cuvettes (Fisherbrand)
- 0.22- $\mu$ m filter (Millipore, cat. No. GSWP14250), 0.22- $\mu$ m bottle-top filter (Millipore, cat. No. SCGPS01RE) or 0.22- $\mu$ m syringe filter (Fisher Scientific, cat. No. 09-719G)

#### **Software**

- NMR data-processing software (e.g., Bruker's TopSpin or MestReNova)
- DLS analysis software (e.g., Zetasizer Nano)
- Confocal analysis software (e.g., ZEN)
- TEM analysis software (e.g., AMT)

#### **Reagent setup**

##### **Preparation of sodium phosphate buffer (500 ml)**

The sodium phosphate buffer is prepared as 50 mM phosphate and 100 mM NaCl at pH 7.4.

1. Place 400 ml of Milli-Q or HPLC-grade water in a beaker along with a magnetic stir bar.
2. Add 2.68 g of sodium phosphate dibasic anhydrous and 0.76 g of sodium phosphate monobasic anhydrous.
3. Add 10 ml of 5 M NaCl solution.
4. Add 50 ml of D<sub>2</sub>O (to achieve 10% of D<sub>2</sub>O for the total volume).



5. Adjust the pH to 7.4 by using HCl or NaOH with the aid of a pH meter.
6. Add water to bring the volume to 500 ml.
7. Filter the solution with a 0.22- $\mu$ m filter.
8. Store at  $-20\text{ }^{\circ}\text{C}$  for  $\leq 1$  month and monitor the pH before use.

### **CRITICAL**

Sodium phosphate dibasic heptahydrate and sodium phosphate monobasic monohydrate may be used, but the amounts must be modified to account for the difference in molecular weight.

### **Preparation of detergent at 10% (wt/vol) (1 ml)**

Examples of detergents we have used are given in Supplementary Table 1.

1. Weigh 0.1 mg of detergent.
2. Add 1 ml of the sodium phosphate buffer made previously.
3. Store at room temperature ( $15\text{--}25\text{ }^{\circ}\text{C}$ ) for  $\leq 1$  month and mix well before use.

### **Preparation of PTA at 3% (wt/vol) (100 ml)**

1. Weigh 3.0 g of phosphotungstic acid and add to a 250-ml beaker equipped with a magnetic stir bar.
2. Add 80 ml of Milli-Q or HPLC-grade water.
3. Adjust the pH to 6.0 by using a 1 N NaOH solution ( $\sim 8$  ml).
4. Add Milli-Q or HPLC-grade water to bring the volume to 100 ml.
5. Store in a refrigerator ( $4\text{ }^{\circ}\text{C}$ ) for  $\leq 6$  months.

### **Preparation of mounting media: 70% (vol/vol) glycerol (1 ml)**

1. Pipette 700  $\mu$ l of glycerol in an Eppendorf tube.
2. Add 300  $\mu$ l of the sodium phosphate buffer.
3. Shake well after closing the cap.
4. Store at room temperature for  $\leq 1$  month.

### **Preparation of 4% (wt/vol) PFA (25 ml)**

1. Add 6.25 ml of 16% (vol/vol) PFA to a 50-ml Falcon tube.

2. Add 18.75 ml of sodium phosphate buffer to the tube and mix well.
3. Store at 4 °C for  $\leq 1$  month.
4. Before usage, let the 4% (vol/vol) PFA solution warm to room temperature for  $\geq 2$  h.

## Equipment setup

### NMR assays

For the NMR  $^1\text{H}$  assay, NMR dilution assay and NMR detergent assay, NMR experiments can be run on a 400-MHz or higher-field instrument. The spectra shown in the figures were run on a 600-MHz Bruker Avance III spectrometer equipped with a helium cryoprobe. A standard Bruker  $^1\text{H}$  pulse sequence can be used. Here, we also used a sequence that implements water suppression via excitation sculpting (zgesgp)<sup>36</sup>. Spectra can be acquired with 32 scans for the NMR  $^1\text{H}$  assay and the NMR detergent assay, whereas 128 scans per spectrum are recommended for the NMR dilution assay. If one wishes to use quantification, then a relaxation delay of 10 s should be used to ensure sufficient relaxation of the aromatic spins before subsequent scans to allow reasonably accurate measurement of the solution concentration.

For the NMR T<sub>2</sub>-CPMG assay, the parameters used here are a modified version of the standard Bruker  $^1\text{H}$  sequence with excitation sculpting (zgesgp) and the addition of a CPMG pulse train after the initial 90-degree excitation pulse. For the results shown here, the total duration of each spin echo was fixed to 1 ms, whereas the number of echoes in the pulse train was varied according to the total time ( $T$ ). The number of scans for all spectra was 4.

### CRITICAL

If one decides to prepare samples with buffer containing 90% (vol/vol) H<sub>2</sub>O and 10% (vol/vol) D<sub>2</sub>O, then  $^1\text{H}$  experiments should use water suppression

### CRITICAL

If the NMR spectrometer is equipped with a cooled (5–6 °C) SampleJet sample changer, samples in DMSO-d<sub>6</sub> solvent will freeze (the melting point for DMSO is  $\sim 18$  °C), and the freeze–thaw cycle could have a negative effect on the NMR data obtained. In this case, it is possible to substitute DMSO-d<sub>6</sub> for a 50:50 (vol/vol) mixture of DMSO-d<sub>6</sub>/€-d<sub>3</sub>, which has a lower freezing point as long as the compound has sufficient solubility in the mixed solvent. Note that the use of a solvent mixture may require tuning of the lock.

## Procedure

## Preparation of a stock solution of a compound

Timing 12 min

### CRITICAL

This section describes how to prepare compound stock solutions at a concentration of 20 mM in DMSO-d<sub>6</sub>.

1. Weigh between 0.3 and 0.6 mg of the compound and place it in a microcentrifuge tube.
2. Use the following formula to calculate the volume of DMSO-d<sub>6</sub> to be added:

$$DMSO \text{ volume (mL)} = \frac{\text{Exact mass weighed (mg)}}{\text{MW of compound} \left(\frac{\text{mg}}{\text{mmol}}\right)} \cdot \frac{1 \times 10^6}{20 \text{ (mM)}}$$

3. Dissolve the compound with DMSO-d<sub>6</sub> in the microcentrifuge tube to give a 20 mM solution and mix. We recommend letting the DMSO stock sit at room temperature for ≥1–2 h to allow better dissolution. If particles are still visible after that time, sonication can be used to improve dissolution.

### PAUSE POINT

If compounds are sufficiently stable, DMSO stock can often be left to sit at room temperature in a closed tube for several hours or even a few days. We recommend freezing the stocks if the compounds are not going to be used during several days. If compound stability is limited, fresh stocks should ideally be prepared before each experiment.

### NMR <sup>1</sup>H assay

Timing 3 min (preparation)

### CRITICAL

Note that use of differing brands of NMR tubes may require slightly different volumes (because of potential differences in glass thickness). Moreover, different types of NMR probes might require the use of additional volumes. The suggested values here represent a general rule of thumb that should work with most spectrometers. Ensure that the appropriate solution volume is used. To have sufficient volume for proper lock and shim, we suggest using ~200 μl for 3-mm tubes and 600 μl for 5-mm tubes. In the following protocols, the volumes required for 5-mm NMR tubes are shown in parentheses.

### CRITICAL

This section of the procedure is summarized in Extended Data Fig. 1.

4. Prepare 200  $\mu\text{M}$  compound solutions in DMSO- $\text{d}_6$  and appropriate buffers as described in options A and B.

**(A) Assays in DMSO- $\text{d}_6$**

- (i) Add 198  $\mu\text{l}$  (594  $\mu\text{l}$ ) of DMSO- $\text{d}_6$  in a microcentrifuge tube.
- (ii) Add 2  $\mu\text{l}$  (6  $\mu\text{l}$ ) of 20 mM stock solution and mix.
- (iii) Transfer into an NMR tube with a Pasteur pipette and analyze.

**(B) Assays in buffer**

- (i) Add 198  $\mu\text{l}$  (594  $\mu\text{l}$ ) of buffer in a microcentrifuge tube.
- (ii) Add 2  $\mu\text{l}$  (6  $\mu\text{l}$ ) of 20 mM stock solution and mix.
- (iii) Transfer into an NMR tube with a Pasteur pipette and analyze.

## **NMR dilution assay**

Timing 27 min (preparation)

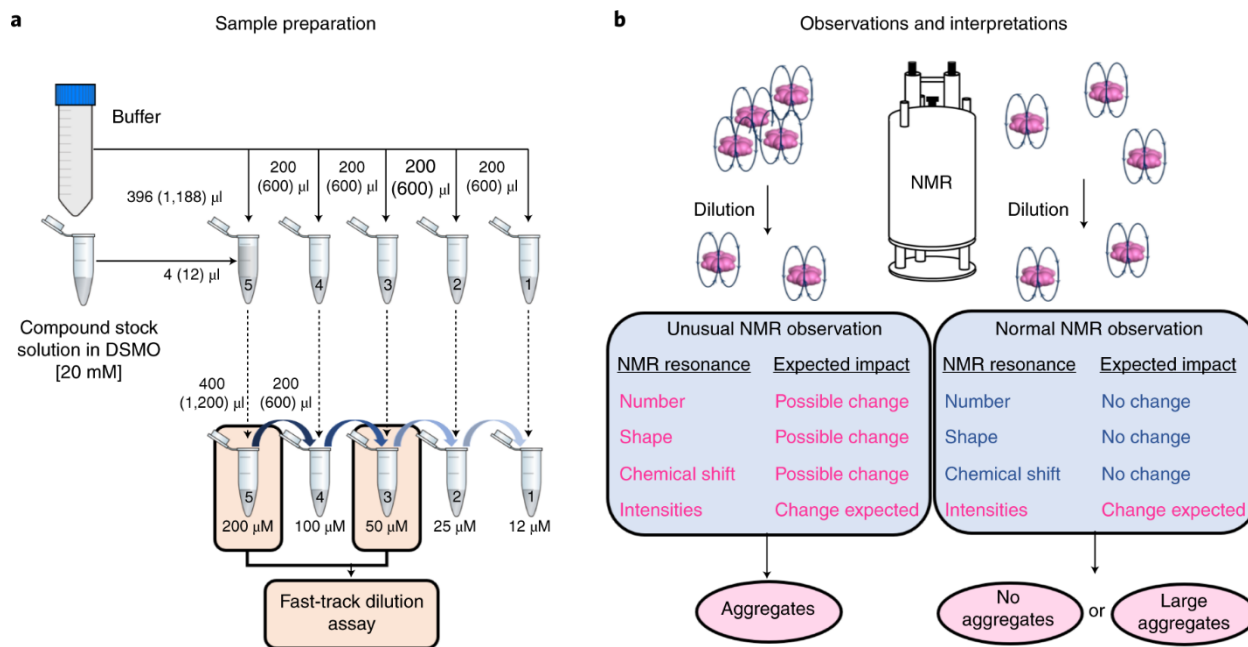
### **CRITICAL**

If undertaking the fast-track dilution strategy (see the beige highlighted segments in Fig. 4a), data acquisition is performed only on the 50  $\mu\text{M}$  and 200  $\mu\text{M}$  samples. Thus, in place of this section, 50 (150)  $\mu\text{l}$  of the 200  $\mu\text{M}$  sample can instead be diluted into 150 (450)  $\mu\text{l}$  of buffer to get the 50  $\mu\text{M}$  concentration point.

### **CRITICAL**

This section of the procedure is summarized in Fig. 4.

1. Add buffer into microcentrifugation tubes: 396 (1,188)  $\mu\text{l}$  in a tube labeled '#5' and 200 (600)  $\mu\text{l}$  each in tubes labeled '#1' to '#4'.



**Fig. 4: NMR dilution assay.**

**A, Two options for preparation of the samples: the ensemble of samples for the full dilution assay and the shorter fast-track strategy as highlighted in beige boxes. B, Interpretation of the results. Shown are volumes suggested for 3-mm NMR tubes, and volumes for 5-mm tubes are in parentheses.**

- Add 4 (12)  $\mu\text{l}$  of the 20 mM stock solution to tube #5. This tube now becomes the 200  $\mu\text{M}$  stock sample in buffer.
- Prepare the remaining samples by serial dilution by adding 200 (600)  $\mu\text{l}$  of the 200  $\mu\text{M}$  stock sample in buffer to tube #4 and mixing well to produce the 100  $\mu\text{M}$  sample.
- Continue serial dilutions by using 200 (600)  $\mu\text{l}$  from the previous sample to get 50, 25 and 12  $\mu\text{M}$  samples.
- Transfer each of the solutions from tubes #1 to #5 into separate NMR tubes and analyze.

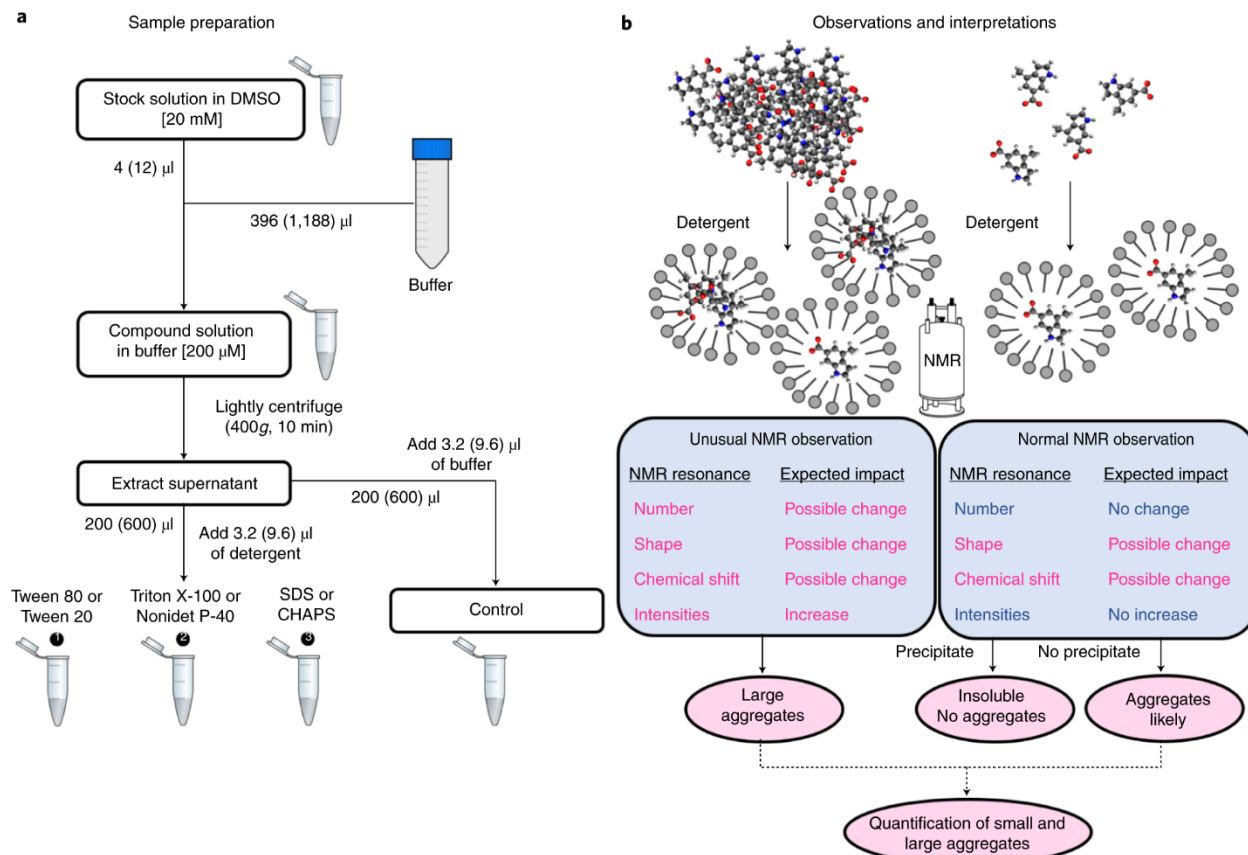
## NMR detergent assay

Timing 30 min (preparation)

### CRITICAL

This section of the procedure is summarized in Fig. 5.

1. Prepare a stock solution of compound at 200  $\mu\text{M}$  in buffer, allowing for 200 (600)  $\mu\text{l}$  for each detergent condition to be tested in addition to a control sample without detergent.



**Fig. 5: NMR detergent assay.**

**A, Preparation of the samples. B, Interpretation of the results.** Shown are volumes suggested for 3-mm NMR tubes, and volumes for 5-mm tubes are in parentheses. CHAPS, 3-((3-cholamidopropyl) dimethylammonium)-1 propane sulfonate.

1. Place 200 (600)  $\mu\text{l}$  in separate microcentrifuge tubes for each detergent.
2. Centrifuge at 400g for 10 min.
3. Take the supernatant and add 3.2 (9.6)  $\mu\text{l}$  of detergent stock solution at 10% (wt/vol) in 200 (600)  $\mu\text{l}$  of the solution. To the control tube, add 3.2 (9.6)  $\mu\text{l}$  of buffer to compensate the slight dilution from the addition of detergent.

### CRITICAL STEP

Different detergent concentrations can be used to cover ranges above and below their critical micelle concentration. The concentrations stated here serve as a practical starting point.

4. Transfer to NMR tubes and analyze.

#### **CRITICAL STEP**

Triton X-100, Nonidet P-40 and 3-((3-cholamidopropyl) dimethylammonium)-1 propane sulfonate have resonances in the aromatic region that may overlap with aromatic signals of the compound. If this is the case, one should analyze the signals that are not overlapping and/or change the detergent.

#### **CRITICAL STEP**

For aggregating compounds, one would expect an increase in intensity or appearance of new peaks. For non-aggregating compounds, one would expect no increased intensity or appearance of peaks; however, it is possible to see a decrease/broadening of the signal as a result of compound interaction with the detergent.

### **NMR T2-CPMG assay**

Timing 15 min (preparation)

#### **CRITICAL**

If this experiment follows the NMR  $^1\text{H}$  assay (Step 4B), there is no need to reprepare the 200  $\mu\text{M}$  sample. The 200  $\mu\text{M}$  sample can be reused to run the T2-CPMG experiments. In this scenario, proceed directly to Step 17.

#### **CRITICAL**

This section of the procedure is summarized in Extended Data Fig. 2.

15. Prepare a solution with compound at 200  $\mu\text{M}$  in buffer.
16. Transfer to an NMR tube.
17. Set up separate NMR T2-CPMG experiments with various delay times (1, 25, 50, 100, 200, 300, 500 and 800 ms), or under the fast-track strategy, acquire only two spectra with delay times of 1 and 800 ms.

## Solvent solubility assay

Timing 15–20 min

### CRITICAL

This section of the procedure is summarized in Extended Data Fig. 3.

18. Prepare a series of samples of the compound at 200  $\mu\text{M}$  in deuterated solvents ( $\text{C}_6\text{D}_6$ , chloroform- $d$ , etc.).

### CRITICAL STEP

The same stock of compound from Step 3 can be used to prepare 200  $\mu\text{M}$  samples as described in Step 4 instead of preparing a new stock.

### CRITICAL

Some solvents may not be compatible with the use of plastic tubes for sample preparation. Glass vials may be more appropriate in such cases.

19. Transfer samples to NMR tubes and visually observe whether the compound appears to be soluble.
20. Acquire a regular 1D  $^1\text{H}$  NMR spectrum of each sample.

## NMR data analysis by TopSpin

Timing 30 min to 2 h depending on the number of datasets

### CRITICAL

There are many options in TopSpin for data processing. The steps given below constitute a practical workflow; however, depending on the dataset, other autophasing algorithms may work better.

1. Load the appropriate data folder containing NMR data in the TopSpin software.
2. Add the first spectrum and process it by typing the following in the command line at the bottom of the graphics interface: `lb 1; efp; apk; abs`.
3. If the automatic phasing routine did not yield satisfactory results, adjust the spectrum phase as needed to achieve a straight baseline.



4. If the ERETIC method has been implemented on the NMR spectrometer, then use it to calculate the actual concentration of the compound in solvent or in buffer. To this end, integration of peaks in the aromatic region is usually simpler, but well-defined aliphatic resonances can also be used. However, aliphatic protons can exhibit relaxation times that could differ significantly from their aromatic counterparts. Therefore, the relaxation delays might require optimization to allow reliable estimation of concentration based on aliphatic resonances. If in doubt, a longer delay can be used at the expense of a longer acquisition time. Ensure that peaks chosen for quantitation are outside the chemical shift range affected by water suppression, if used.
5. In Integration mode, left-click on the desired integral to select the ERETIC command and calculate the observed concentration of the compound.

## **Orthogonal assays**

### **CRITICAL**

This section of the procedure is summarized in Extended Data Fig. 4.

26. Further analyze compound behavior as appropriate by following options A, B and/or C to analyze by DLS, TEM and/or CLSM, respectively.

#### **(A) DLS**

Timing 30–50 min depending on the number of samples to be evaluated

- (i) Prepare a solution of the compound at a concentration of 200  $\mu\text{M}$  in buffer (~1 ml, although smaller sizes of cuvette exist).

### **CRITICAL STEP**

It is recommended to filter the buffer before use with a 0.22- $\mu\text{m}$  filter to reduce the contamination risks, because small dust particles can significantly skew results. The solvent used to dissolve the compound should also ideally be filtered before use with a solvent-compatible 0.02- $\mu\text{m}$  filter. Preparation of samples under a laminar flow cabinet is also strongly suggested.

- (ii) Centrifuge the sample tubes (400g, 10 min).
- (iii) Transfer the solution supernatant to a cuvette for DLS measurements.

### **CRITICAL**

The use of a compressed air duster is recommended to clear cuvettes of possible particles before use.

- (iv) Use the Zetasizer software to perform the measurements, choosing the following parameters: sample name, temperature, cuvette type and number of measurements (a minimum of three experiments is recommended).
- (v) Place the cuvette in the Zetasizer Nano (or other DLS equipment used).
- (vi) Use the Zetasizer software (Malvern Panalytical) to analyze the data.

### **CRITICAL STEP**

Results can be viewed according to various functions. The Count Rate display allows monitoring of sample quality. The Correlation Function can be useful to identify potential problems with the sample. The default view is Intensity PSD (particle size distribution), which is considered the best choice for most scenarios. Additional information can be found in the instrument user manual.

### **(B) TEM**

Timing 1–4 h depending on the number of samples to be analyzed

- (i) Prepare a sample of the compound with a concentration of 200  $\mu\text{M}$  in buffer or medium (as required for the function of the desired experiment).
- (ii) Centrifuge at 400g for 10 min at room temperature.
- (iii) Deposit 100  $\mu\text{l}$  (1 drop) of the compound supernatant solution on the Parafilm.
- (iv) Add the copper grid (200 mesh) on the drop and wait 5–10 min.

### **CRITICAL STEP**

The shiny side of the grid must be appropriately placed to be facing the droplet (not the matte side).

- (v) Remove the grid and place it on a drop of 3% (wt/vol) PTA for 15 s.
- (vi) Remove the grid and dry it with an absorbent paper.
- (vii) Place the grid in the transmission electron microscope for analysis.

### **CRITICAL STEP**

Higher concentrations of sample might be required to observe self-assemblies by TEM. However, bear in mind that, as discussed in the Introduction, the physical sizes of the aggregates may differ

from those seen by DLS, CLSM and NMR; thus, results will need to be interpreted with caution in such cases.

### (C) CLSM

Timing 2–3 d depending on the required incubation time

#### **CAUTION**

Before starting, determine whether the compound is fluorescent with a fluorometer or directly on the microscope by using the compound alone. Note the fluorescent wavelength. If the compound is not fluorescent, CLSM cannot be used.

#### **CRITICAL**

If cells are already growing in culture, in place of Step 26C(i and ii), obtain a suspension of cells.

#### **Cell thawing**

- (i) If required, remove the cell vial (frozen at 1 million cells/vial) from the freezer and thaw the cells in a water bath at 37 °C. This usually takes <1 min.
- (ii) When the sample is thawed, transfer the cells to a centrifuge tube.
- (iii) Centrifuge at 125g for 7 min at room temperature.
- (iv) Remove the supernatant.
- (v) Add 5 ml of medium and resuspend the cell pellet.
- (vi) Transfer all of the cells and medium to a 25-cm<sup>2</sup> flask.
- (vii) Incubate at 37 °C in an incubator with 5% CO<sub>2</sub> for 3 or 4 d or until cells reach 80% confluency.

#### **CRITICAL STEP**

Check for cell contamination and occasionally monitor the growth of the cells.

- (viii) Make a stock solution with the compound at 20 mM in DMSO, as detailed in Steps 1–3.

#### **CRITICAL STEP**

Although deuterated DMSO is not required for this assay, the same DMSO-d<sub>6</sub> stocks used for NMR assays can be used (from Step 3).

(ix) From the stock solution, make a new 100  $\mu\text{M}$  sample in medium (e.g., DMEM). Ensure that sufficient 100  $\mu\text{M}$  sample is made up;  $\geq 500 \mu\text{l}$  is required for each well of a 24-well plate that requires compound treatment.

### **CRITICAL STEP**

Bear in mind that if DMEM with phenol red is used, fluorescence from the additive may overlap/interfere with any produced by the compound. It is therefore recommended to use DMEM without phenol red.

## **Cell seeding**

(x) When the cells reach 80% confluency, remove the old medium and wash with sterile PBS (use 2 ml if using a 25-cm<sup>2</sup> flask).

(xi) Remove the PBS solution.

(xii) Add trypsin (use 1 ml if using a 25-cm<sup>2</sup> flask) and put the flask in an incubator at 37 °C with 5% CO<sub>2</sub> for 5–10 min.

### **CRITICAL STEP**

After 5 min, check to see if the cells have started to detach. If so, proceed to the next step.

(xiii) Centrifuge the cell suspension at 800g for 7 min. Discard the supernatant.

(xiv) Add medium (use 4 ml if using a 25-cm<sup>2</sup> flask) and resuspend the cells.

(xv) Count the cells and then dilute them to obtain 600,000 cells/well (1.2 million cells/ml) and 500  $\mu\text{l}$ /well.

(xvi) Add one glass coverslip on the bottom of each well to be used in a 24-well plate.

(xvii) Add 500  $\mu\text{l}$  of the cell stock to each well and put these plates in the incubator at 37 °C with 5% CO<sub>2</sub> overnight.

## **Cell treatment**

(xviii) On the next day, if the cells are growing well, remove the old medium.

### **CRITICAL STEP**

Cells can alternatively be left for longer before proceeding with treatment.

(xix) Wash cells with PBS.

(xx) Add 500  $\mu$ l of the compound solution (at 100  $\mu$ M in medium).

### **CRITICAL STEP**

Compound concentrations used for other techniques may not directly transpose to CLSM because they can result in higher cell toxicity. Testing a range of concentrations might be required depending on the molecules tested.

(xxi) Put the plate in an incubator for the desired incubation time (e.g., 2, 3 or 4 h or longer).

### **Staining and fixation**

(xxii) Remove the old medium and wash cells with PBS.

(xxiii) Add 300  $\mu$ l of membrane stain (WGA-555) at a concentration of 3.33  $\mu$ g/ml and wait 5 min.

(xxiv) Remove the membrane stain solution and wash cells once with PBS.

(xxv) Add 350  $\mu$ l of the fixative PFA at 4% (vol/vol) and wait another 10 min.

(xxvi) Remove PFA and wash cells twice with PBS.

(xxvii) Add 300  $\mu$ l of the nucleus stain (DRAQ-5) and wait 15 min.

(xxviii) Remove the staining solution and add 1 ml of PBS.

(xxix) Add one drop of mounting medium (glycerol at 70–90% (vol/vol)) on the slide.

(xxx) Remove the coverslip from the well and dry it on the absorbent paper.

(xxxi) Put the coverslip on the slide with mounting medium on the appropriate side (cells should be between the slide and the coverslip, with the mounting medium).

(xxxii) Let dry for 5–10 min under a ventilated hood.

(xxxiii) Visualize under a confocal microscope.

### **Troubleshooting**

Troubleshooting advice can be found in Table 2.

**Table 2: Troubleshooting table**

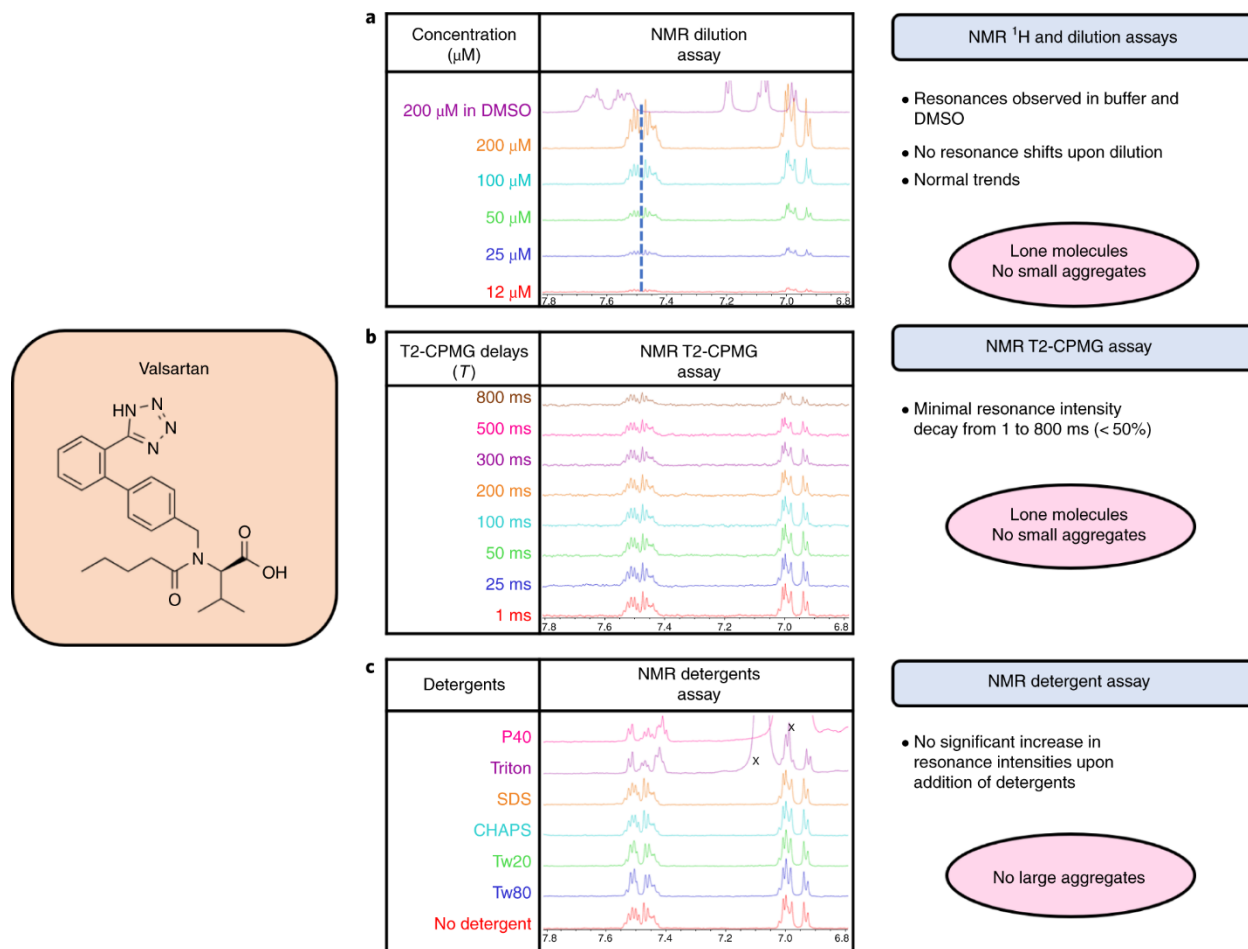
Step	Problem	Possible reason	Possible solution
4B(iii) (NMR <sup>1</sup> H assay: sample preparation in buffer)	Precipitate formation	Insolubility of the compound	Use the supernatant after centrifugation at 400g for 10 min. If this is done, any observable precipitate should be noted
9 (NMR dilution assay: NMR spectra)	No signal or weak signals for samples used in the fast-track method	Solubility is too low	Run the NMR dilution assay by using lower concentrations
14 (NMR detergent assay: addition of detergent to the samples)	The compound forms a suspension	Centrifugation speed is too low, or duration is not long enough	Increase the time and/or speed of the centrifugation. Note any observations
19 (Solvent solubility assay: compound dissolution in solvent)	Precipitate formation	Insolubility of the compound	Centrifuge at 400g for 10 min. If this is not enough, increase the time and/or speed
26A(iv) (DLS assay)	Failure of measurement	Incompatible sample (color or polydispersity)	Trying a different laser color (if your instrumentation allows it) may help circumvent problems with colored samples. Reducing sample concentration might help reduce polydispersity, although results might not directly compare with techniques that used different concentrations
26A(vi) (DLS assay: data analysis)	Inconclusive results	Inappropriate sample preparation	Wash the cuvette with alcohol and dry with a pressurized air canister to remove any contaminants. Try using different compound concentrations
26B(vii) (TEM assay: micrograph observation)	Nothing observable in TEM micrograph	Low concentration or precipitate	Try different compound concentrations. Some compounds have distinct aggregation behavior that varies with concentration, depending on the type of buffer, temperature, detergents, etc

Step	Problem	Possible reason	Possible solution
26C(xxxiii) (CLSM assay: visualisation under the microscope)	No coloration or insufficient fixation	Incubation time with the stains or fixative was too short, or the concentration of the reagents was too low	Increase the incubation time or increase the concentration of the stains
	No aggregates observable	Addition of FBS may sequester some or all aggregates, leading to no fluorescence being observed inside the cells (false negatives)	Use culture medium without FBS

## Anticipated results

To date, we have used this procedure (in whole or in part) to monitor the solution behavior of thousands of compounds in both confidential projects and some published papers<sup>1,6,7,8,9</sup>. Here, we show four examples of results we have obtained that demonstrate the type of data acquired, how they can be displayed and how to interpret the results obtained. We also include further examples of results we have obtained in Supplementary Figures 1–6. Full details of all compounds used are given in Supplementary Table 3.

In the first example, we evaluated the solution behavior of valsartan. Sharp NMR resonances were observed by NMR <sup>1</sup>H assay for valsartan in the spectra acquired in buffer and DMSO-d<sub>6</sub> at 200 μM nominal concentration (Fig. 6a). In addition, all samples were clear with no precipitate upon visual inspection, confirming that valsartan is soluble in both solvents.



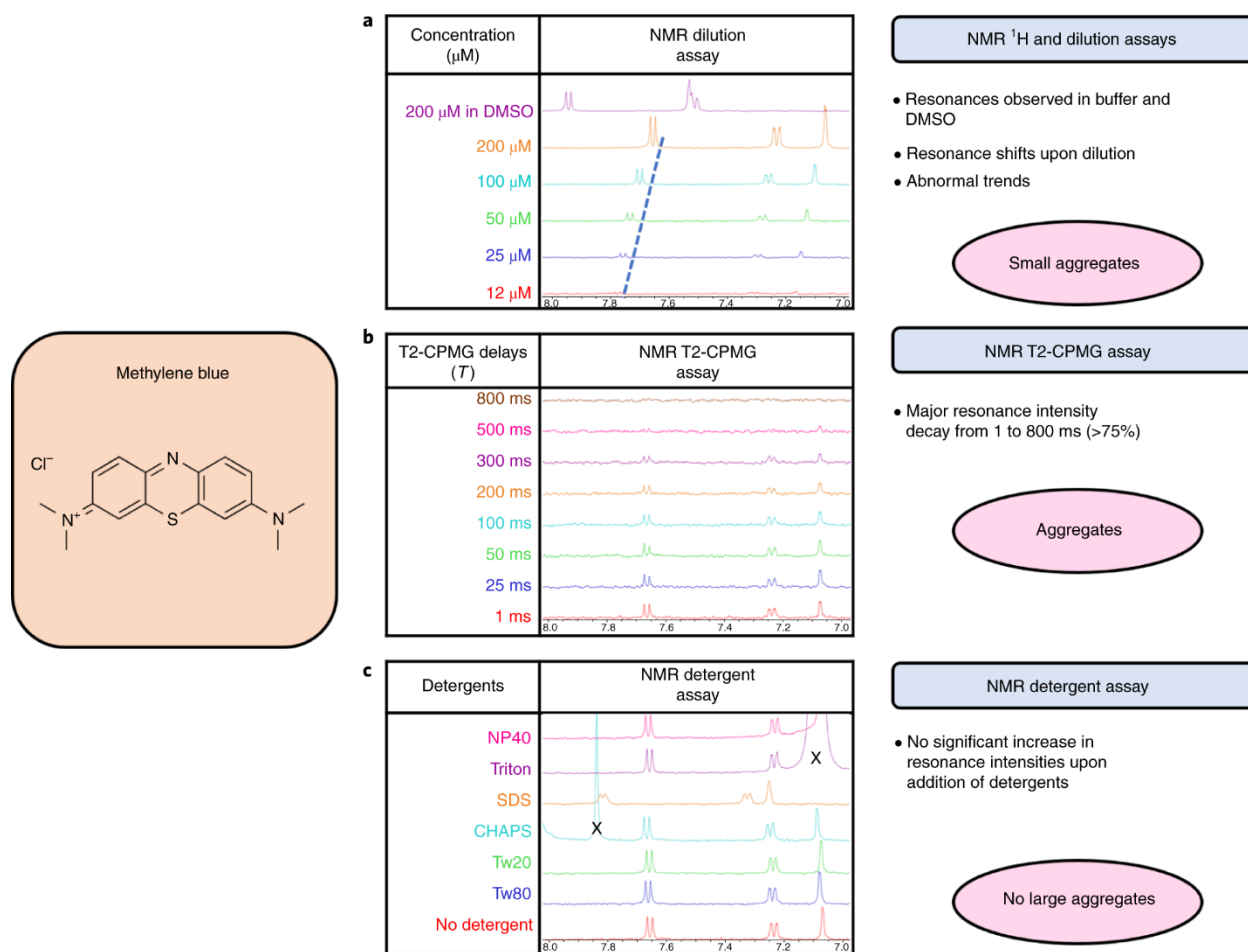
**Fig. 6: Probing the solution behavior of valsartan.**

**A–c, Shown are NMR data from the NMR <sup>1</sup>H and dilution assays (a), the T2-CPMG assay (b) and the NMR detergent assay (c). The chemical structure is shown on the left. The ‘X’s in c denote resonances that arise from the detergent and not from the compound. Tw20, Tween 20; Tw80, Tween 80.**

Furthermore, upon dilution in buffer from 200 to 12  $\mu\text{M}$ , normal trends were seen (i.e., Fig. 6a: decreases in intensity with no changes in peak shapes or chemical shifts). These observations were consistent with the behavior of lone tumbling molecules with no self-association tendencies. These conclusions were corroborated by data from the NMR T2-CPMG assay (Fig. 6b), where the resonance intensities were minimally affected upon comparison of the resonances of the spectra by using 1-ms versus 800-ms delays (small resonance intensity decay <50%). Furthermore, no significant increases in resonance intensities or appearance of new peaks were noted upon addition of detergents in the NMR detergent assay (Fig. 6c), supporting the conclusion that no large aggregates exist.



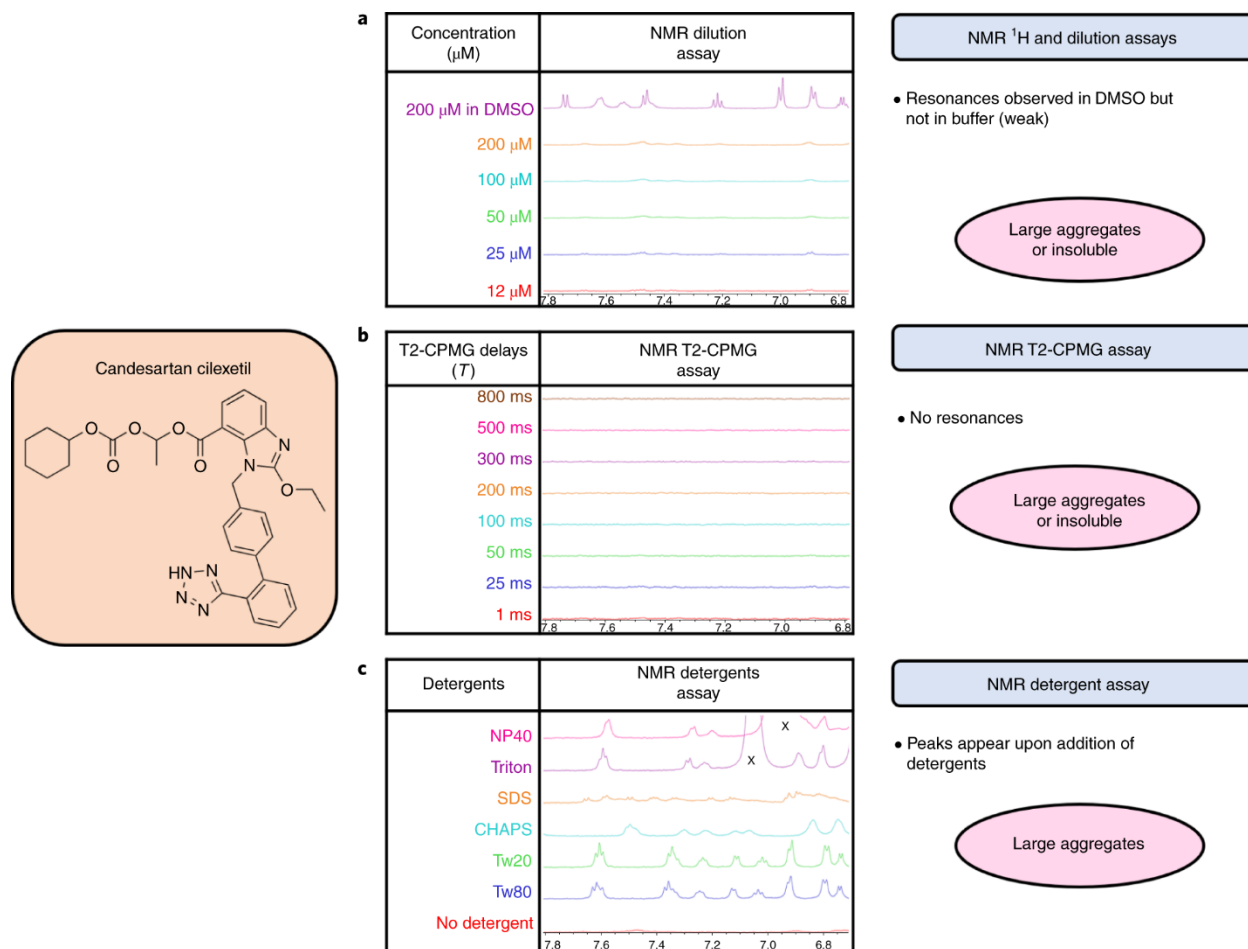
In the second example, we evaluated the solution behavior of methylene blue. NMR resonances are observed for methylene blue in both buffer and DMSO- $d_6$  by NMR  $^1H$  assay (Fig. 7a). In addition, the samples were blue and clear with no precipitate upon visual inspection, confirming that methylene blue was soluble in both conditions. However, upon dilution in buffer from 200 to 12  $\mu M$ , abnormal trends were seen. Indeed, concentration-dependent changes in chemical shifts are observable, consistent with the behavior of self-association into small aggregates. These conclusions were corroborated by data from the NMR T2-CPMG assay (Fig. 7b), where the resonance intensities were appreciably different upon comparison between the resonances of the spectra by using 1-ms versus 800-ms delays (major resonance intensity decay  $>75\%$ ). Interestingly, no significant increases in resonance intensities were noted upon addition of detergents in the NMR detergent assay (Fig. 7c), suggesting that no large aggregates exist. Thus, methylene blue probably self-associates into small aggregates and not into large aggregates.



**Fig. 7: Probing the solution behavior of methylene blue.**

**A–c, Shown are NMR data from the NMR  $^1\text{H}$  and dilution assays (a), the T2-CPMG assay (b) and the NMR detergent assay (c). The chemical structure is shown on the left. The 'X's in c denote resonances that arise from the detergent and not from the compound.**

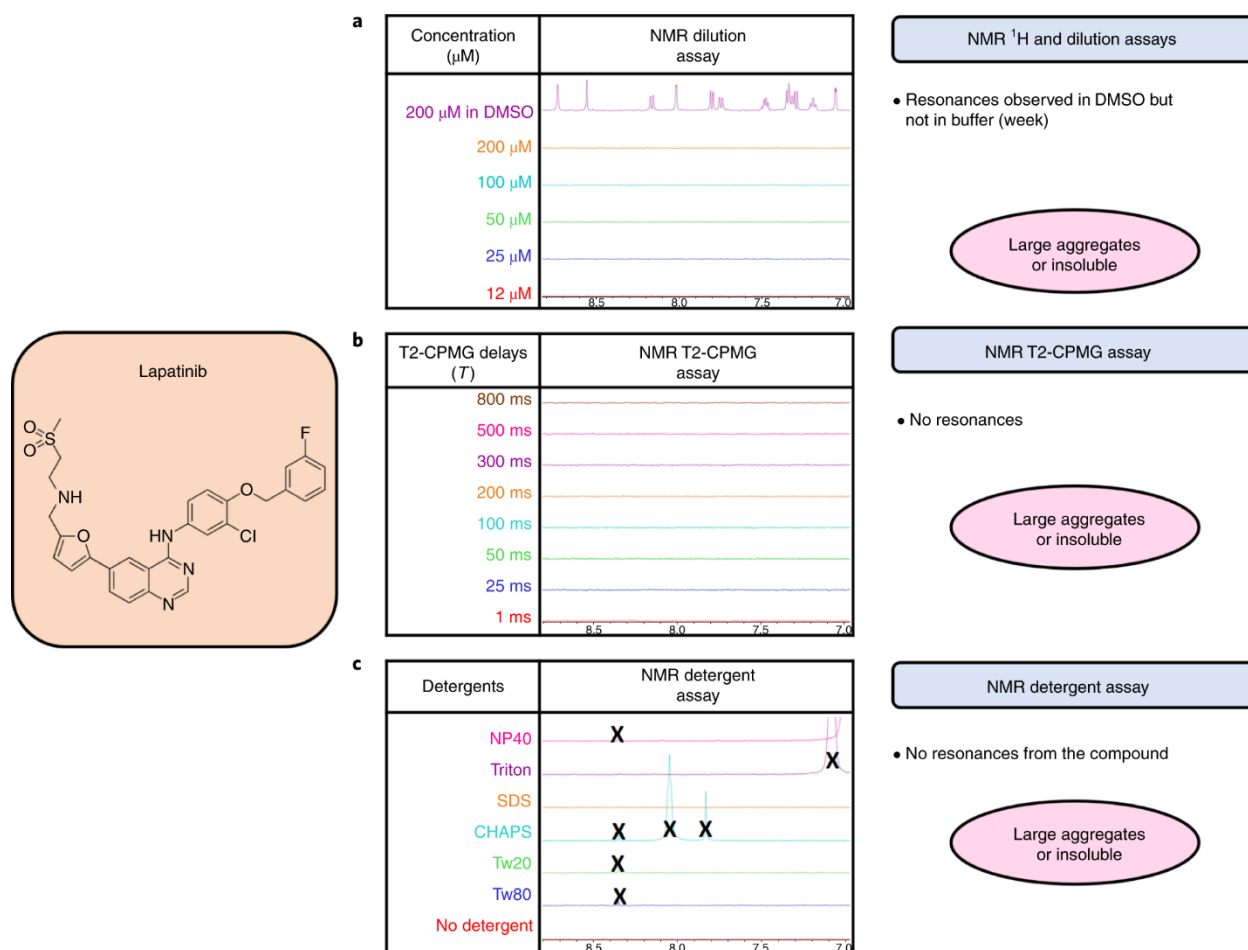
In the third example, we evaluated candesartan cilexetil (referred to here as candesartan). NMR resonances are present for the DMSO- $d_6$  sample but non-existent in buffer by NMR  $^1\text{H}$  assay (Fig. 8a). In addition, the samples were clear and neither were turbid nor had precipitate upon visual inspection, which confirms that candesartan is soluble in both conditions. Furthermore, upon dilution in buffer from 200 to 12  $\mu\text{M}$ , no resonances were noted (Fig. 8a), and no cloudiness or precipitate was notable. The absence of resonances obviously led to no resonances being observed in the NMR T2-CPMG assay (Fig. 8b). Taken together, these observations were consistent with the lack of lone tumbling molecules and small aggregates and suggestive of the presence of large aggregates. Interestingly, significant increases in resonance intensities were noted upon addition of detergents in the NMR detergent assay (Fig. 8c), demonstrating that candesartan exists as large soft aggregates in buffer that can be disrupted by the addition of detergents.



**Fig. 8: Probing the solution behavior of candesartan cilexetil.** A–c, Shown are NMR data from the NMR <sup>1</sup>H and dilution assays (a), the T2-CPMG assay (b) and the NMR detergent assay (c). The chemical structure is shown on the left. The ‘X’s in c denote resonances that arise from the detergent and not from the compound.

In the fourth example, we evaluated the solution behavior of lapatinib, showing unexpected trends. NMR resonances were observed for lapatinib in DMSO-d<sub>6</sub> at 200 μM nominal concentration but not in buffer by NMR <sup>1</sup>H assay (Fig. 9a). The buffer samples were slightly opaque with some precipitate upon visual inspection. No resonances appeared when the compound was diluted in buffer from 200 to 12 μM (Fig. 9a), and no resonances were seen in the NMR T2-CPMG assay (Fig. 9b). These observations are consistent with a lack of lone molecules or small aggregates and suggest that the compound is either partially insoluble or that it forms large aggregates. To our surprise, no significant increases in resonance intensities were noted upon addition of detergents in the NMR detergent assay (Fig. 9c). (The peak appearing at 8.3 p.p.m. is one of an impurity.) Given these observations, it could not be confirmed whether the lack

of resonances was due to the compound being partially insoluble or the fact that large hard aggregates were being formed in the solution. We thus proceeded to further orthogonal methods to confirm the solution behaviors.

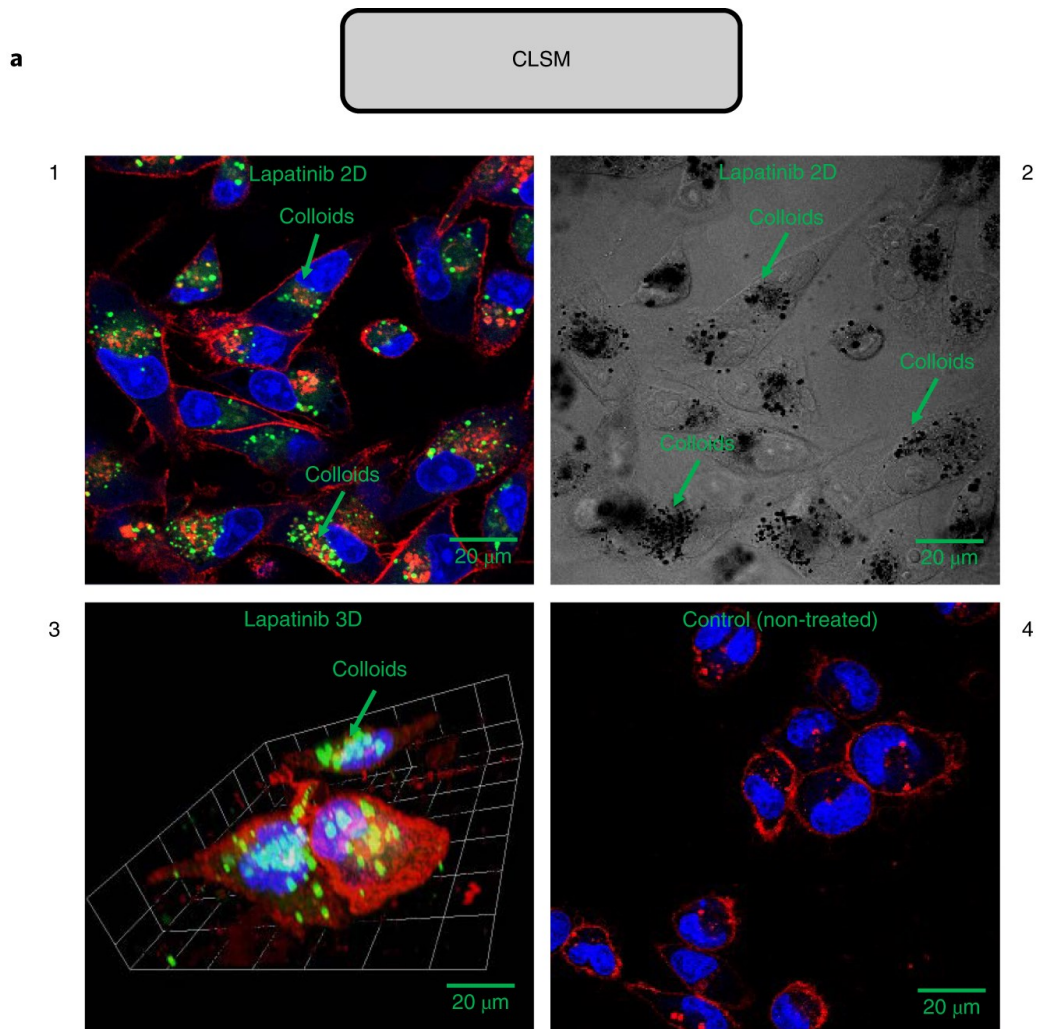


**Fig. 9: Probing the solution behavior of lapatinib.**

**A–c, Shown are NMR data from the NMR  $^1\text{H}$  and dilution assays (a), the T2-CPMG assay (b) and the NMR detergent assay (c). The chemical structure is shown on the left. The ‘X’s in c denote resonances that arise from the detergent or impurities and not from the compound.**

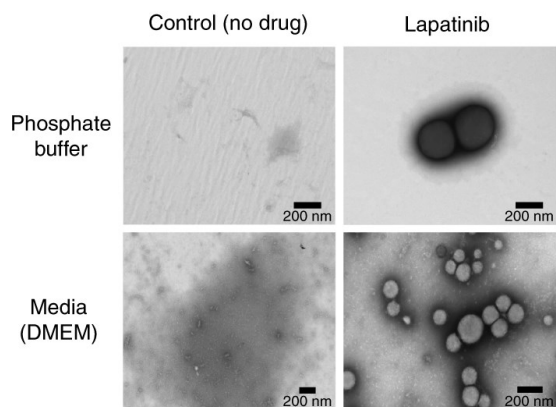
Large aggregates were seen by CLSM (Fig. 10a) and TEM (Fig. 10b) and further confirmed by using DLS (Fig. 10c). It is, however, intriguing that addition of Tween 80 led to a loss of detection of lapatinib aggregates by DLS (Fig. 10c), suggesting breaking of the entity, but no signal was observed by NMR in the presence of detergent. Because of the use of a shared high-throughput sample handler (SampleJet) on the NMR, sample acquisition can sometimes only be done several hours after sample submission. Therefore, it is possible that lapatinib aggregates precipitate out

of solution over time. Visual precipitate was observable upon sample preparation, and storage in a refrigerated NMR SampleJet such as was used here could also 124eglerated sample precipitation.



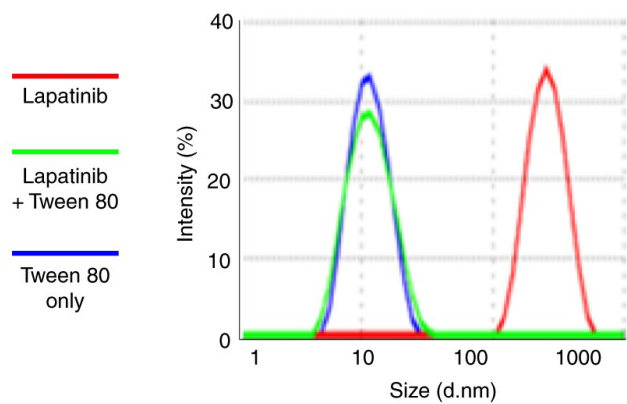
**b**

TEM



**c**

DLS



**Fig. 10: Probing the solution behavior of lapatinib by using orthogonal techniques.**

**A–c, Shown are data from the orthogonal methods CLSM (100  $\mu\text{M}$ ) (a), TEM (50  $\mu\text{M}$ ) (b) and DLS (200  $\mu\text{M}$ ) (c). a, CLSM images of lapatinib in Raw 264.7 cells. The aggregates are in green (fluorescence of compound), the membrane of the cells in red (WGA 555 staining) and the nucleus in blue (DRAQ-5 staining). The three primary images show the aggregates in cells in 2D (with fluorescence (1) and without fluorescence (2)) and 3D (3). The last picture (4), shows the cells without compound (control). The 3D picture (3) is a zoom on one cell. B, TEM images of lapatinib in phosphate buffer and DMEM with 5% (vol/vol) FBS and its controls. C, Lapatinib at 200  $\mu\text{M}$  in sodium phosphate buffer in the presence and absence of 0.025% (vol/vol) Tween 80 for 24 h. The size of lapatinib aggregates can be different between TEM (b) and DLS (c), possibly because of dehydration of the sample during the preparation for the observation with the TEM or the differences in concentrations used. D.nm, diameter values in nanometers.**

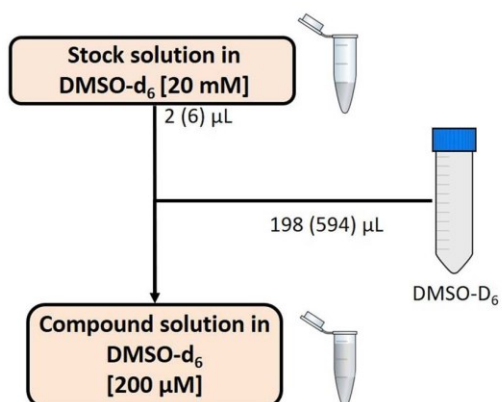
## **Conclusions**

The examples above show that no one technique can detect all the types of aggregates that exist; thus, a combination of strategies is necessary as proposed in this protocol. One of the examples demonstrates that no single method can expose the full range of nano-entities that exist, highlighting the importance of using coordinated orthogonal strategies. We hypothesize that there are distinct attributes of aggregates that can be further identified and anticipate that further attributes such as these will have distinct properties that merit revealing and exploitation. We hope that the wide use of this protocol along with future improved versions will advance knowledge of the largely unexplored world of drug nano-entities. We suspect that there are many types of nano-entities that exist that have various sizes and architectures, and that these will be correlated with properties that can potentially be exploited, for example, for drug delivery, anti-aggregates, cell penetrators and bioavailability enhancers.

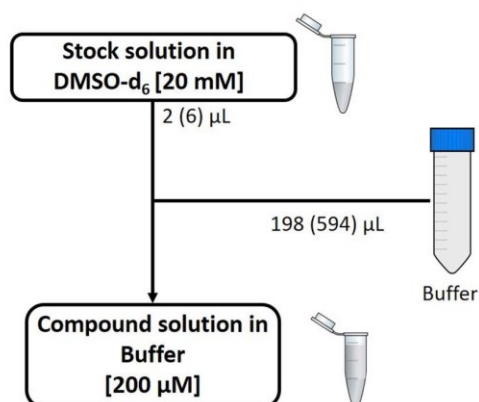
## Extended data

### NMR $^1\text{H}$ Assay

#### A. DMSO- $d_6$ assay preparation



#### B. Buffer assay preparation

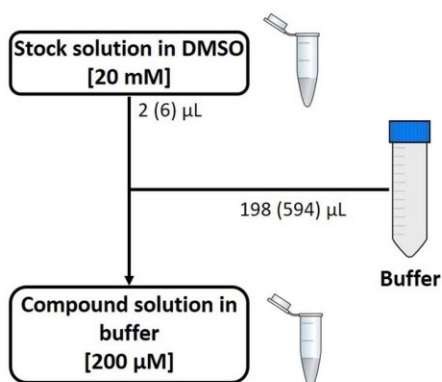


### Extended Data Fig. 1: NMR $^1\text{H}$ assay.

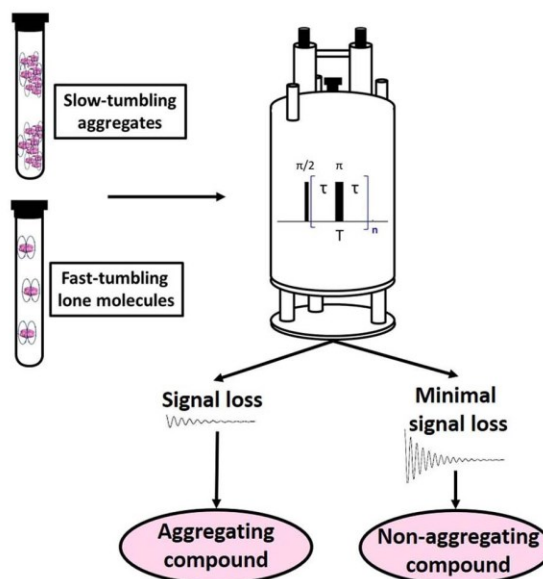
A, Preparation for assay in DMSO- $d_6$ . B, Preparation for assay in buffer. Shown are volumes suggested for 3-mm NMR tubes, and volumes for 5-mm tubes are in parentheses.

### NMR T2-CPMG Assay

#### A. Sample preparation



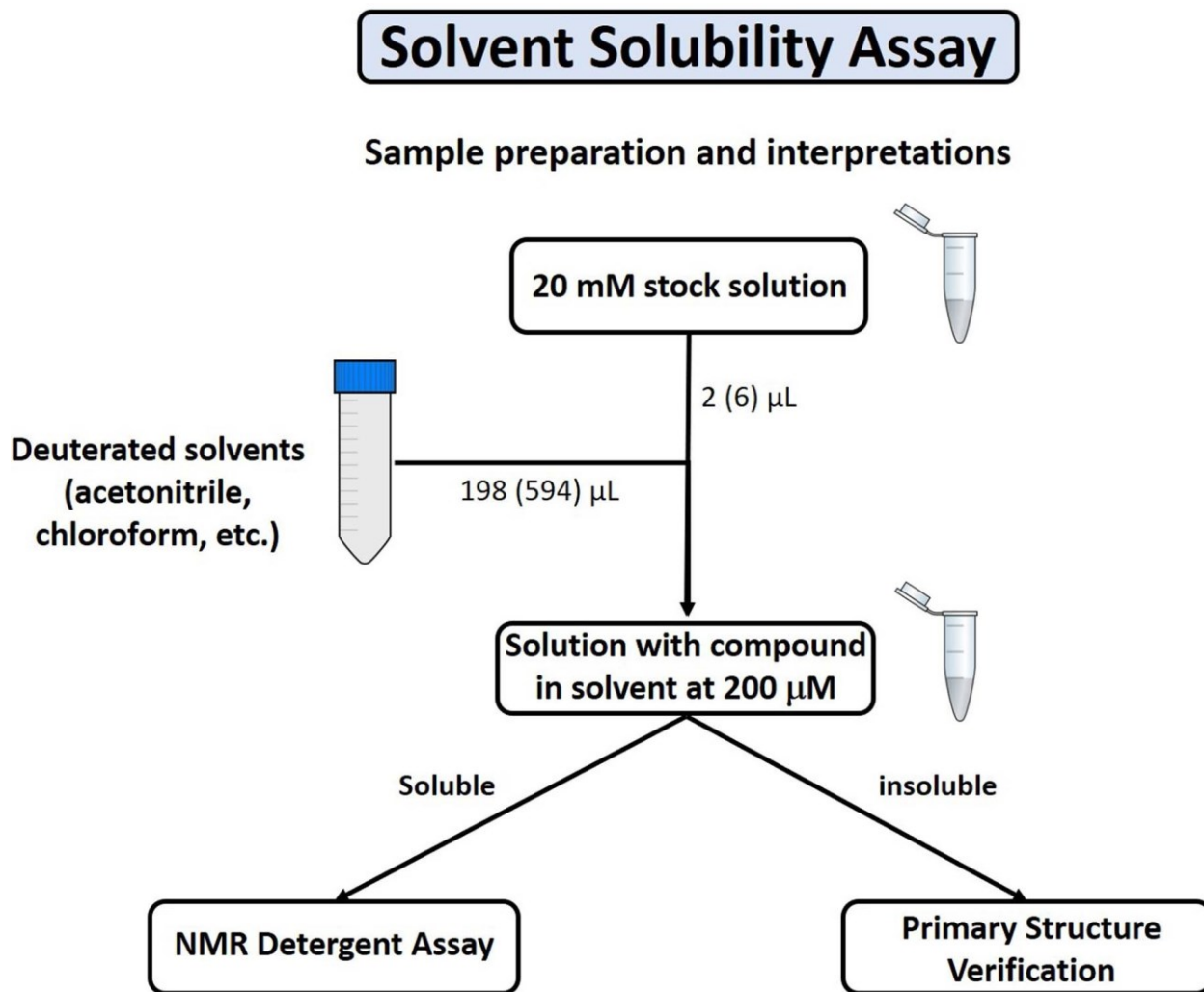
#### B. Observations and interpretations



### Extended Data Fig. 2: NMR T2-CPMG assay.

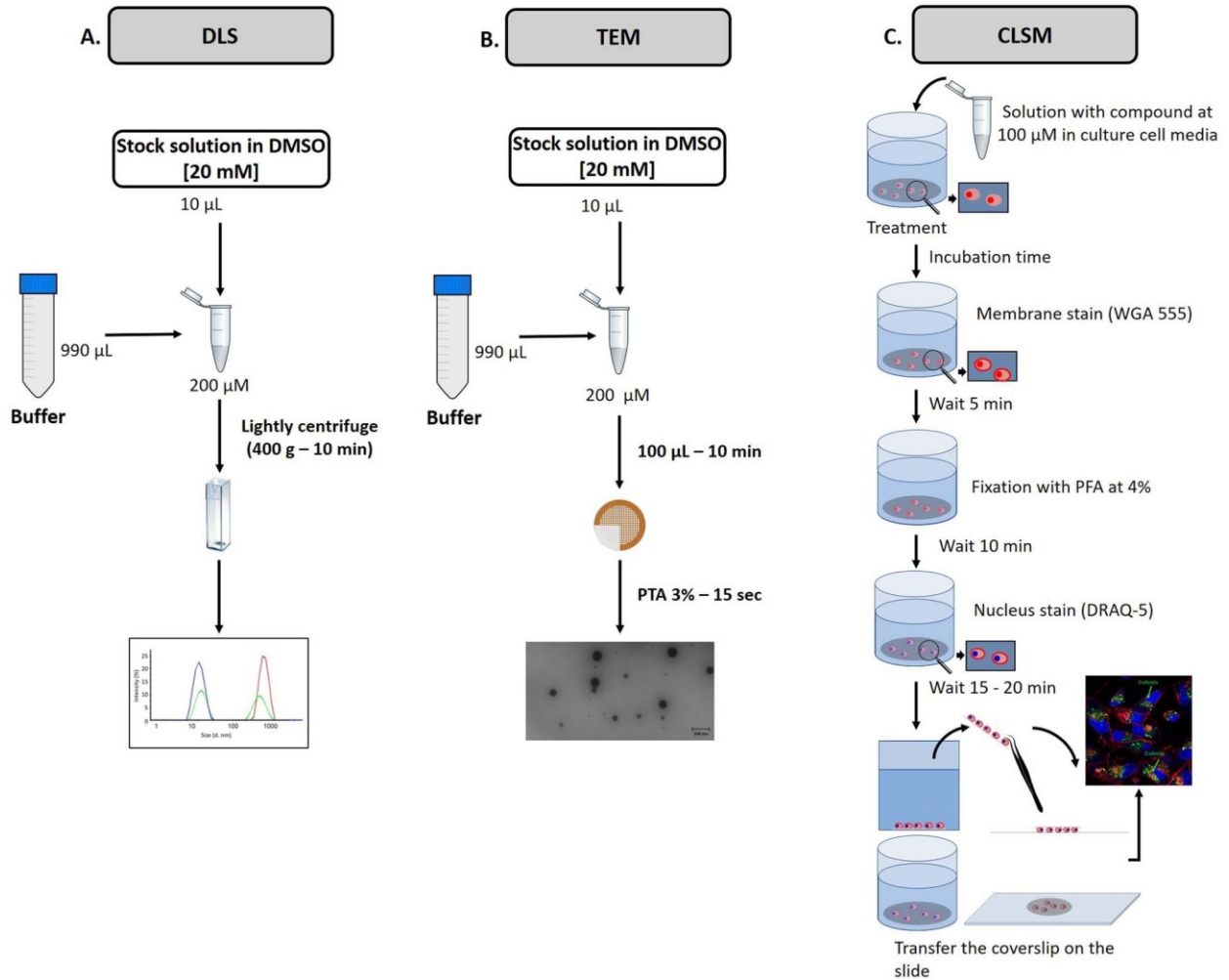


A, Preparation of the sample. B, Interpretation of the results.



Extended Data Fig. 3: Solvent solubility assay. Preparation of the sample and interpretations.

## Orthogonal Assays



**Extended Data Fig. 4: Orthogonal assays.**

**A–c,** Preparation of the samples for DLS (a), TEM (b) and CLSM (c).

## Acknowledgements

We thank the following agencies for helping fund this research: NSERC (Natural Sciences and Engineering Research Council of Canada), CQDM (Quebec Consortium for Drug Discovery), CFI (Canada Foundation for Innovation), Mitacs, INRS (Institut national de la recherche scientifique), Institut Pasteur, la 130e région Auvergne-Rhône-Alpes, le ministère de l'enseignement supérieur et de la recherche (France) and NMX Research and Solutions Inc. We also thank our colleagues for their help, suggestions and encouragement: P. Bouchard, N. Girard, D. Bendahan, D. Girard, J. Tremblay and A. Nakamura.

## Contributions

S.R.L.P. conceived the concepts described in this report. S.R.L.P. and V.R. wrote the paper. V.R., F.S., G.L.P., M.M.D., S.R.L.P., S.T.L., Y.A. and S.W. performed the experiments or helped with interpretations. Y.A. and S.T.L. implemented some of the experiments used herein.

## Corresponding author

Correspondence to [Steven R. LaPlante](#).

## Ethics declarations

Competing interests

The authors declare no competing interests.

## Data availability

The NMR data that support Figs. 6–9 and Supplementary Figs. 1–5 are available in figshare (<https://doi.org/10.6084/m9.figshare.15019755.v1>).

## References

1. LaPlante, S. R. et al. Monitoring drug self-aggregation and potential for promiscuity in off-target in vitro pharmacology screens by a practical nmr strategy. *J. Med. Chem.* **56**, 7073–7083 (2013).

2. Duan, D. et al. Internal structure and preferential protein binding of colloidal aggregates. *ACS Chem. Biol.* **12**, 282–290 (2017).
3. Feng, B. Y. & Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors. *Nat. Protoc.* **1**, 550–553 (2006).
4. Feng, B. Y. et al. A high-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* **50**, 2385–2390 (2007).
5. Owen, S. C., Doak, A. K., Wassam, P., Shoichet, M. S. & Shoichet, B. K. Colloidal aggregation affects the efficacy of anticancer drugs in cell culture. *ACS Chem. Biol.* **7**, 1429–1435 (2012).
6. Dlim, M. M., Shahout, F. S., Khabir, M. K., Labonté, P. P. & Laplante, S. R. Revealing drug self-associations into nano-entities. *ACS Omega* **4**, 8919–8925 (2019).
7. Ayotte, Y. et al. Exposing small-molecule nanoentities by a nuclear magnetic resonance relaxation assay. *J. Med. Chem.* **62**, 7885–7896 (2019).
8. LaPlante, S. R. et al. Compound aggregation in drug discovery: implementing a practical NMR assay for medicinal chemists. *J. Med. Chem.* **56**, 5142–5150 (2013).
9. Beaulieu, P. L. et al. Multi-parameter optimization of aza-follow-ups to BI 207524, a thumb pocket 1 HCV NS5B polymerase inhibitor. Part 2: impact of lipophilicity on promiscuity and in vivo toxicity. *Bioorg. Med. Chem. Lett.* **25**, 1140–1145 (2015).
10. Owen, S. C. et al. Colloidal drug formulations can explain ‘bell-shaped’ concentration-response curves. *ACS Chem. Biol.* **9**, 777–784 (2014).
11. Tres, F., Posada, M. M., Hall, S. D., Mohutsky, M. A. & Taylor, L. S. The effect of promiscuous aggregation on in vitro drug metabolism assays. *Pharm. Res.* **36**, 1–9 (2019).
12. Ganesh, A. N. et al. Colloidal drug aggregate stability in high serum conditions and pharmacokinetic consequence. *ACS Chem. Biol.* **14**, 751–757 (2019).
13. Frenkel, Y. V. et al. Concentration and pH dependent aggregation of hydrophobic drug molecules and relevance to oral bioavailability. *J. Med. Chem.* **48**, 1974–1983 (2005).
14. Frenkel, Y. V., Gallicchio, E., Das, K., Levy, R. M. & Arnold, E. Molecular dynamics study of non-nucleoside reverse transcriptase inhibitor 4-[[4-[[4-(2-cyanoethenyl)-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino] benzonitrile (TMC278/rilpivirine) aggregates:

- correlation between amphiphilic properties of the drug and oral bioavailability. *J. Med. Chem.* **52**, 5896–5905 (2009).
15. Ganesh, A. N., Donders, E. N., Shoichet, B. K. & Shoichet, M. S. Colloidal aggregation: from screening nuisance to formulation nuance. *Nano Today* **19**, 188–200 (2018).
  16. Hoo, C. M., Starostin, N., West, P. & Mecartney, M. L. A comparison of atomic force microscopy (AFM) and dynamic light scattering (DLS) methods to characterize nanoparticle size distributions. *J. Nanopart. Res.* **10**, 89–96 (2008).
  17. Tomaszewska, E. et al. Detection limits of DLS and UV-Vis spectroscopy in characterization of polydisperse nanoparticles colloids. *J. Nanomater.* **2013**, 313081 (2013).
  18. Akoka, S., Barantin, L. & Trierweiler, M. Concentration measurement by proton NMR using the ERETIC method. *Anal. Chem.* **71**, 2554–2557 (1999).
  19. Seidler, J., McGovern, S. L., Doman, T. N. & Shoichet, B. K. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J. Med. Chem.* **46**, 4477–4486 (2003).
  20. Doak, A. K., Wille, H., Prusiner, S. B. & Shoichet, B. K. Colloid formation by drugs in simulated intestinal fluid. *J. Med. Chem.* **53**, 4259–4265 (2010).
  21. Hassan, P. A., Rana, S. & Verma, G. Making sense of Brownian motion: colloid characterization by dynamic light scattering. *Langmuir* **31**, 3–12 (2015).
  22. Ganesh, A. N., McLaughlin, C. K., Duan, D., Shoichet, B. K. & Shoichet, M. S. A new spin on antibody-drug conjugates: trastuzumab-fulvestrant colloidal drug aggregates target HER2-positive cells. *ACS Appl. Mater. Interfaces* **9**, 12195–12202 (2017).
  23. Wiest, J. et al. Geometrical and structural dynamics of imatinib within biorelevant colloids. *Mol. Pharm.* **15**, 4470–4480 (2018).
  24. Hong, Y., Lam, J. W. Y. & Tang, B. Z. Aggregation-induced emission. *Chem. Soc. Rev.* **40**, 5361–5388 (2011).
  25. Hong, Y., Lam, J. W. Y. & Tang, B. Z. Aggregation-induced emission: phenomenon, mechanism and applications. *Chem. Comm.* (29), 4332–4353 (2009).

26. Kestens, V., Bozatzidis, V., De Temmerman, P. J., Ramaye, Y. & Roebben, G. Validation of a particle tracking analysis method for the size determination of nano- and microparticles. *J. Nanopart. Res.* **19**, 271 (2017).
27. Bevan, C. D. & Lloyd, R. S. A high-throughput screening method for the determination of aqueous drug solubility using laser nephelometry in microtiter plates. *Anal. Chem.* **72**, 1781–1787 (2000).
28. Vom, A. et al. Detection and prevention of aggregation-based false positives in STD-NMR-based fragment screening. *Aust. J. Chem.* **66**, 1518–1524 (2013).
29. Boulton, S. et al. Mechanisms of specific versus nonspecific interactions of aggregation-prone inhibitors and attenuators. *J. Med. Chem.* **62**, 5063–5079 (2019).
30. Yang, Z. Y. et al. Structural analysis and identification of colloidal aggregators in drug discovery. *J. Chem. Inf. Model.* **59**, 3714–3726 (2019).
31. Irwin, J. J. et al. An aggregation advisor for ligand discovery. *J. Med. Chem.* **58**, 7076–7087 (2015).
32. Meiboom, S. & Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* **29**, 688–691 (1958).
33. Ryan, A. J., Gray, N. M., Lowe, P. N. & Chung, C. W. Effect of detergent on ‘promiscuous’ inhibitors. *J. Med. Chem.* **46**, 3448–3451 (2003).
34. Feng, B. Y., Shelat, A., Doman, T. N., Guy, R. K. & Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* **1**, 146–148 (2005).
35. Pellecchia, M., Sem, D. S. & Wüthrich, K. NMR in drug discovery. *Nat. Rev. Drug Discov.* **1**, 211–219 (2002).
36. Hwang, J. & T., L. S. Water suppression that works. *J. Magn. Reson. A* **112**, 275–279 (1995).

## 5 SUPPORTING INFORMATION - ARTICLE 2

---

### **Probing the free-state solution behavior of drugs and their tendencies to self-aggregate into nano-entities**

Steven R. LaPlante<sup>1,2\*</sup>, Valérie Roux<sup>1</sup>, Fatma Shahout<sup>1</sup>, Gabriela LaPlante<sup>2</sup>, Simon Woo<sup>1</sup>, Maria M. Denk<sup>2</sup>, Sacha T. Larda<sup>2</sup>, Yann Ayotte<sup>1</sup>

<sup>1</sup> Université du Québec, INRS-Centre Armand-Frappier Santé Biotechnologie, 531 Boulevard des Prairies, Laval, Québec, H7V 1B7, CANADA

<sup>2</sup> NMX Research and Solutions Inc., 500 Boulevard Cartier Ouest, Laval, Québec, H7V 5B7, CANADA

Published in : Nature Protocols, 2021, 16, 5250-5273.

DOI : 10.1038/s41596-021-00612-3.

#### **Table of contents:**

**Supplementary Figure 1.** Probing the solution behavior of etodolac

**Supplementary Figure 2.** Probing the solution behavior of riluzole

**Supplementary Figure 3.** Probing the solution behavior of imatinib

**Supplementary Figure 4.** Probing the solution behavior of lansoprazole

**Supplementary Figure 5.** Probing the solution behavior of pranlukast

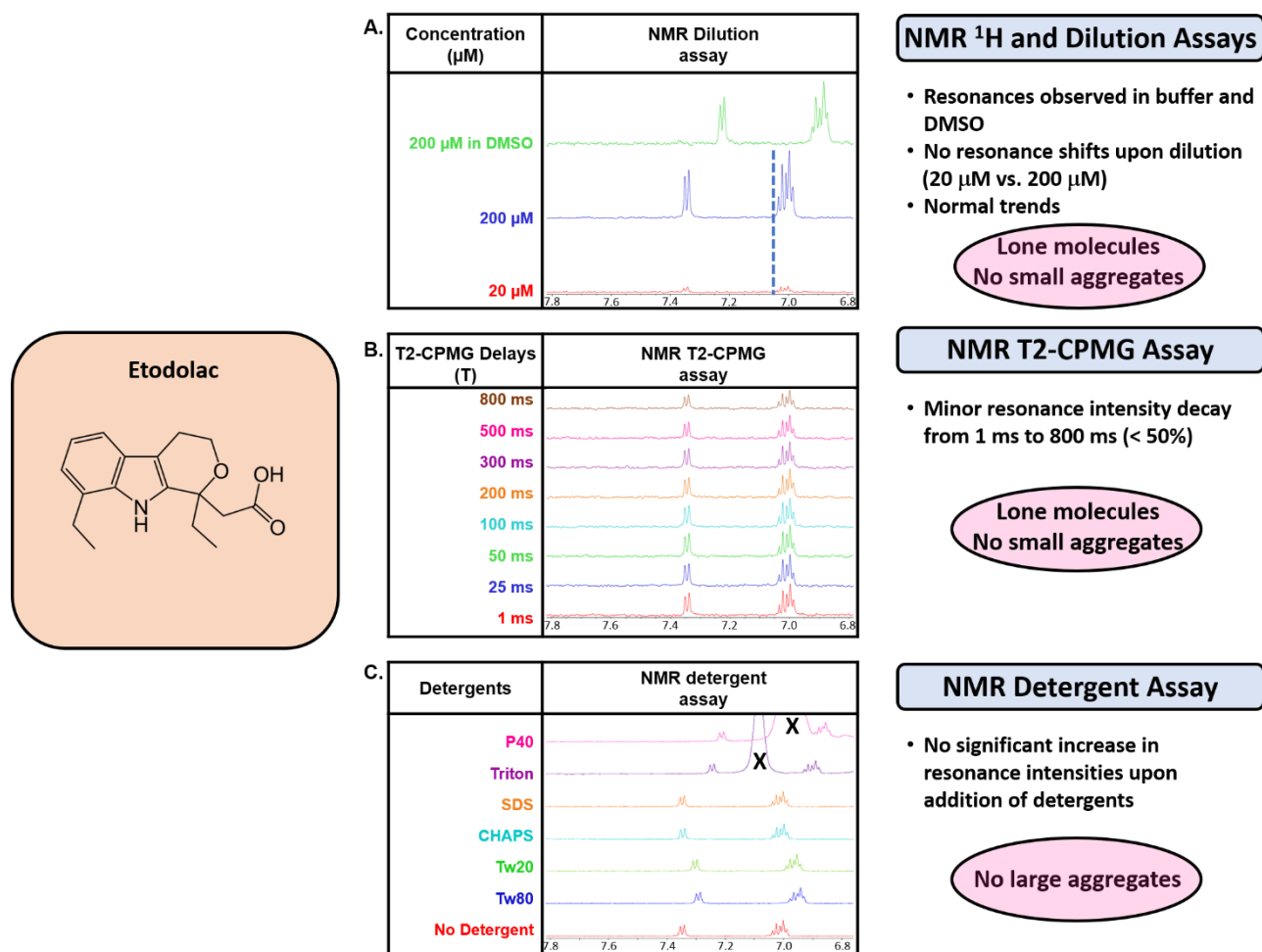
**Supplementary Figure 6.** Example of a possible DMSO aggregator

**Supplementary Table 1.** Detergents properties

**Supplementary Table 2.** Experimental guidelines for various sample set sizes

**Supplementary Table 3.** Compounds information

## Probing the solution behavior of etodolac



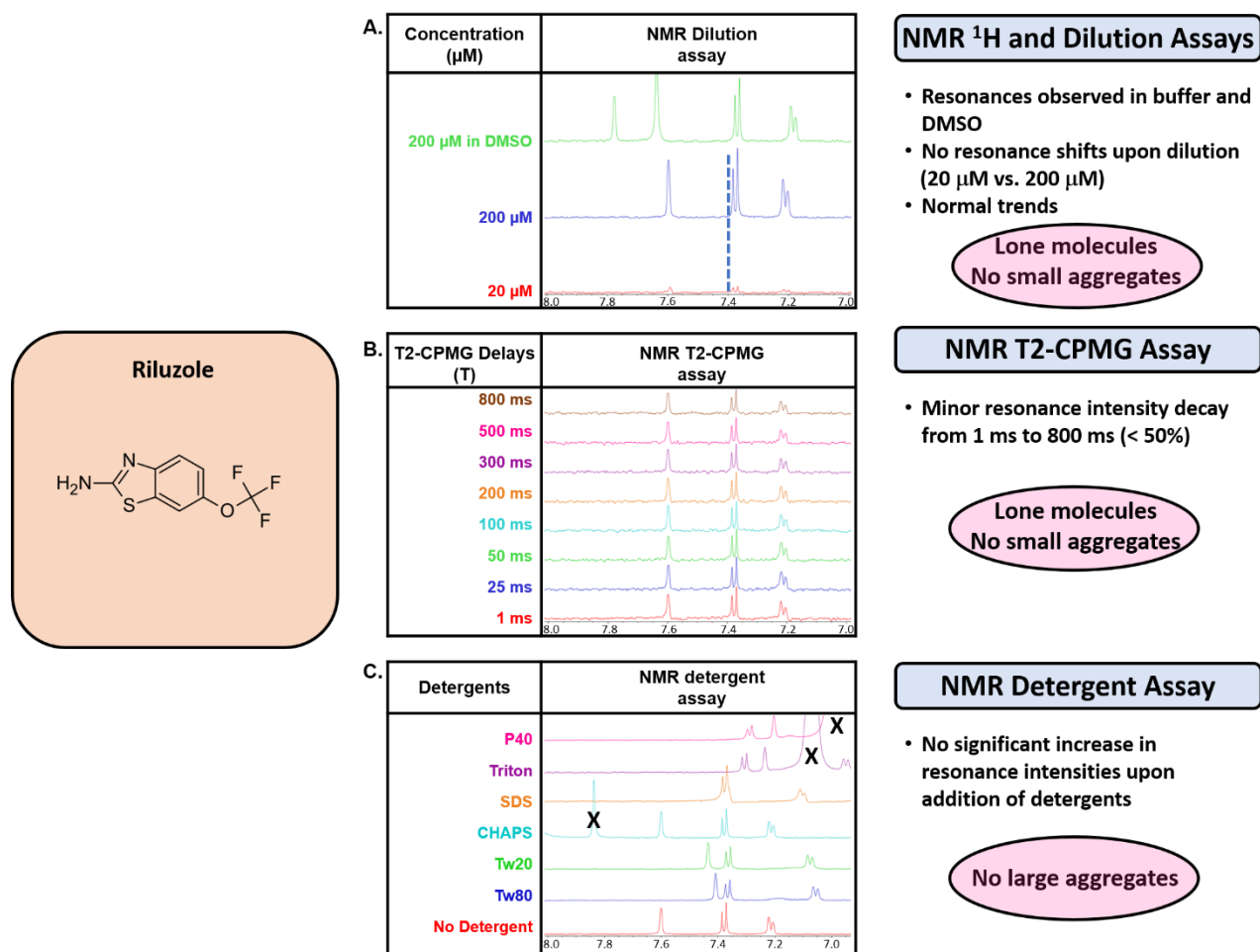
**Supplementary Figure 1 | Probing the solution behavior of etodolac.** Shown are NMR data from, (A) the NMR <sup>1</sup>H and Dilution Assays, (B) the T2-CPMG Assay, and (C) The NMR Detergent Assay. The chemical structure is shown on the left. “x” in (C) denote resonances that arise from the detergent and not from the compound.

The NMR <sup>1</sup>H Assay for etodolac in Supplementary Figure 1A resulted in observable resonances in both buffer and DMSO-d<sub>6</sub> at 200 μM nominal concentration. Also, the samples were clear with no precipitate upon visual inspection, confirming that etodolac was soluble in both solvents. Furthermore, upon dilution in buffer from 200 μM to 20 μM, normal trends were notable (i.e. Supplementary Figure 1A - decreases in signal intensity with no change in chemical shifts). These observations were consistent with a behavior of lone tumbling molecules with no self-association tendencies. These conclusions were corroborated by data from the NMR T2-CPMG Assay



(Supplementary Figure 1B) where the resonance intensities were minimally affected upon comparison of the resonances of the spectra employing delay times of 1 ms versus 800 ms (minor resonance intensity decay < 50%). Additionally, no significant increases in resonance intensities were noted upon addition of detergents in the NMR Detergent Assay (Supplementary Figure 1C) reporting that no large aggregates exist. The same finding can be observed with riluzole (Supplementary Figure 2). None of these two drugs exhibit evidence of aggregation at these concentrations.

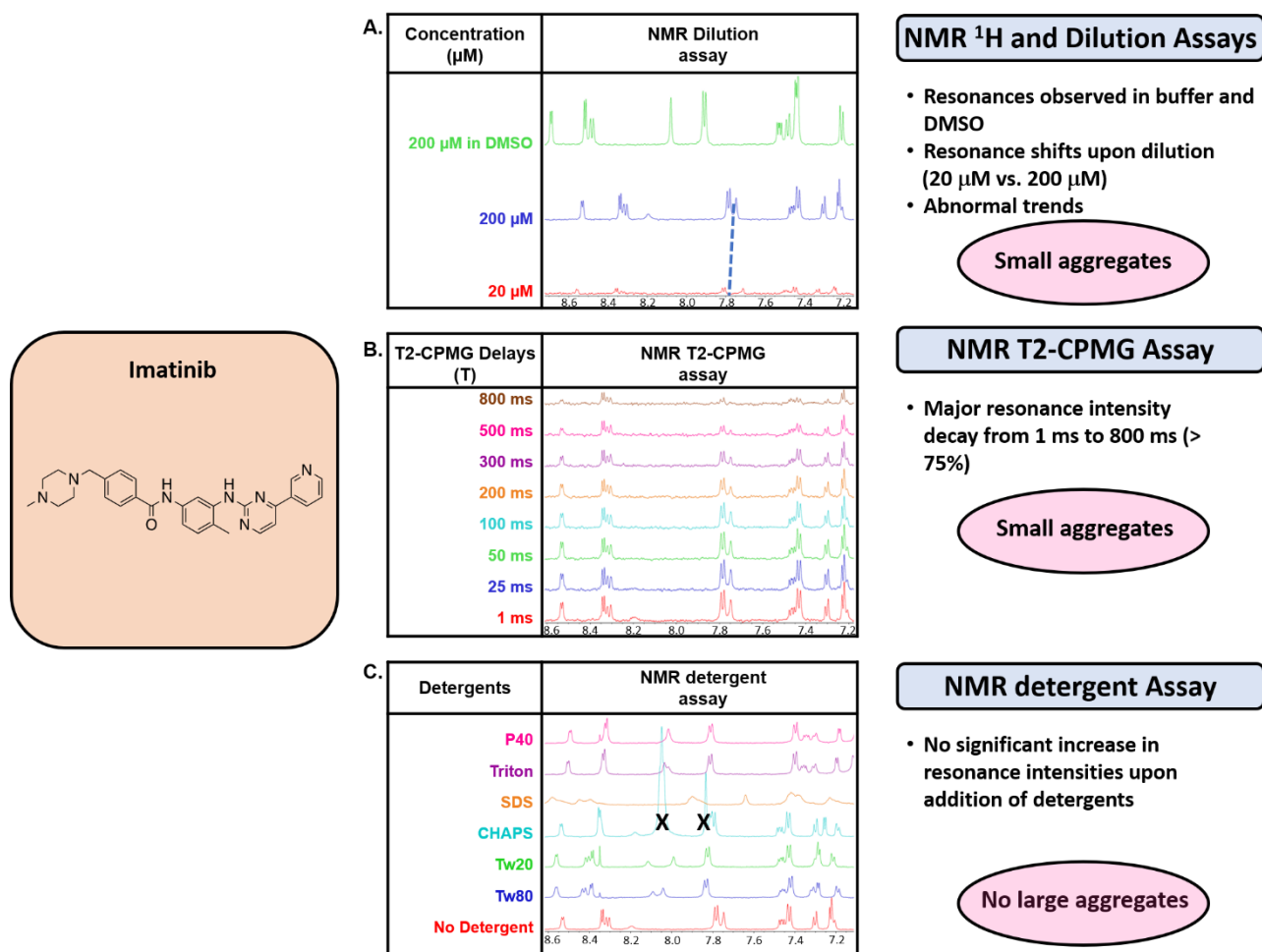
### Probing the solution behavior of riluzole



**Supplementary Figure 2 | Probing the solution behavior of riluzole.** Shown are NMR data from, (A) the NMR <sup>1</sup>H and Dilution Assays, (B) the T2-CPMG Assay, and (C) The NMR Detergent Assay. The chemical structure is shown on the left. “x” in (C) denote resonances that arise from the detergent and not from the compound.

## Probing the solution behavior of imatinib

NMR resonances were observed for imatinib in the NMR  $^1\text{H}$  Assay acquired in buffer and DMSO- $d_6$  at 200  $\mu\text{M}$  nominal concentration (Supplementary Figure 3A). Again, the samples were clear with no precipitate upon visual inspection. Dilution in buffer from 200  $\mu\text{M}$  to 20  $\mu\text{M}$  results in observation of abnormal trends (i.e. Supplementary Figure 3A - decreases in intensity along with changes in chemical shifts). This would be consistent with a behavior of self-association into small aggregates. The NMR T2-CPMG Assay (Supplementary Figure 3B) supports these conclusions as resonance intensities were significantly reduced after 800 ms (intensity decay > 75%). Interestingly, no significant increases in resonance intensities were noted upon addition of detergents in the NMR Detergent Assay (Supplementary Figure 3C) reporting that no large aggregates exist. Thus, the data suggests that imatinib self-associates into small aggregates and not into large aggregates.

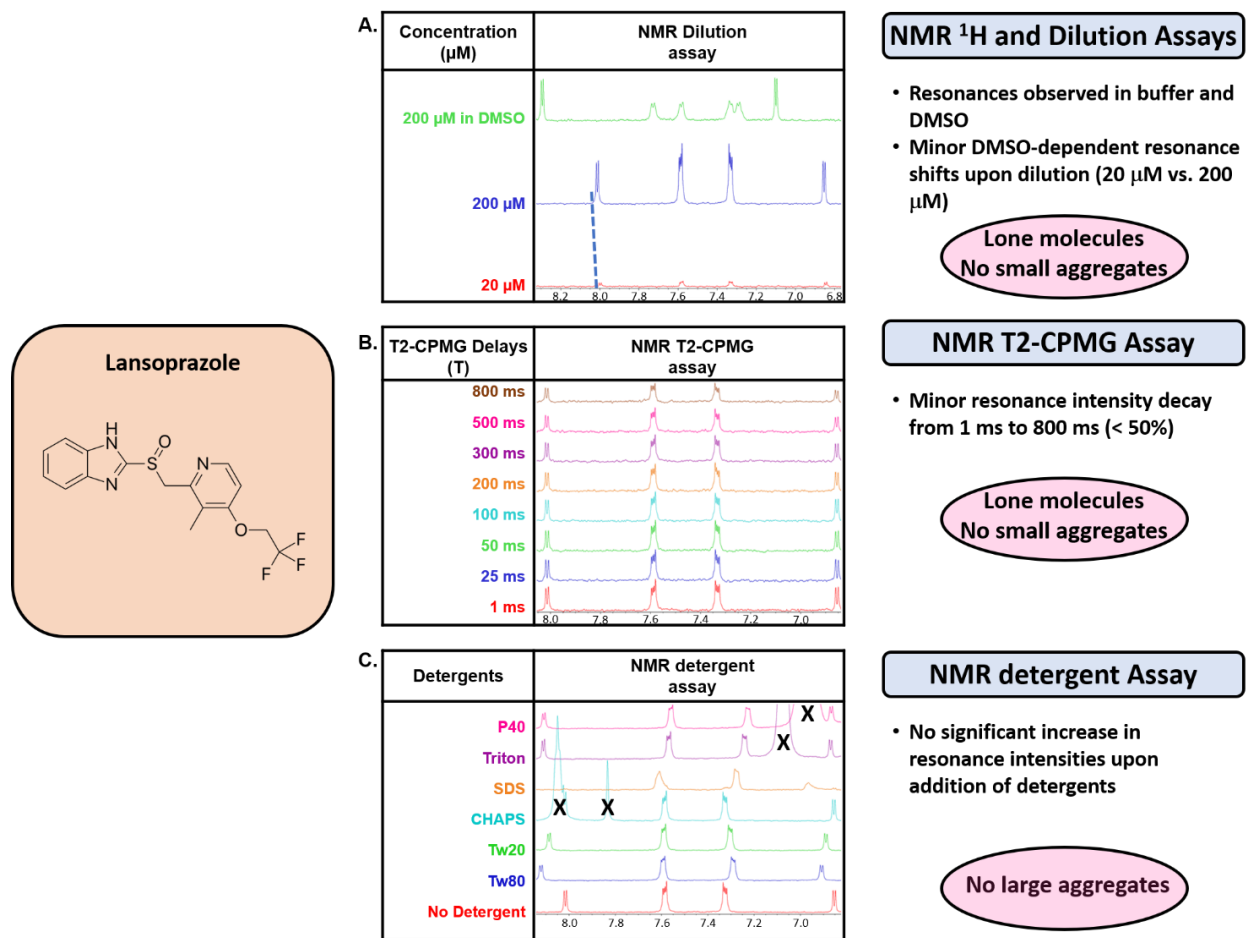


**Supplementary Figure 3 | Probing the solution behavior of imatinib.** Shown are NMR data from, (A) the NMR  $^1\text{H}$  and Dilution Assays, (B) the T2-CPMG Assay, and (C) The NMR Detergent

Assay. The chemical structure is shown on the left. "x" in (C) denote resonances that arise from the detergent and not from the compound.

### **Probing the solution behavior of lansoprazole**

Here is a particular example where caution must be taken upon interpretation of the data. T2-CPMG and detergent assays of lansoprazole do not suggest any aggregation phenomenon. However, upon looking at the dilution assay, small changes in chemical shifts can be observed for some resonances upon dilution from 200 to 20  $\mu$ M. For the sake of simplicity and speed, the dilution assay involves diluting the samples while also diluting the amount of DMSO at the same time. Therefore, some cases as this one may arise where observed shift can be attributed to such variation in DMSO concentration. Retesting of lansoprazole dilutions with constant amount of DMSO results in no observable changes in chemical shifts (data not shown). The fact that only some lansoprazole resonances would shift upon dilution can be an indication that care must be taken with interpretation. It is likely that some hydrogens are more sensitive to the changes in chemical environment caused by this variation in DMSO. Although, it is conceivable that some aggregates could also preferentially induce changes in chemical shifts for specific resonances. This highlights the importance of looking at the data as a whole.

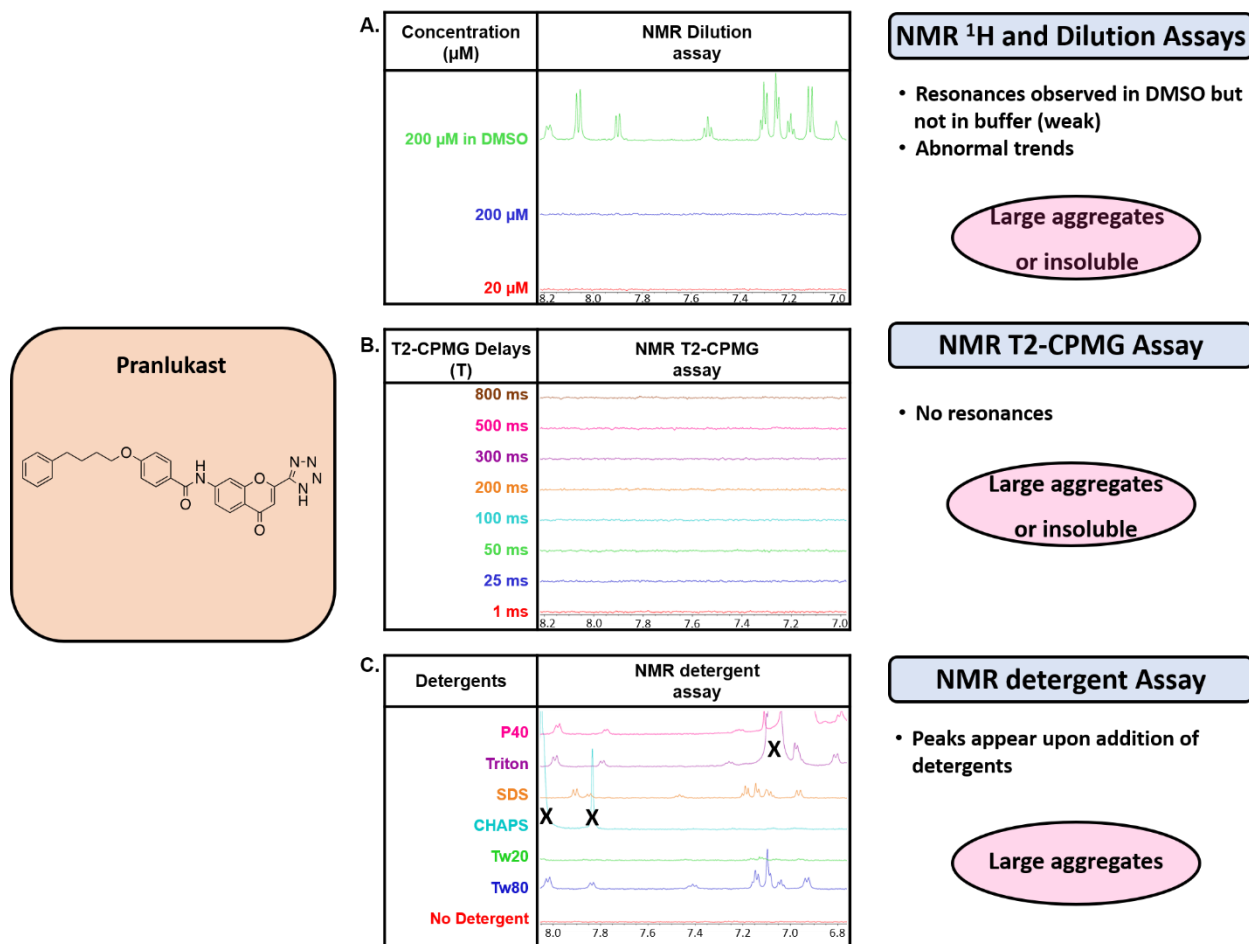


**Supplementary Figure 4 | Probing the solution behavior of lansoprazole.** Shown are NMR data from, (A) the NMR <sup>1</sup>H and Dilution Assays, (B) the T2-CPMG Assay, and (C) The NMR Detergent Assay. The chemical structure is shown on the left. “x” in (C) denote resonances that arise from the detergent and not from the compound.

### Probing the solution behavior of pranlukast

NMR resonance for pranlukast were observed in DMSO-d<sub>6</sub>, but not in buffer at 200 μM nominal concentration in the NMR <sup>1</sup>H Assay (Supplementary Figure 5A). However, samples in buffer had cloudiness upon inspection but no precipitate, confirming that the compound was still in solution. Dilution in buffer from 200 μM to 20 μM did not result in appearance of signal but resulted in significant decrease in sample cloudiness without any observable precipitate. The NMR T2-CPMG Assay therefore did not provide any information due to the lack of resonances. The NMR

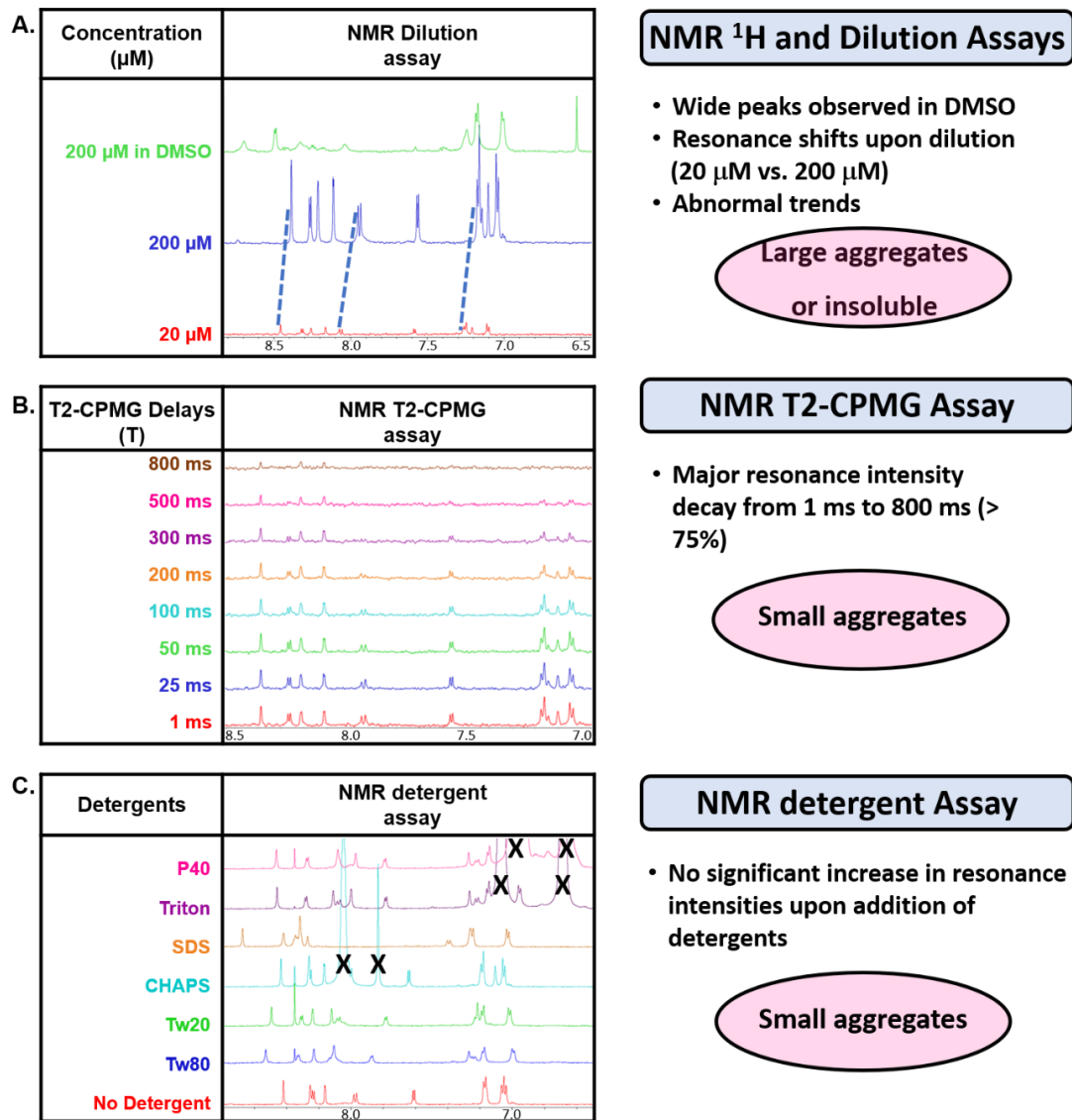
Detergent Assay confirms the presence of large aggregates since the addition of several detergents results in appearance of resonances (Supplementary Figure 5C).



**Supplementary Figure 5 | Probing the solution behavior of pranlukast.** Shown are NMR data from, (A) the NMR <sup>1</sup>H and Dilution Assays, (B) the T2-CPMG Assay, and (C) The NMR Detergent Assay. The chemical structure is shown on the left. “x” in (C) denote resonances that arise from the detergent and not from the compound.

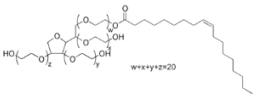
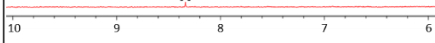
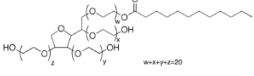
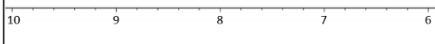
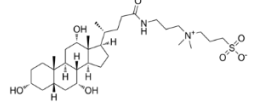
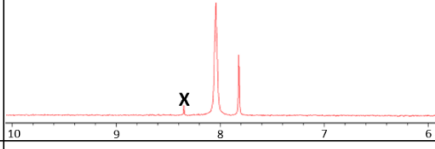
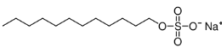
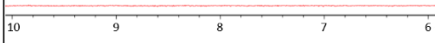
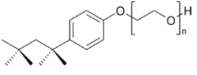
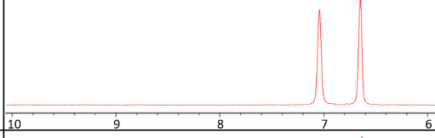
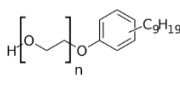
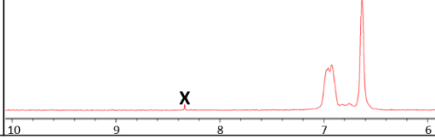
### **Example of a possible DMSO aggregator**

NMR resonances for an undisclosed compound were observed in both DMSO-d<sub>6</sub> and buffer. However, the NMR profile in DMSO would suggest that the compound is aggregating (broad peaks, low signal intensity) in this solvent (Supplementary Figure 6A). Although the NMR peaks appear much sharper in buffer, the dilution from 200 μM to 20 μM results in changes in chemical shifts for all the observed aromatic resonances. No precipitate or cloudiness could be visible in the samples. The NMR T<sub>2</sub>-CPMG Assay supports the dilution results as the signals exhibit a significant decay after 800 ms (Supplementary Figure 6B). Finally, no significant increase in resonance intensities could be observed in the NMR Detergent Assay (Supplementary Figure 6C). The overall data suggests the presence of small aggregates in buffer, with apparent aggregation of the molecule in DMSO as well.



**Supplementary Figure 6| Example of a compound that possibly aggregates in both buffer and DMSO.** Shown are NMR data from, (A) the NMR <sup>1</sup>H and Dilution Assays, (B) the T2-CPMG Assay, and (C) The NMR Detergent Assay. The chemical structure is shown on the left. “x” in (C) denote resonances that arise from the detergent and not from the compound.

## Detergents properties

Detergent	Chemical structure	Class	CMC	Spectrum (aromatic region)
Tween 80		Non-ionic	0.028 mM	
Tween 20		Non-ionic	0.042 mM	
CHAPS		Zwitterionic	8-10 mM	
SDS		Ionic	6-8 mM	
Triton X-100		Non-ionic	0.2 mM	
Nonidet P-40		Non-ionic	0.29 mM	

**Supplementary Table 1 | Properties of detergents used in this report.** “x” in NMR spectra denote resonances that arise from the impurities and not from detergents.



## Experimental guidelines for various sample set sizes

Number of compounds	Time per compound	Methods
1 - 20	<ul style="list-style-type: none"> <li>• <b>Preparation time:</b> 30-40 min</li> <li>• <b>NMR time:</b> For one sample               <ul style="list-style-type: none"> <li>• T2 CPMG : 5.5 min (4 scans)</li> <li>• <sup>1</sup>H NMR : 4 min (16 scans), 7 min (32 scans), 13.3 min (64 scans) and 26 min (128 scans)</li> </ul> </li> <li>• <b>Analysis time:</b> 40-50 min</li> </ul>	<ul style="list-style-type: none"> <li>• <b>NMR dilution assay:</b> All concentrations</li> <li>• <b>NMR T2-CPMG assay:</b> All delay times</li> <li>• <b>NMR detergent assay:</b> 3 detergents or more (e.g. Tween 80, NP40 and CHAPS)</li> <li>• <b>Note:</b> The number of scans should be same for all dilution experiments. However they can be 16, 32, 64, 128 or more (for dilution and detergent assays)</li> </ul>
20 - 50	<ul style="list-style-type: none"> <li>• <b>Preparation time:</b> 15-20 min</li> <li>• <b>NMR time:</b> For one sample               <ul style="list-style-type: none"> <li>• T2 CPMG : 5.5 min (4 scans)</li> <li>• <sup>1</sup>H NMR : 4 min (16 scans) or 7 min (32 scans)</li> </ul> </li> <li>• <b>Analysis time:</b> 30 min</li> </ul>	<ul style="list-style-type: none"> <li>• <b>NMR dilution assay:</b> Fast track method</li> <li>• <b>NMR T2-CPMG assay:</b> All delay times</li> <li>• <b>NMR detergent assay:</b> 1 or 2 detergents only (e.g. Tween 80 and NP40)</li> <li>• <b>Number of scans</b> (dilution and detergent assays): 16 or 32</li> <li>• <b>Note:</b> The number of scans should be same for all dilution experiments. However they can be 16 or 32 (for dilution and detergent assays)</li> </ul>
> 50	<ul style="list-style-type: none"> <li>• <b>Preparation time:</b> 15-20 min</li> <li>• <b>NMR time:</b> For one sample               <ul style="list-style-type: none"> <li>• T2 CPMG : 5.5 min (4 scans)</li> <li>• <sup>1</sup>H NMR : 4 min (16 scans)</li> </ul> </li> <li>• <b>Analysis time:</b> 25 min</li> </ul>	<ul style="list-style-type: none"> <li>• <b>NMR dilution assay:</b> Fast track method</li> <li>• <b>NMR T2-CPMG assay:</b> 1 ms and 800 ms delays only</li> <li>• <b>NMR detergent assay:</b> 1 or 2 detergents only (e.g. Tween 80 and NP40)</li> <li>• <b>Note:</b> The number of scans should be same for all dilution experiments. 16 scans often result in satisfactory signal-to-noise (for dilution and detergent assays)</li> </ul>

**Supplementary Table 2 | Experimental guidelines according to various sample set sizes.** Presented here are suggestions in order to mitigate increases in NMR acquisition time with larger sample sets.

## Compounds information

Compounds	SMILES	Supplier	Catalog No.	CAS	Drug class	References in the manuscript
Valsartan	<chem>CCCCC(=O)N(C1=CC=C(C=C1)C2=CC=CC=C2C3=NNN=N3)C(C)C(=O)O</chem>	BETAPHARMA	56-02004	137862-53-4	Antihypertensive	Figures 6
Etodolac	<chem>CCC1=C2C(=CC=C1)C3=C(N2)C(OCC3)(CC)CC(=O)O</chem>	SIGMA	E0516-50MG	41340-25-4	NSAID	Figure S1
Riluzole	<chem>C1=CC2=C(C=C1)OC(F)(F)F)SC(=N2)N</chem>	SIGMA	R116-250MG	1744-22-5	Amyotrophic lateral sclerosis treatment	Figure S2
Methylene blue	<chem>CN(C)C1=CC2=C(C=C1)N=C3C=C(C=[N+](C)C)C=C3S2.[Cl-]</chem>	SIGMAALDRICH	M9140-25G	7220-79-3	Antidote for methemoglobinizant poisoning	Figure 7
Imatinib	<chem>CC1=C(C=C(C=C1)NC(=O)C2=CC=C(C=C2)CN3CCN(CC3)C)NC4=NC=CC(=N4)C5=CN=CC=C5</chem>	BETAPHARMA	86-33437	152459-95-5	Chemotherapy	Figure S3
Lansoprazole	<chem>CC1=C(C=CN=C1)CS(=O)C2=NC3=CC=CC=C3N2)OCC(F)(F)F</chem>	SIGMA	L8533-250MG	103577-45-3	Proton pump inhibitor	Figure S4

Pranlukast	<chem>C1=CC=C(C=C1)CCCCOC2=CC=C(C=C2)C(=O)N(C3=CC=CC4=C3OC(=CC4=O)C5=NNN=N5)</chem>	BETAPHARMA	56-05418	10317 7-37-3	Antiasthmatic agent (anti-inflammatory)	Figure S5
Candesartan Cilexetil	<chem>CCOC2=NC1=CC=CC(=C1[N]2CC3=CC=C(C=C3)C4=CC=CC=C4C5=N[N]N=N5)C(=O)OC(C)OC(=O)OC6CCCCC6</chem>	BETAPHARMA	14-31650	14504 0-37-5	Antihypertensive	Figure 8
Lapatinib	<chem>CS(=O)(=O)CCN(C1=CC=C(O1)C2=CC3=C(C=C2)N=CN=C3NC4=C(C=C(C=C4)OCC5=CC(=CC=C5)F)Cl</chem>	Ontario Chemicals Inc	L1034	23127 7-92-2	Chemotherapy	2, 10
Clofazimine	<chem>CC(C)N=C1C=C2C(=NC3=CC=CC=C3N2C4=CC=C(C=C4)Cl)C=C1N(C5=CC=C(C=C5)Cl</chem>	Sigma	C8895	2030- 63-9	Anti-leprosy	2
Sorafenib	<chem>CNC(=O)C1=NC=CC(=C1)OC2=CC=C(C=C2)NC(=O)NC3=CC(=C(C=C3)Cl)C(F)(F)F</chem>	SynChem	BAM66401	47520 7-59-1	Chemotherapy	2
Curcumin	<chem>COC1=C(C=CC(=C1)C=CC(=O)CC(=O)C=CC2=CC(</chem>	Sigma	C1386	458- 37-7	NA	2

	<chem>=C(C=C2)O)OC</chem> O					
--	--------------------------------	--	--	--	--	--

**Supplementary Table 3 | Compounds information.**



## 6 ARTICLE 3 – IMPLEMENTATION AND VALIDATION OF AN NMR-CENTRIC PLATFORM TO GUIDE EARLY LEAD OPTIMIZATION

---

### A Robust Strategy for Hit-to-Lead Discovery: NMR for SAR

Sacha T. Larda<sup>1,‡</sup>, Yann Ayotte<sup>1,2,‡</sup>, Maria Denk<sup>1,2</sup>, Paul Coote<sup>1,3</sup>, Gregory Heffron<sup>1,3</sup>, David Bendahan<sup>2</sup>, Fatma Shahout<sup>2</sup>, Nicolas Girard<sup>1</sup>, Mustapha Iddir<sup>2</sup>, Patricia Bouchard<sup>1</sup>, Francois Bilodeau<sup>1</sup>, Simon Woo<sup>1,2</sup>, Luc J. Farmer<sup>1,2</sup>, Steven R. LaPlante<sup>1,2\*</sup>

<sup>1</sup>NMX Research and Solutions Inc., Laval, Canada

<sup>2</sup>INRS – Centre Armand-Frappier Santé Biotechnologie, Laval, Canada

<sup>3</sup>Harvard Medical School, Boston, Massachusetts, United States.

\*Corresponding author

‡These authors contributed equally

Published in : Journal of Medicinal Chemistry (In press).

DOI : 10.1021/acs.jmedchem.3c00656.

#### Authors contributions :

Sacha Larda and Yann Ayotte contributed equally and are co-first authors.

The concept of scoring ligand-observed effects was brought forward by Sacha Larda and Steven Laplante. Most of the NMR experiments were designed and performed by Yann Ayotte, with support from Sacha Larda, Maria Denk and David Bendahan. Protein plasmids designs, expression and purification were done by Yann Ayotte. Medicinal chemistry design was done by Luc Farmer, Simon Woo, and Sacha Larda. Synthesis of some in-house compounds was done by David Bendahan. SPR and MST experiments were performed by Sacha Larda, Maria Denk and Mustapha Iddir. The nucleotide release assay was implemented and performed by Yann Ayotte. Cellular assays were performed by Fatma Shahout. The manuscript was written by Sacha Larda with support from Yann Ayotte. Manuscript figures were generated by Yann Ayotte and Sacha Larda. Revision of the manuscript was done by Sacha Larda, Yann Ayotte, Steven Laplante, Simon Woo, and Luc Farmer. Software was developed by Nicolas Girard, Paul Coote, Sacha Larda and Yann Ayotte. Project management was done by Sacha Larda, Luc Farmer, Yann Ayotte, and Simon Woo. Implementation of methodologies was done by Sacha Larda, Nicolas Girard, Yann Ayotte, Gregory Heffron, Luc Farmer, and Simon Woo. Funding acquisition was done by François Bilodeau, Steven Laplante and Sacha Larda. Project administration was done by Sacha Larda, Steven Laplante, François Bilodeau, Patricia Bouchard, Yann Ayotte, Luc Farmer, and Simon Woo.

## ABSTRACT

Establishing robust structure-activity relationships (SAR) is key to successful drug discovery campaigns, yet often remains elusive due to screening and hit-validation artifacts (false positives, false negatives) which frequently result in unproductive downstream expenditures of time and resources. To address this issue, we developed an integrative biophysics-driven strategy that expedites hit-to-lead discovery, mitigates false positives/negatives and common hit-validation errors, and provides a robust approach to obtain accurate binding and affinity measurements. The advantage of this method is that it vastly improves clarity and reproducibility for affinity-driven SAR by monitoring and eliminating confounding factors. We demonstrate the ease at which high quality micromolar binders can be generated from initial millimolar fragment screening hits against an "undruggable" protein target, HRas.

## INTRODUCTION

Most early stage drug discovery efforts begin with one or more compound screening campaigns. These may include high throughput biochemical screens (HTS), virtual / computational screens, or biophysics-based screens such as X-ray crystallography, surface plasmon resonance (SPR), nuclear magnetic resonance (NMR), differential scanning fluorimetry (DSF), etc. The objective of HTS and virtual screening efforts is to identify relatively strong binders to a given disease-relevant target (protein or nucleic acid). In contrast, many biophysics-based screening approaches tend to identify relatively weak binding ligands, such as fragments, that are subsequently elaborated and optimized across repeated rounds of analoging to improve potency. The challenge, no matter what type of screening strategy taken, is that most of these approaches are susceptible to high numbers of false positives which can confound not only initial hit identification, but also follow-up validation studies by misleading interpretations of structure-activity relationships. The result is that significant resources can often be invested on non-productive hit-optimization pathways. In addition, most follow-up studies for initial hits involve orthogonal evaluation of the hit potency or affinity, often followed by efforts to obtain a structure of the ligand bound to the target in order to support or mount rational design efforts at the very early stages of a program. Applying these criteria onto initial screening hits can be a significant problem given the aforementioned false positives along with the fact that many of the initial hits from most biophysical screening campaigns are weak binders, have poor solution behavior, or are identified under different solution conditions than what is needed for biophysical follow-up characterization studies.

An alternative approach, which aims to circumvent the previously mentioned challenges, is to establish early on a biophysics-based SAR cycle that closely monitors and tightly controls factors that contribute to false positives, negatives, and common issues that arise in screening and hit-validation assays. This is achieved by comprehensive monitoring of the ligand solution behavior, aqueous solubility, structure,

and stability, whilst simultaneously evaluating target stability and conformation in presence of the ligand. This platform aims to provide a rapid and robust framework from which to evaluate systematic improvements in potency of the binders via simple, sensitive one-dimensional nuclear magnetic resonance (NMR) experiments. We demonstrate our NMR for SAR biophysics strategy with an “undruggable” protein target, HRas, and show the ease with which high quality  $\mu\text{M}$  binders can be developed from an initial  $\sim 7\text{-}10$  mM fragment screening hit.

Harvey-RAS (HRas) belongs to the Ras family of small GTPases which activate the RAS–RAF–MEK–ERK pathway. HRas functions as molecular switch in signaling pathways involved in cell growth, proliferation and differentiation.<sup>1,2</sup> As an oncogene, HRas mutations predominate in bladder cancer and neck squamous cell carcinoma with substitutions on residues G12, G13 and Q61 being the most common.<sup>3</sup> Ras family proteins are notoriously difficult to drug with one of the challenges being few and shallow binding pockets to target.<sup>4</sup> Currently no drugs that directly inhibit HRas have been approved. The lead-optimizable protein-protein interaction (PPI) inhibitors we developed are shown to inhibit a key interaction between Ras and a physiologically relevant binding-partner, SOS (Son-of-Sevenless). We demonstrate this mechanism *via* a functional assay which monitors disruption of SOS-mediated nucleotide exchange.

Our strategy is unlike previously described SAR by NMR approaches,<sup>5</sup> in that our method: i) is built upon a core set of one-dimensional NMR experiments which afford vastly higher throughput, ii) focuses on monitoring ligand solution behavior, structure, and solubility - key to the quality control process and facilitate robust interpretations of SAR, iii) captures a much larger set of information that describe both the ligand and target, and how they interact (all using the same sample), iv) is not limited by target size given that it does not rely exclusively on multi-dimensional NMR methods, nor does it require labeled protein, and finally v) requires far less material – one can perform robust experiments at volumes of only 200  $\mu\text{L}$  and at target protein concentrations of 10-15  $\mu\text{M}$ .

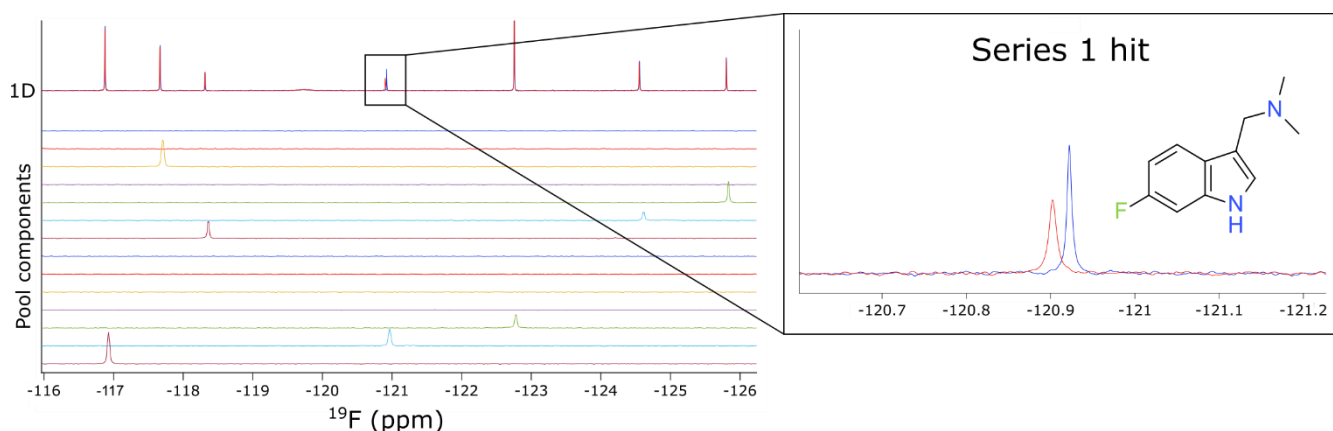
From  $\sim 200$  compounds, our approach has delivered inhibitors that serve as suitable entry points for subsequent lead-optimization (LO). Overall, the NMR for SAR strategy is an efficient means by which to fast-track early stage drug discovery efforts for a wide variety of protein targets and has been vetted within the context of contract research organization (CRO) services provided to  $>70$  pharma/biotech clients.



## RESULTS AND DISCUSSION

### $^{19}\text{F}$ Fragment Screen

The fluorine nucleus is extremely sensitive to chemical environment, making it ideal as a probe for ligand binding interactions,<sup>6,7</sup> especially for difficult to drug targets such as Ras-family GTPases. The NMX Classic Fluorine ( $^{19}\text{F}$ ) library<sup>8,9</sup> of 461 compounds was screened as cocktails of 11-15 compounds per pool against GDP-HRas<sup>G12V</sup> (hereafter “HRas”). The initial screen yielded a low hit rate, which was expected given the shallowness of available surface pockets for this target, and the magnitude of the binding effects were thus very subtle even for the best screening hit. The screen was therefore repeated at a higher protein concentration of 50  $\mu\text{M}$  (instead of the initial 15  $\mu\text{M}$ ), highlighting one of the advantages of fragment screening by  $^{19}\text{F}$  NMR: namely, that there is no interference from signal arising from the protein in  $^{19}\text{F}$  NMR. In contrast, this high protein concentration would have affected the reliability of screening by  $^1\text{H}$  NMR due to protein resonance overlap with the ligand. The fragment screens were performed in a matter of  $\sim 5$  hrs each and led to the identification of 14 hits by  $^{19}\text{F}$  differential line width (DLW) (Figure 1). All 14 hits were submitted for subsequent testing as singletons and two of those hits exhibited binding scores (magnitude of changes in line broadening (DLW) between compounds in the absence and presence of protein) that justified more thorough follow-up.



**Figure 1. Identification of initial hits from a  $^{19}\text{F}$  NMR fragment screen against GDP-HRas<sup>G12V</sup>.** 1D  $^{19}\text{F}$  spectra of a pool of fragments in the absence (blue) and the presence (red) of protein. Below is shown the spectrum of each fragment present in the pool. A zoom is presented on the right for the series 1 hit.

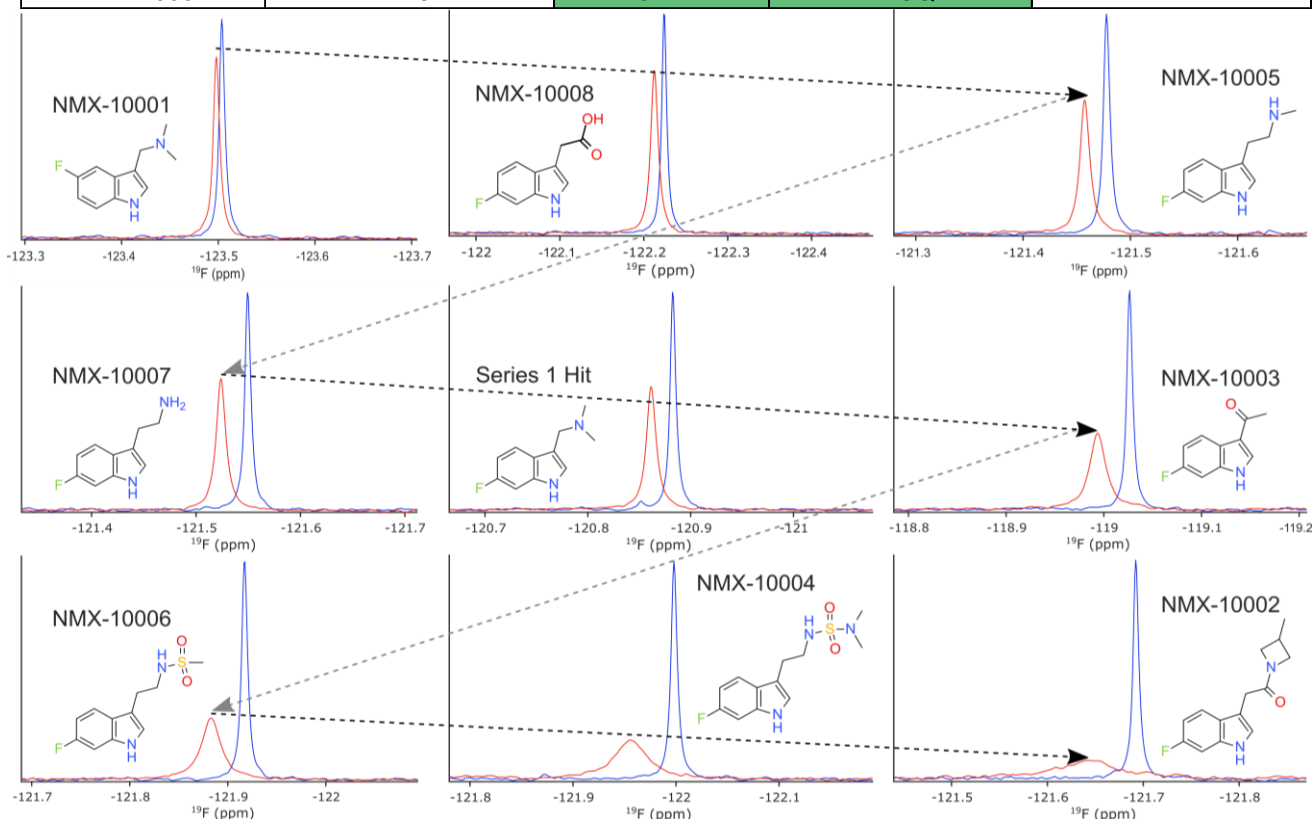
Note that screening by  $^{19}\text{F}$  NMR was crucial to the success of this program as no binding was observed by ligand-detected  $^1\text{H}$  NMR experiments for the initial hits (Figure S1). This phenomenon can be attributed to the relative sensitivity of the fluorine nucleus, which is far more susceptible to subtle differences in chemical environment and relaxation effects than proton. It should be noted, however, that the NMR for SAR strategy is equally feasible using  $^1\text{H}$  ligand-detected NMR experiments when the

sensitivity of the assay permits - the use of fluorine was necessitated here due to the shallowness of the HRas binding pocket which resulted in very weak initial fragment hits under the conditions used.

## Early Analoging

The best binding fragment was used as a template for subsequent analoging efforts. The first round of analogs was procured from commercially available sources and one of the two identified binders (series 1 discussed here) was prioritized for further elaboration to explore SAR. Figure 2 summarizes some of the results obtained using our strategy. Binding scores for both  $^{19}\text{F}$  DLW and  $^{19}\text{F}$  T2-CPMG ligand-detected NMR methods were extracted and used to monitor improvements in binding affinity and guide interpretations of SAR. Ranking of compounds by the magnitude of ligand-detected changes in this way is valid as the scores are directly related to the fraction of bound ligand provided the fast-exchange condition is satisfied<sup>10-13</sup>. Ligand concentrations are directly measured by  $^1\text{H}$  NMR acquired on the same samples using the ERETIC method<sup>14</sup> (which provides an electronic concentration reference), in order to control for variations between nominal and true ligand concentration in aqueous solution. Concentrations were measured for both the free ligand samples and the samples containing ligand and protein where appropriate. Compound aggregation is also monitored on each sample within this workflow using a previously reported  $^1\text{H}$  T2-CPMG assay<sup>15,16</sup>.

Compound ID	Measured conc. ( $\mu\text{M}$ )	DLW score	T2-CPMG score	Aggregator
NMX-10001	172	0.6	0.7	No
NMX-10010	118	1.6	1.8	No
NMX-10009	145	1.3	2.3	No
NMX-10008	193	2.0	3.4	No
NMX-10007	162	2.2	5.3	No
Series 1 Hit	199	2.5	6.4	No
NMX-10011	152	1.7	6.4	No
NMX-10005	154	2.2	7.8	No
NMX-10003	193	6.6	15.6	No
NMX-10006	182	8.6	27.9	No
NMX-10004	149	14.6	31.0	No
NMX-10002	140	31.4	>LOQ	No



**Figure 2. Systematic improvements in relative binding affinity.** Top: Table reporting measured concentration, aggregation status and binding scores extracted for each ligand. Protein concentration was 50  $\mu\text{M}$  for early analoging and subsequently reduced to 15  $\mu\text{M}$  for later NMR binding assays. Note that the ligand:protein ratio is key to accurate interpretations of rank order. The measured concentration was taken into consideration throughout each round of analoging. Aggregator status was established using a previously reported NMR assay<sup>14</sup>. The T2-CPMG score is above the limit of quantitation for NMX-10002 (exceeds maximum binding score) at the concentrations tested. Bottom: 1D  $^{19}\text{F}$  NMR data obtained for a series of analogs showing sequential improvement of relative binding affinity, with free ligand spectra (blue) and ligand with protein (red).

## **Solution Behavior**

Accurate affinity measurements are often confounded by, and are highly susceptible to, differences between nominal and true solution concentrations of the ligand. The importance of this cannot be overstated as the majority of dose-response assays performed in laboratories employ nominal ligand concentrations which often do not accurately reflect the realities of compound solution behavior or true stock solution concentrations<sup>15-17</sup>. By precisely measuring true aqueous ligand concentrations throughout a titration (Figure S2), we can significantly reduce experimental error and correct dose-response curves such that they accurately reflect the affinity of the ligand for the target of interest. Failure to correct dose-response curves according to true aqueous concentrations can lead to large errors in common biophysical affinity measurements and often mislead interpretations of relative affinities and SAR.

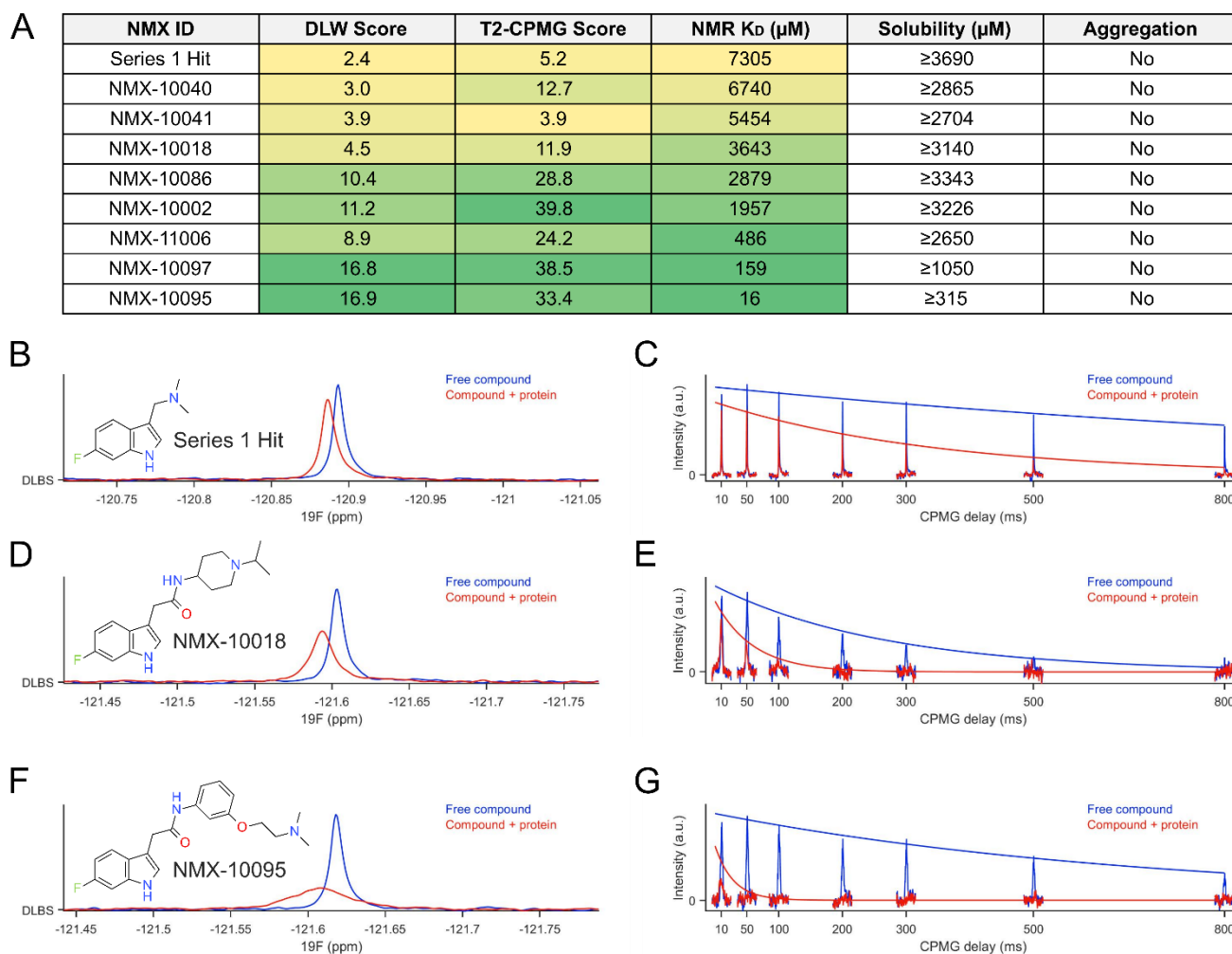
## **Orthogonal Biophysical Methods for $K_D$**

To validate our strategy,  $K_D$  values were also orthogonally obtained for select compounds using NMR, microscale thermophoresis (MST), and SPR, demonstrating excellent agreement with the relative rank order established by the ligand-detected NMR data (Figure 3A & S7).

## **$^1\text{H}$ Ligand- and Protein-Detected NMR**

Using the same NMR samples as for the  $^{19}\text{F}$  ligand-detected NMR experiments, we also recorded both  $^1\text{H}$  1D ligand- and  $^1\text{H}$  1D protein-detected NMR spectra to: i) accurately quantify ligand concentration in solution, and ii) monitor changes in the protein conformation in the presence of ligand, respectively. Compounds that induced larger magnitude changes in the protein spectrum were prioritized. Likewise, by tracking systematic differences in the protein “fingerprint”, we were able to obtain dose-response curves upon ligand titration (Figure S4).

The 1D protein-detected NMR experiments have the added advantage of providing exquisite insight into protein stability, as protein precipitation, aggregation, or loss (significant sources of false positives or negatives in binding assays) can be easily monitored and controlled for each sample. The protein requirements are no different from our standard ligand-detected NMR experiments ( $\sim 15\ \mu\text{M}$ ) and take 11 minutes to complete with usual one-dimensional proton NMR techniques, using unlabeled protein.

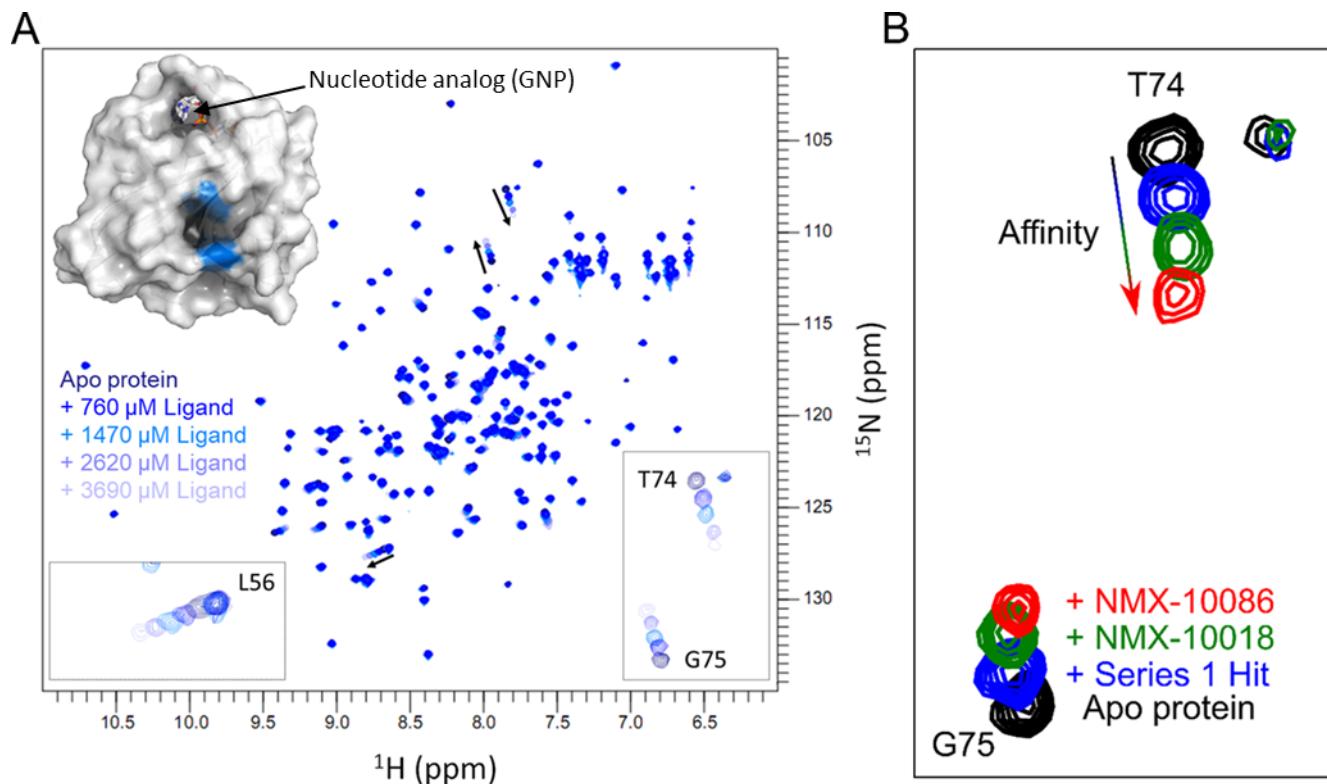


**Figure 3. The NMR for SAR Strategy** (A) Rank ordering by ligand-detected NMR correlates with affinity constants determined by NMR. Maximum solubility is also assessed by NMR on compounds selected for follow-up in order to guide subsequent experiments. (B-G) Examples of  $^{19}\text{F}$  ligand-detected methods used to rank the compounds in (A). (B, D, F)  $^{19}\text{F}$  DLW and (C, E, G)  $^{19}\text{F}$  T2-CPMG spectra for series 1 hit, 10018 and 10095, respectively. DLW: Differential line width; CPMG: Carr-Purcell-Meiboom-Gill. Protein concentration used was 15  $\mu\text{M}$ . DLW scores approaching 17 and T2-CPMG scores approaching 40 are near the limit at which subtle differences in relative affinity can be discerned, under the conditions tested. This can be ameliorated by changing the ligand:protein ratio to make the assay more or less sensitive, for resolution of weaker or stronger binders, respectively. Conditions were optimized for evaluating binding across a wide range of relative affinities.

## 2D NMR for Monitoring Binding Site

Note that while the NMR for SAR workflow can be successfully employed to drive improvements in potency in the absence of structural information, we opted to employ 2D NMR to confirm the binding site for the most promising compounds from each round of analoging. This was not a requirement but did offer orthogonal validation for the hits and ensured that the binding site was preserved throughout SAR optimizations. Backbone amide chemical shift assignments were available for HRas<sup>G12V</sup> and were readily transferred to the 2D  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC<sup>18</sup>. Chemical shift perturbations (CSP) were then used to map the

binding site of the compounds (Figure 4 (A)) which was found to be in agreement with a conserved Ras-GTPase pocket, previously reported for KRas<sup>19–22</sup>.

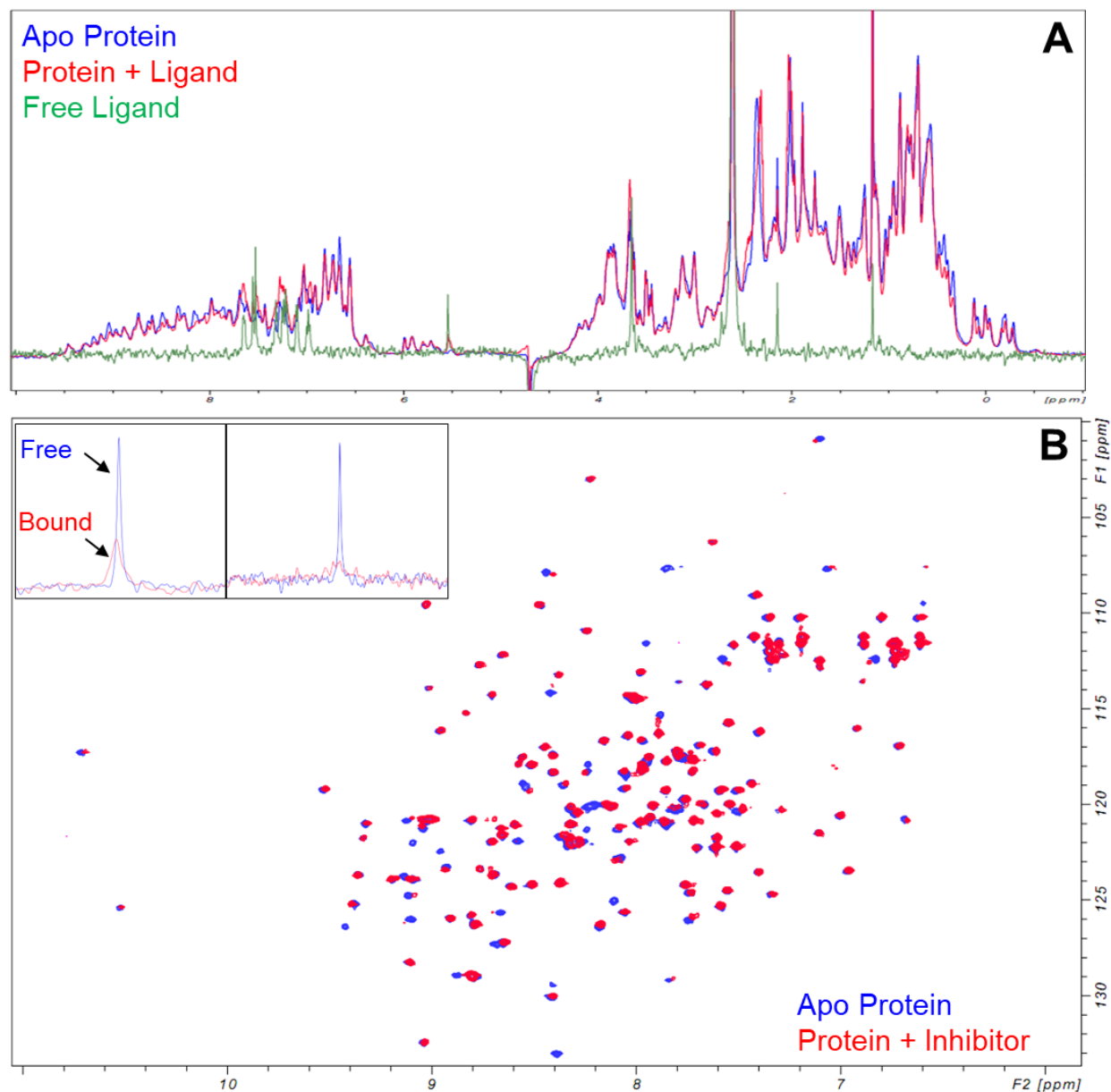


**Figure 4. Assessment of binding site and affinity ranking by 2D NMR.** (A) 2D  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC spectra reveal the binding site of the initial screening hit, as protein is tested against increasing concentrations of ligand. Three backbone residues showing the largest CSPs, indicated by arrows, are highlighted in blue on the protein surface (PDB:3OIW). Measured ligand concentrations: 0, 760, 1470, 2620, and 3690  $\mu\text{M}$ . The nucleotide binding pocket, which can be observed at the top of the protein structure, is occluded and inaccessible to small molecule ligands due to the presence of either GDP or GTP under physiological conditions. (B) Magnified regions of 2D  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC spectra for analogs with the best binding scores from each round of analoging. Compounds were tested at the same ligand-to-protein ratios (5:1 depicted here) to evaluate both ligand binding site and to prioritize based on magnitude of target engagement.

Following each round of optimization, the solution properties of the best scoring analogs were further evaluated across a range of concentrations to ensure that they would be amenable for titration studies by 1D and 2D NMR, as well as other orthogonal biophysical experiments. Based on the results of these experiments, the maximum concentration used for the compound in any of the subsequent assays was set to maintain good compound behavior in solution; that is, concentrations were used at which there was no evidence of aggregation by T2-CPMG (example application in Figure S3) and where linear relationships between nominal and observed concentrations were maintained (Figure S2).

To prioritize compounds that would be tested by subsequent 2D NMR dose-response titrations (and for orthogonal biophysical validation),  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC experiments were performed with a fixed compound

concentration to achieve equivalent ligand-to-protein ratios (Figure 4B) across the analog series. Compounds that exhibited the largest chemical shift perturbations (CSPs) in these comparative experiments were then prioritized for  $K_D$  determination by NMR.



**Figure 5. Identification of a low micromolar binder.** (A) <sup>1</sup>H protein-detected NMR changes for a relatively potent low  $\mu$ M binder to HRas. Apo protein spectrum in blue, protein + ligand spectrum in red, free ligand spectrum in green. (B) <sup>15</sup>N, <sup>1</sup>H HSQC spectrum of apo HRas (blue) and protein with inhibitor at equimolar stoichiometry (red) demonstrating significant conformational perturbation. Inlays show ligand-detected <sup>19</sup>F 1D spectra of free ligand (blue) vs ligand + protein (red) where significant broadening of the ligand signal is consistent with intermediate chemical exchange. Leftmost inlay corresponds to a ligand:protein ratio of  $\sim$ 7:1 whereas the rightmost inlay corresponds to a ligand:protein ratio of 1:1. At lower stoichiometry of ligand:protein, the broadening of the ligand resonance is greater.

## **K<sub>D</sub> by 1D and 2D NMR**

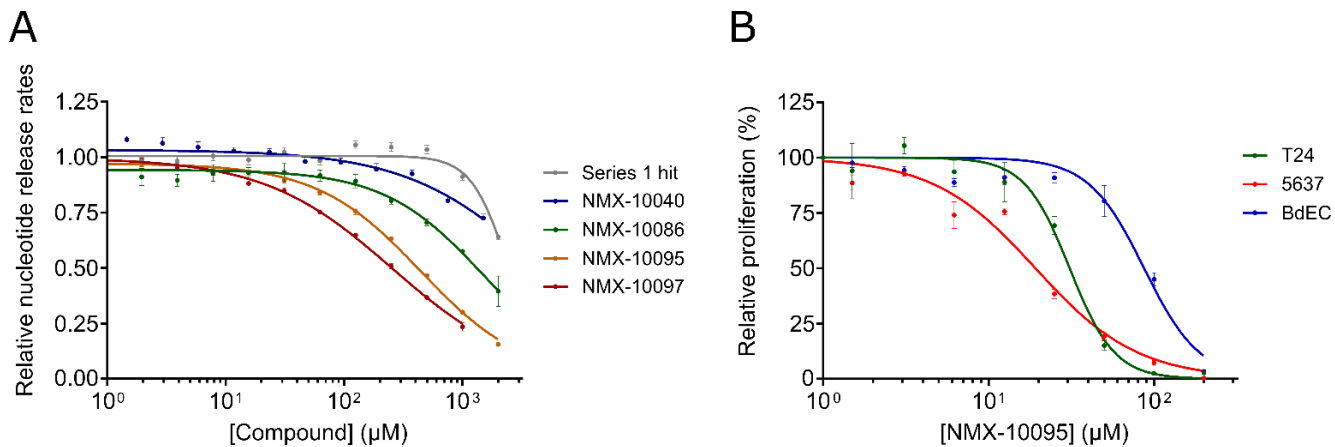
K<sub>D</sub> measurements were obtained via both i) 1D NMR by monitoring changes in <sup>1</sup>H upfield methyl protein resonances (Figure S4), and ii) 2D NMR (where appropriate) (Figure S5).

The lower protein concentration requirements for <sup>1</sup>H 1D protein-observed experiments reduces the concentration of ligand needed to achieve sufficient saturation of the protein and can therefore be used to measure K<sub>D</sub> across a wide range of affinities (μM-mM). The use of lower protein concentrations also permits K<sub>D</sub> evaluation for more poorly soluble compounds. For low μM to pM binders, <sup>1</sup>H 1D protein-detected NMR can also be used as a way of validating productive binding to the target simply by monitoring changes in the protein spectrum “fingerprint” relative to the apo-protein (Figure 5). These experiments have the added advantage of not requiring isotopically-labeled protein and can therefore be done using protein expressed from mammalian cells, yeast, bacteria, or any other expression system. These <sup>1</sup>H 1D protein-detected experiments were performed in parallel to the 1D NMR ligand-observed binding assays (on the same samples).

## **Biochemical / Functional Assay**

Systematic improvements in binding affinity were thus established and ultimately led to the development of low μM-nM binders toward a key binding site that is important for a physiologically relevant binding event between SOS and HRas. Ligands were tested according to a previously established functional assay that monitors SOS-mediated exchange of fluorescently labeled (BODIPY) nucleotide<sup>20,23</sup>. We demonstrate that our binders are able to act as functional PPIs by impeding SOS-mediated nucleotide exchange (Figure 6A). Compounds that were most active in the nucleotide release assay were tested in cell proliferation assays. Figure 6B illustrates that NMX-10095 exhibits differential inhibition of cell proliferation between T24 and 5637 bladder cancer cell lines and healthy bladder epithelial cells (BdEC). These results suggest that this early lead compound possesses some selectivity against cancerous cells over their healthy counterparts.





**Figure 6. Evaluation of the effect of compounds on biochemical and cellular activity.** (A)  $\text{SOS}^{\text{cat}}$ -catalyzed nucleotide release assay showing potency of compounds. Experiments were performed in duplicate with 500 nM  $\text{SOS}^{\text{cat}}$  and 1  $\mu\text{M}$   $\text{HRas}^{\text{G12V-BODIPY-GDP}}$  and plots show the release rates normalized to that of the DMSO control. (B) Anti-proliferative activity of NMX-10095 against T24 (bladder carcinoma), 5637 (bladder carcinoma), and BdEC (healthy, primary bladder epithelial) cell lines. Inhibition curves are normalized to the DMSO control. Error bars: SEM.

## **Distinctive Features of the NMR for SAR Strategy**

Our approach to compound screening and hit-validation satisfies a number of key criteria that are often overlooked by more traditional strategies: i) By evaluating early SAR (mM to  $\mu$ M range), we can establish clear systematic changes in binding affinity and build confidence in productive ligand-target engagement, ii) By monitoring the solution behavior and structure of the ligand at all stages of follow-up, we ensure that the binding event is not a false positive (i.e. aggregation induced, or degraded compound), iii) The accuracy and reproducibility of affinity measurements is significantly improved by correcting nominal concentrations for true aqueous solubilities, iv) By building up from fragment-sized scaffolds, this approach affords a direct forward-driven approach to improving potency. This is in contrast to HTS hits which are often potent (whether productive binders or simply non-specific) but are usually non-optimized (poor ligand efficiency) and may need to be de-constructed into the minimal binding scaffold, v) The target is monitored under truly "label-free" conditions (unlike many other biophysical methods that require tethering of the target or ligand or fluorophore labeling; which can alter the behavior of the protein target relative to its non-immobilized / non-tagged state)<sup>24</sup>, vi) The techniques employed here provide direct, atomic-level insight into conformational changes in the target protein upon ligand binding and can be simultaneously used to monitor target stability, even for unlabeled protein at low concentrations, and vii) The approach is higher throughput (comparable to other biophysical methods), requiring only a reference sample and a test sample per compound whilst providing a far greater wealth of information on both the compound and target than is offered by orthogonal techniques - all within <1 hr of experiment time per hit/analog.

## **CONCLUSIONS**

Using the NMR for SAR strategy, we have been successful in guiding medicinal chemistry efforts to improve binding affinity from an initial ~7-10 mM fragment screening hit to potent  $\mu$ M binders showing PPI inhibition between HRas and SOS as well as selective inhibition of cancerous cell proliferation. The compounds bind to an allosteric site in HRas and disrupt SOS-mediated nucleotide release, which is critical for cycling of Ras between its inactive and active conformations. Key to our approach is precise control and monitoring of compound solution behavior, aqueous solubility, and target conformational changes; accurate relative affinity

measurements; and robust high-throughput 1D NMR assays. Nearly all measurements are done on single matched pair samples of free ligand and ligand with protein, affording clear insight into true binding events and structure-activity-relationships. We demonstrate that the relative binding affinities obtained by ligand-detected NMR binding assays agree exquisitely well with the affinity rank order obtained using orthogonal biophysical measurements and, in combination with 1D protein-detected NMR, provide much greater insight and control than traditional strategies for ranking by affinity or potency. This strategy has been developed specifically to address many of the challenges encountered in early stage drug discovery and screening, both for academic and industrial clients alike. We believe this approach can significantly expedite hit-to-lead discovery and can help mitigate attrition and waste expenditures for a wide range of drug discovery programs.

## **EXPERIMENTAL SECTION**

### **Fragment library**

Fragment screening was performed using a library containing 461 fragments that had been designed within Rule of Three constraints<sup>25</sup> and filtered to remove undesirable members, including common pan-assay interference compound (PAINS) substructures.<sup>26</sup> All fragments have been curated by <sup>1</sup>H and <sup>19</sup>F NMR for structure verification, purity, solubility in phosphate buffer and lack of apparent aggregation under these conditions. Fragments were pooled based on chemical compatibility and sufficient separation of <sup>19</sup>F NMR signals, resulting in 31 different pools to increase screening throughput. The library was provided by NMX Research and Solutions Inc. (<https://www.nmxresearch.com/>). Further information about this library is available through Key Organics (<https://www.keyorganics.net/bionet-products/fragment-libraries/>) as the BIONET Fluorine Fragment Library.

### **NMR**

NMR experiments were performed with the following hardware: 600 MHz Bruker Avance III NMR spectrometer equipped with a QCI HFCN helium cryoprobe, SampleJet, and ATMA autotune system. All samples were prepared in 3 mm Bruker SampleJet tubes to volumes of 180-200  $\mu$ L with assay buffer (25 mM sodium phosphate pH 7.4, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM TCEP-d<sub>16</sub>, 10% D<sub>2</sub>O). DMSO-d<sub>6</sub> content was kept constant at 3.6%.

<sup>19</sup>F screening was performed using the standard <sup>19</sup>F 1D NMR pulse sequence zgfhigqn.2 (1D sequence for F-19 observe with inverse gated H-1 decoupling). Protein concentration was 50 μM and fragment pools were added such that each ligand (~11-15 per pool) was ~240 μM in solution. <sup>19</sup>F 1D Ligand-Detected NMR was performed using the zgfhigqn.2 sequence. Protein concentration was 50 μM for the initial round of analoging but was subsequently reduced to 15 μM when the magnitude of ligand binding effects (line broadening) exceeded that observed for compound 10002.

The <sup>19</sup>F T2-CPMG sequence was a modified version of a 1D CPMG pulse sequence with broadband excitation and refocusing shaped pulses.

The <sup>19</sup>F excitation and refocusing pulses were generated using the gradient ascent pulse engineering (GRAPE) framework for designing NMR shaped pulses.<sup>27</sup> Specifically, we followed the methodology that is outlined in Coote et al., 2017 for generating <sup>19</sup>F T2-CPMG pulses tailored to specific sets of NMR resonance frequencies.<sup>28</sup>

We optimized the two pulses to achieve high-fidelity excitation and refocusing, respectively, for the populated spectral locations (NMR peak positions) present in our <sup>19</sup>F fragment library. An ensemble of resonance frequencies was selected based on our in-house library characterization data. Each pulse was optimized for the desired point-to-point rotation (for excitation) or universal-rotation (for refocusing) over the selected ensemble of resonance frequencies. These pulses are not intended to function correctly for other resonance frequencies that are not populated in our fragment library.

Pulse durations are 300 usec (excitation) and 600 usec (refocusing), with average RF amplitude of 8.5 kHz and peak amplitude of 22.7 Khz. The peak amplitude is equivalent to an 11 μsec hard pulse excitation and within hardware specifications.

<sup>1</sup>H 1D Protein-Detected NMR was performed using the zgesgp pulse sequence.<sup>29</sup> Since these experiments were performed on the same sample used for the ligand detected experiment for each compound, the protein concentration matched those (typically 15 μM, but 50 μM for early analogs).

2D <sup>1</sup>H, <sup>15</sup>N NMR experiments were performed using the fHSQC sequence.<sup>30</sup>

Software: Topspin 3.5pl7 was used to control the spectrometer, for data acquisition and processing. Figures were prepared with one or more of the following software: MestreNova, CCPNMR Analysis, Bruker Topspin, Pymol, Microsoft Excel, and NMX's in-house software packages. NMX's custom software were employed in automated and semi-supervised analysis of both fluorine screening and  $^1\text{H}/^{19}\text{F}$  ligand-binding data.

### **DLW Score**

Each  $^{19}\text{F}$  peak is fit by least-squares to a series of common NMR multiplet peak shapes corresponding to different coupling topologies (singlet Lorentzian, doublet, triplet, quartet, doublet-of-doublets, triplet-of-doublets). The other peak shapes are superpositions of singlets that are shifted left and right according to the couplings. The highest quality (lowest square-error) fit is chosen, and the others are discarded. The fit yields measurements of NMR parameters such as resonance frequency, linewidth, peak amplitude, and J-splitting(s). The  $^{19}\text{F}$  Differential Line Width (DLW) score has units of Hz and is defined as the difference of fit linewidth, in the presence vs absence of the protein. Normalization of scores can be done with respect to a known binder (control compound) but is not obligatory when comparing scores obtained within the same preparation (using identical sample conditions).

$$DLW\ Score = w_{obs} - w_{free}$$

The score is directly related to the fraction of bound ligand via the following relationship<sup>13</sup>:

$$x_b = \frac{w_{obs} - w_{free}}{w_{bound} - w_{free}}$$

Where  $x_b$  is the fraction of bound ligand,  $w_{obs}$  is the linewidth of the ligand in the presence of protein,  $w_{free}$  is the linewidth of the free ligand, and  $w_{bound}$  is the linewidth of the ligand when bound to the target protein.

Likewise, the fraction of ligand bound can be related to the measured  $T_2$  relaxation rate (and thus the  $T_2$ -CPMG score) given that the linewidth ( $w$ ) at half height (FWHM) is inversely proportional to  $T_2$  via:

$$w = \frac{1}{\pi * T_2}$$

Where  $w$  is the linewidth at half height (“full width at half maximum”).

The fraction of bound ligand is equal to the ratio of bound vs total ligand, and this can be described in terms of protein-ligand complex by:

$$x_b = \frac{[L_b]}{[L_t]} = \frac{[PL]}{[L_t]}$$

Where  $[L_b]$  is the concentration of bound ligand,  $[L_t]$  is the total concentration of ligand, and  $[PL]$  is the concentration of protein-ligand complex.

Assuming the fast exchange condition whereby ligands exhibit fast off-rates ( $k_{off}$ ), a reasonable assumption for most  $\mu\text{M}$  to  $\text{mM}$  binders, and where ligand concentrations are in significant excess relative to protein (such that  $L_f \cong L_t$ ), we can then invoke a simplified 1:1 binding model to relate the amount of protein-ligand complex ( $PL$ ) to  $K_D$ :

$$[PL] = \frac{[P_t][L_f]}{K_D + L_f} \cong \frac{[P_t][L_t]}{K_D + L_t}$$

Thus, we can relate the fraction of bound ligand to  $K_D$  by relation:

$$x_b * [L_t] = [PL] = \frac{[P_t][L_t]}{K_D + L_t}$$

In cases where the amount of free ligand does not closely approximate the total ligand ( $L_f \neq L_t$ ), a more complete solution is required:

$$[PL] = \frac{[P_t][L_f]}{K_D + L_f} = \frac{[P_t][L_t - PL]}{K_D + [L_t - PL]}$$

Solving for  $[PL]$  gives:

$$[PL] = \frac{(L_t + K_D + P_t) - \sqrt{(-P_t - L_t - K_D)^2 - 4(P_t L_t)}}{2}$$

In the same way,  $X_b$  can be related to  $K_D$  via  $[PL]$  and is therefore a direct measure of relative affinity.<sup>13</sup>

The equation that correlates chemical shift changes to  $K_D$  is written as follows (for 1:1 binding model):

For target-detected NMR<sup>31</sup>:

$$\frac{(\delta_{obs} - \delta_{free})}{(\delta_{bound} - \delta_{free})} = \frac{(L_t + K_D + P_t) - \sqrt{(P_t + L_t + K_D)^2 - 4(P_t L_t)}}{2(P_t)}$$

### **T2-CPMG Score**

The CPMG score has units of seconds and is obtained by taking the difference between the observed  $R_2$  relaxation rate for the ligand in the presence of protein and the  $R_2$  of the free ligand. As with the DLW score, normalization can be done with a known binder (control).

$$T2\text{-CPMG Score} = R_{2_{obs}} - R_{2_{free}}$$

The score is determined using two samples, one with free ligand and one with ligand in the presence of the target. For each sample, a series of NMR spectra are acquired with eight different relaxation delays (1, 25, 50, 100, 200, 300, 500, and 800 ms). NMR regions of interest (ROIs) containing peaks, multiplets or clusters of peaks are extracted. Within each ROI, the signal from the first spectrum is projected onto each of the other seven spectra, to create a series of scalar projection weights ( $M_k$  for  $k = 1-8$ ). These are fit by least-squares to the theoretical relaxation equation, i.e. a standard exponential decay curve:

$$M_k = A * \exp(-R_2 t)$$

Where  $t$  is the CPMG relaxation delay for each of the eight spectra, and  $R_2$  and  $A$  are free parameters that are determined by the fitting routine.

This is calculated separately for each ROI, and then averaged to generate an overall score for each ligand.

## 2D <sup>1</sup>H, <sup>15</sup>N HSQC

<sup>1</sup>H, <sup>15</sup>N Fast HSQC experiments<sup>30</sup> were recorded at a protein concentration of 50 or 100 μM. Dissociation constants were obtained by monitoring changes in chemical shifts as a function of ligand concentration. The weighted average changes in chemical shifts (d) were calculated according to the following equation<sup>31</sup>:

$$d = \sqrt{\frac{1}{2} [\delta_H^2 + (\alpha \cdot \delta_N)^2]}$$

Where δ are the changes in chemical shift in ppm for <sup>1</sup>H and <sup>15</sup>N and the correction factor (α) was set at 0.15.

## Biochemical Assay

Nucleotide release rates were measured using 1 μM BODIPY GDP loaded HRas<sup>G12V</sup> in 25 mM Tris, pH 7.4, 50 mM NaCl, 1 mM DTT, 10 mM MgCl<sub>2</sub>. The nucleotide release reaction was then initiated by addition of either DMSO (control) or compounds in DMSO (across a range of concentrations), SOScat and unlabeled GTP to final concentrations of 500 nM and 20 μM, respectively. DMSO content was kept constant at 3% in all conditions.

Changes in fluorescence were measured by a fluorescence spectrometer (Tecan Infinite M1000 Pro) in a black 384 well plate (Greiner). Fluorescence excitation was carried out at λ = 485 nm and emission was measured at λ = 510 nm every 30 s for 30 minutes at 28 °C. Release rates were determined by fitting the decrease in fluorescence over time to a single exponential decay. The derived rates were normalized to the DMSO treated sample and plotted against compound concentration as mean ± SEM. The IC<sub>50</sub> for each compound was calculated by fitting the normalized rates to a four-parameter dose response curve.

## Cloning, expression and purification of proteins

The codon optimized sequences for HRas<sup>G12V</sup> (aa 1-166) and human SOS1 (SOS<sup>cat</sup>, aa 564-1049) were prepared and cloned into the pET-28a(+) plasmid at GenScript (<https://www.genscript.com>). Proteins were expressed using E. coli BL21 (DE3) cells in Terrific Broth (TB) media and induced with 0.5 mM IPTG at 25 °C overnight. Cells were harvested by centrifugation (5,000 g for 20 minutes at 4°C) and cell pellets were extracted by sonication in 50 mM Tris, pH 8,



500 mM NaCl, 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP). Cell lysates were centrifuged at 20,000 g for 30 minutes at 4 °C and supernatant was diluted in TCEP-free lysis buffer to reduce TCEP concentration to 1 mM. The diluted solution was loaded onto a HisTrap HP column (GE Healthcare). The column was washed with washing buffer (50 mM Tris, pH 8, 500 mM NaCl, 1 mM TCEP) and the bound proteins were eluted using a linear gradient of elution buffer (50 mM Tris, pH 8, 500 mM NaCl, 1 mM TCEP, 300 mM imidazole). For HRas NMR studies, HRas N-terminal His-tag was cleaved by incubation with tobacco etch virus protease (TEV) overnight at 4 °C in a SnakeSkin Dialysis Tubing (Thermo Scientific) to remove excess imidazole. TEV was removed using a nickel column and the proteins were further purified on a size-exclusion (SEC) Superdex 75 column (GE Healthcare) in buffers containing 25 mM sodium phosphate, pH 7.4, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM TCEP; and 25 mM Tris, pH 7.4, 50 mM NaCl, 1 mM dithiothreitol (DTT) for HRas and SOS, respectively. Fractions containing the respective proteins were pooled, concentrated using Amicon centrifugal filters (Millipore), flash frozen in liquid nitrogen and stored at -80 °C. The purity of HRas and SOS was greater than 95% by SDS-PAGE. Uniformly <sup>15</sup>N-labeled HRas was purified using the same steps as described above but was expressed in M9 minimal media with <sup>15</sup>NH<sub>4</sub>Cl as the sole nitrogen source.

For biophysical assessments, HRas was loaded with GDP before the size-exclusion chromatography step by incubating the protein with 20 mM EDTA and 5 mM GDP at room temperature for 30 minutes. The solution was then buffer-exchanged in EDTA- and nucleotide-free buffer with 5 mM MgCl<sub>2</sub> before loading into the SEC column.

For the nucleotide release assay, HRas was buffer exchanged in MgCl<sub>2</sub>-free buffer and incubated with 20 mM EDTA and 2 mM BODIPY FL GDP (Invitrogen) for 1.5 h at room temperature. The reaction was then supplemented with 10 mM MgCl<sub>2</sub> and incubated for another 30 minutes at room temperature. EDTA and the excess nucleotides were removed by desalting the protein into 25 mM Tris, pH 7.4, 50 mM NaCl, 1 mM DTT, 10 mM MgCl<sub>2</sub>.

### **Cell lines and cell culture**

Bladder cancers (T24 and 5637) and primary bladder epithelial/normal cells (BdEC) were obtained from American Type Culture Collection (ATCC). T24 cells were cultured in McCoy's medium, and 5637 cells in RPMI-1640, supplemented with 10% FBS. BdEC cells were cultured in healthy bladder epithelial basal medium supplemented with a growth kit as recommended by the manufacturer. Cells were incubated in a humidified atmosphere of 5% CO<sub>2</sub> at 37 °C to ensure growth and viability.

### **Proliferation assay**

Antiproliferative effects were evaluated by using an MTT (3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide) assay. When cells reached 80-90% confluency, they were transferred and cultured overnight in 96 well plates, in a humidified atmosphere of 5% CO<sub>2</sub> at 37 °C. Cells were then starved with media without FBS for 4 hours. After starvation, cells were treated with various concentrations of each compound diluted in fresh media without FBS for 72h. DMSO was used as a vehicle control for each cell type for normalization of the data. After incubation, 20 µl of MTT solution (prepared at 5 mg/mL) was then added to each well and incubated for 4 h in the dark at 37 °C. The formazan crystals were then solubilized by removing the media and adding DMSO. Cell viability was determined by measuring absorbance at 570 nm and subtracting the 650 nm background absorbance. Relative proliferation for each cancer/normal cell lines was plotted using GraphPad Prism.

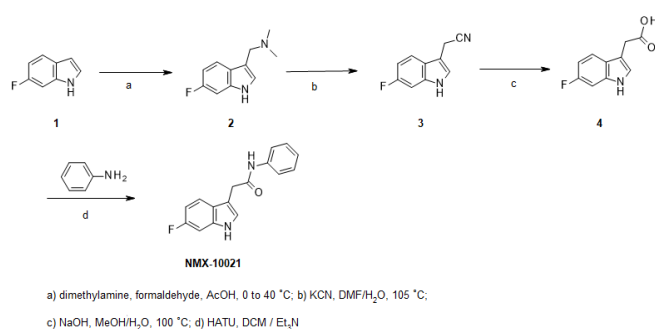
### **Compound Preparation and Characterization.**

All commercially available reagents and anhydrous solvents were used without further purification. Purity assessments for final compounds were determined by analytical HPLC: 4.6 mm × 50 mm Waters YMC Pro-C18, 5 µm column, 120A. Mobile phases were as follows: A, H<sub>2</sub>O with 0.2% formic acid; B, acetonitrile with 0.2% formic acid. Gradient: 10–90% B in 3 min with a 5 min run time. The flow rate was 1.5 mL/min. Unless specified otherwise, all compounds were ≥95% pure. Mass spectrometry samples were analyzed on a Micro Mass ZQ, ZMD, Quattro LC, or Quattro II mass spectrometer operated in a single MS mode with electrospray ionization. Samples were introduced into the mass spectrometer using flow injection (FIA) or chromatography. The mobile phase for all mass analysis consisted of acetonitrile–water mixtures with either 0.2% formic acid or ammonium formate. <sup>1</sup>H NMR spectra were recorded using either a Bruker Avance 400 (400 MHz) or a Bruker Avance II- 300 (300 MHz) instrument. Column chromatography was performed using Teledyne ISCO RediSep Normal Phase (35–70 µm) or RediSep Gold Normal Phase (25–40 µm) silica flash columns using a Teledyne ISCO Combiflash Companion or Combiflash Rf purification system. Preparative reversed-phase chromatography was carried out using a Gilson 215 liquid handler coupled to a UV–vis 156 Gilson detector and an Agilent Zorbax SB-C18 column, 21.2 mm × 100 mm. A linear gradient from 10% to 90% CH<sub>3</sub>CN in H<sub>2</sub>O over 10 min (0.1% trifluoroacetic acid) was used; the flow rate was 20 mL/min.

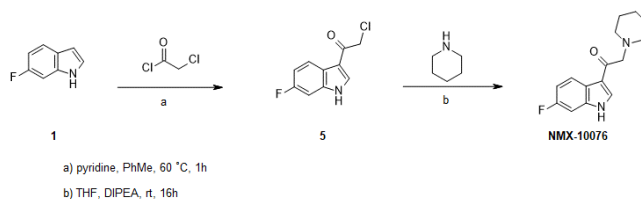
High-resolution mass spectrometry data were collected on a Thermo Scientific QExactive mass spectrometer coupled to a Waters Acquity UPLC system. Samples were analyzed from a 100 µM DMSO solution with a 3 µL injection volume. The chromatographic column was a Waters Acquity

CSH C18, 2.1 × 50 mm, 1.7 μm particle size. Gradient elution was employed using 0.1% formic acid in water as mobile phase A and 0.1% formic acid in CH<sub>3</sub>CN as mobile phase B. The gradient began at 10% B and increased to 60% B over 0.8 min and to 100% B over the next 0.2 min, followed by a 0.5 min re-equilibration at the initial conditions. The mass spectrometer was run in full MS mode, positive polarity, with the resolution set to 35 000. A heated electrospray source was used with settings of 3.5 kV and 400 °C.

All compounds shown in Figures 1 and 2, **FS-1255** and **NMX-10001** to **NMX-10011** were purchased from Key Organics Limited and Enamine Limited.



**Scheme 1. Synthesis of 2-(1H-Indol-3-yl)-N-methyl-acetamide series.**



**Scheme 2. Synthesis of 1-(1H-Indol-3-yl)-2-(amino) ethenone series.**

## Synthesis

Synthesis of indole amides such as **NMX-10021** relied on the readily accessible 2-(6-fluoro-1H-indol-3-yl) acetic acid **4**. This intermediate was prepared on multigram scale by the two-step cyanomethylation of 6-fluoro indole **1**, followed by nitrile hydrolysis to the corresponding pivotal acid intermediate **4**, as outlined in Scheme 1.

Compounds NMX- **10014**, **10016**, **10018**, **10021**, **10040**, **10041**, **10078**, **10086**, **10095** and **10097** were prepared via amide couplings under standard conditions using indole acid **4** and the corresponding amines (ex. aniline) in the presence of HATU in dichloromethane as shown for compound **NMX-10021** (Scheme 1).

Indole ketone **NMX-10076** was prepared according to Scheme 2. For the synthesis of **NMX-10076**, 3-acylation of 6-fluoro indole **1** with 2-chloroacetyl chloride afforded the chloromethyl ketone intermediate **5**, which upon N alkylation with piperidine furnished the desired analog **NMX-10076**.

#### **1-(6-fluoro-1H-indol-3-yl)-N,N-dimethylmethanamine (2)**

A solution of 40 percent aq. dimethylamine (9.1 g, 80.7 mmol) was cooled to 5°C, and glacial acetic acid (6.1 mL) was added dropwise while maintaining the temperature at ~10°C. After stirring for 20 minutes, 37 percent aqueous formaldehyde (6.1 mL, 80.7 mmol) was slowly added to the above solution while keeping the temperature between 0~10°C, followed by the addition of 6-fluoroindole **1** (10 g, 74.0 mmol). The reaction was exothermic and reached a final temperature of ~40°C, and it was then cooled down to ~20°C. The reaction solution was then slowly added to 160 mL of aqueous NaOH solution (3M). The suspension was stirred for about 30 minutes, and then collected by filtration. The cake was rinsed with water (50mL X 2) and dried to afford a yellow solid (12.1 g, > 99%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.22 (br s, 1H), 7.63-7.59 (m, 1H), 7.08-7.07 (m, 1H), 7.03-7.00 (m, 1H), 6.91-6.86 (m, 1H), 3.60 (s, 2H), 2.28 (s, 6H).

#### **2-(6-fluoro-1H-indol-3-yl)acetonitrile (3)**

A solution of 1-(6-fluoro-1H-indol-3-yl)-N,N-dimethylmethanamine **2** (12 g, 62.4 mmol) and KCN (6.7 g, 102.9 mmol) in DMF (36 mL) and water (19 mL) was heated to 105°C for 10 hours. The reaction mixture was then cooled down to 25°C, and water (145 mL) and toluene (80 mL) were added and stirred for 3 hours. The organic and aqueous layers were separated. The organic layer was washed with aqueous sodium bicarbonate (80 mL) and brine (80 mL), and then dried over sodium sulfate. After filtration, the filtrate was concentrated, and the residue was purified by flash column chromatography on silica gel to get the desired product as a yellow oil (5.7 g, 52.4% yield).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.19 (s, 1H), 7.52-7.49 (m, 1H), 7.22-7.21 (m, 1H), 7.08 (dd, *J* = 9.4, 2.2 Hz, 1H), 6.98-6.93 (m, 1H), 3.82 (d, *J* = 1.1 Hz, 2H).

#### **2-(6-fluoro-1H-indol-3-yl)acetic acid (4)**

A mixture of 2-(6-fluoro-1*H*-indol-3-yl)acetonitrile **3** (2 g, 11.5 mmol), sodium hydroxide (2.6 g, 65.0 mmol), methanol (15 mL) and water (45 mL) was stirred at 100 °C for 3 hours. The reaction was then cooled to 0°C and acidified with 6 N aqueous HCl to pH~1. The solid formed was collected by filtration, which was then washed twice with water and dried to give the title compound **4** as a yellow solid (1.7 g, 76.6% yield).

MS (ESI+) *m/z* 194 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 12.18 (s, 1H), 10.97 (s, 1H), 7.50-7.46 (m, 1H), 7.23-7.22 (m, 1H), 7.12 (dd, *J* = 10.2, 2.3 Hz, 1H), 6.87-6.82 (m, 1H), 3.63 (s, 2H).

**General procedures for amide compounds:** (See Scheme 1)

#### **2-(6-fluoro-1*H*-indol-3-yl)-*N*-phenylacetamide (NMX-10021)**

A solution of 2-(6-fluoro-1*H*-indol-3-yl)acetic acid **4** (100 mg, 0.52 mmol), aniline (53.0 mg, 0.57 mmol), HATU (216.5 mg, 0.57 mmol) and Et<sub>3</sub>N (68.1 mg, 0.67 mmol) in DCM (8 mL) was stirred at room temperature for 16 hours. The mixture was diluted with dichloromethane (30 mL), washed with HCl (15 mL, 1.0 N) and brine (20 mL), and dried over sodium sulfate. After filtration, the filtrate was concentrated, and the residue was purified by flash column chromatography on silica gel to obtain the desired product **NMX-10021** as a colorless oil (80 mg, 57.6% yield).

MS (ESI+) *m/z* 269 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 10.98 (s, 1H), 10.09 (s, 1H), 7.76-7.48 (m, 3H), 7.38-7.20 (m, 3H), 7.12 (dd, *J* = 10.2, 2.3 Hz, 1H), 7.08-6.96 (m, 1H), 6.88-6.83 (m, 1H), 3.71 (s, 2H).

Using the above procedures, the following compounds were prepared:

#### ***N*-(3-acetamidophenyl)-2-(6-fluoro-1*H*-indol-3-yl)acetamide (NMX-10014)**

MS (ESI+) *m/z* 326 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 10.97 (s, 1H), 10.09 (s, 1H), 9.90 (s, 1H), 7.90 (s, 1H), 7.61-7.57 (m, 1H), 7.41-7.02 (m, 5H), 6.87-6.82 (m, 1H), 3.70 (s, 2H), 2.02 (s, 3H).

#### **2-(6-fluoro-1*H*-indol-3-yl)-1-morpholino-ethanone (NMX-10016)**

MS (ESI+) *m/z* 263 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 8.14 (s, 1H), 7.53 (dd, *J* = 8.7, 5.3 Hz, 1H), 7.06-7.02 (m, 2H), 6.92-6.87 (m, 1H), 3.82 (s, 2H), 3.65 (s, 4H), 3.49 (s, 4H).

**2-(6-fluoro-1H-indol-3-yl)-N-(1-isopropyl-4-piperidyl)acetamide (NMX-10018)**

MS (ESI+)  $m/z$  318 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  8.70 (s, 1H), 7.45-7.42 (m, 1H), 7.15-6.99 (m, 2H), 6.92-6.87 (m, 1H), 5.54 (d,  $J$  = 7.9 Hz, 1H), 3.80-3.75 (m, 1H), 3.68 (s, 2H), 2.68-2.65 (m, 3H), 2.22-2.16 (m, 2H), 1.85-1.81 (m, 2H), 1.26-1.16 (m, 2H), 0.96 (d,  $J$  = 6.6 Hz, 6H).

**1-(4-acetylpiperazin-1-yl)-2-(6-fluoro-1H-indol-3-yl)ethanone (NMX-10040)**

MS (ESI+)  $m/z$  304 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  10.97 (s, 1H), 7.54 (s, 1H), 7.29-7.06 (m, 2H), 6.95-6.74 (m, 1H), 3.79 (s, 2H), 3.65-3.43 (m, 8H), 1.98 (s, 3H).

**N-[(3S)-1-[2-(6-fluoro-1H-indol-3-yl)acetyl]pyrrolidin-3-yl]acetamide (NMX-10041)**

MS (ESI+)  $m/z$  304 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  8.46-8.31 (m, 1H), 7.60-7.51 (m, 1H), 7.10-6.99 (m, 2H), 6.92-6.85 (m, 1H), 5.85-5.81 (m, 0.5H), 5.76-5.69 (m, 0.5H), 4.48-4.36 (m, 1H), 3.79-3.65 (m, 3H), 3.64-3.52 (m, 2H), 3.44-3.34 (m, 1H), 2.24-2.00 (m, 1H), 1.94-1.86 (m, 3H), 1.82-1.72 (m, 1H).

**2-(6-fluoro-1H-indol-3-yl)-1-(4-hydroxy-1-piperidyl)ethanone (NMX-10086)**

MS (ESI+)  $m/z$  277 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  10.95 (s, 1H), 7.58-7.49 (m, 1H), 7.22-7.16 (m, 1H), 7.15-7.05 (m, 1H), 6.89-6.75 (m, 1H), 4.66 (d,  $J$  = 4.1 Hz, 1H), 3.99-3.86 (m, 1H), 3.84-3.68 (m, 3H), 3.65-3.55 (m, 1H), 3.21-3.09 (m, 1H), 3.01-2.89 (m, 1H), 1.69-1.49 (m, 2H), 1.23-1.02 (m, 2H).

**N-(3-(2-(dimethylamino)ethoxy)phenyl)-2-(6-fluoro-1H-indol-3-yl)acetamide (NMX-10095)**

MS (ESI+)  $m/z$  356 (M+H)<sup>+</sup>.

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  10.98 (s, 1H), 10.10-10.03 (m, 1H), 7.61-7.54 (m, 1H), 7.34-7.29 (m, 1H), 7.25 (d,  $J$  = 2.3 Hz, 1H), 7.20-7.07 (m, 3H), 6.89-6.81 (m, 1H), 6.63-6.57 (m, 1H), 3.98 (t,  $J$  = 5.8 Hz, 2H), 3.70 (s, 2H), 2.59 (t,  $J$  = 5.8 Hz, 2H), 2.19 (s, 6H).

**N-[3-[[2-(dimethylamino)acetyl]amino]phenyl]-2-(6-fluoro-1H-indol-3-yl)acetamide (NMX-10097)**

MS (ESI+)  $m/z$  369 (M+H)<sup>+</sup>.

$^1\text{H}$  NMR (400 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  10.97 (s, 1H), 10.10 (s, 1H), 9.66 (s, 1H), 8.00-7.93 (m, 1H), 7.63-7.54 (m, 1H), 7.36-7.31 (m, 1H), 7.28-7.23 (m, 2H), 7.21-7.15 (m, 1H), 7.12 (dd,  $J = 10.2, 2.3$  Hz, 1H), 6.89-6.80 (m, 1H), 3.70 (s, 2H), 3.04 (s, 2H), 2.26 (s, 6H).

**Procedures for indole ketone NMX-10076:** (See Scheme 2)

### **2-chloro-1-(6-fluoro-1*H*-indol-3-yl)ethan-1-one (5)**

To a stirred solution of 6-fluoro-1*H*-indole **1** (1 g, 7.4 mmol) and pyridine (0.3 mL, 7.4 mmol) in toluene (12 mL) at 60°C was added 2-chloroacetyl chloride (0.84 g, 7.4 mmol) dropwise. After the addition was complete, the reaction mixture was stirred at 60°C for 1 hour. After cooling to room temperature,  $\text{H}_2\text{O}$  (18 mL) and MeOH (4 mL) were added. The mixture was stirred at room temperature for an additional 1 hour. The precipitate was filtered through a sintered glass funnel and washed with  $\text{H}_2\text{O}$  to obtain compound **5** as a brown solid (300 mg, 19% yield).

### **1-(6-fluoro-1*H*-indol-3-yl)-2-(1-piperidyl)ethanone (NMX-10076)**

To a solution of 2-chloro-1-(6-fluoro-1*H*-indol-3-yl)ethan-1-one **5** (200 mg, 0.94 mmol) and piperidine (120 mg, 1.42 mmol) in tetrahydrofuran (6 mL) was added DIPEA (0.3 mL, 1.9 mmol) dropwise, and the resulting mixture was stirred at room temperature for 16 hours. The solvent was removed under vacuum (<25°C) and the residue was treated with ethyl acetate (10 mL) to produce a precipitate. The precipitate was collected by filtration and was further purified by flash column chromatography to obtain the title product **NMX-10076** as a white solid.

MS (ESI+)  $m/z$  261 ( $\text{M}+\text{H}$ )<sup>+</sup>.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  8.70 (s, 1H), 8.43-8.40 (m, 1H), 8.39-8.33 (m, 1H), 7.13-7.01 (m, 2H), 3.57 (s, 2H), 2.55 (s, 4H), 1.66-1.60 (m, 4H), 1.50-1.43 (m, 2H).

## **ASSOCIATED CONTENT**

**Supporting Information.** Supplementary figures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Comparison of Proton- vs Fluorine- Ligand Detected NMR for Series 1 hit FS-1255; Compound behavior assay for monitoring differences between nominal and true solution concentrations of various ligands prior to dose-response assays; Monitoring aggregation properties of related compounds via a high-throughput 1D NMR assay; HRas apo spectrum and  $K_D$  by 1D protein observed NMR;  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC titrations; HRas stability by 1D protein-observed NMR over 96 hrs; Orthogonal biophysical data for selected HRas binders; Binding site mapping based on backbone amide chemical shift perturbations for compound 10097; Supplementary methods.

## AUTHOR INFORMATION

### Corresponding Author

\* Steven R. Laplante, [steven.laplante@inrs.ca](mailto:steven.laplante@inrs.ca)

### Author Contributions

All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

### Funding Sources

Quebec Consortium for Drug Discovery (CQDM)  
Ministère de l'Économie et de l'Innovation (MEI)  
NMX Research and Solutions Inc.

Mitacs

3 Point Bio

Sprout BioVentures

Viva

NSERC-USRA

### Notes

A patent application has been filed: WO2023060362A1.

## ACKNOWLEDGMENT

We would like to thank the following persons and institutions:

Weihong Zheng and Zhixiong Ye from Viva.

Mike Serrano-Wu and Brian Hubbard.

Andrew Lowerson and Key Organics.

## ABBREVIATIONS

NMR, Nuclear Magnetic Resonance; SAR, Structural-Activity relationship

## REFERENCES

- (1) Lu, S.; Jang, H.; Gu, S.; Zhang, J.; Nussinov, R. Drugging Ras GTPase: A Comprehensive Mechanistic and Signaling Structural View. *Chem. Soc. Rev.* **2016**, *45* (18), 4929–4952. <https://doi.org/10.1039/C5CS00911A>.
- (2) Simanshu, D. K.; Nissley, D. V.; McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **2017**, *170* (1), 17–33. <https://doi.org/10.1016/j.cell.2017.06.009>.
- (3) Cox, A. D.; Fesik, S. W.; Kimmelman, A. C.; Luo, J.; Der, C. J. Drugging the Undruggable RAS: Mission Possible? *Nat. Rev. Drug Discov.* **2014**, *13* (11), 828–851. <https://doi.org/10.1038/nrd4389>.
- (4) Lu, S.; Jang, H.; Muratcioglu, S.; Gursoy, A.; Keskin, O.; Nussinov, R.; Zhang, J. Ras Conformational Ensembles, Allostery, and Signaling. *Chem. Rev.* **2016**, *116* (11), 6607–6665. <https://doi.org/10.1021/acs.chemrev.5b00542>.



- (5) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274* (5292), 1531–1534. <https://doi.org/10.1126/science.274.5292.1531>.
- (6) Dalvit, C. NMR Methods in Fragment Screening: Theory and a Comparison with Other Biophysical Techniques. *Drug Discov. Today* **2009**, *14* (21), 1051–1057. <https://doi.org/10.1016/j.drudis.2009.07.013>.
- (7) Dalvit, C.; Flocco, M.; Veronesi, M.; Stockman, B. J. Fluorine-NMR Competition Binding Experiments for High-Throughput Screening of Large Compound Mixtures. *Comb Chem High Throughput Screen* **2002**, *5* (8), 605–611. <https://doi.org/10.2174/1386207023329923>.
- (8) Lowerson, A.; McCarren, P.; LaPlante, S. R.; Serrano-Wu, M. BIONET 2nd Generation Premium Fragment Library - Building a Diverse and Experimentally-Curated Fragment Library. <https://cdn.technologynetworks.com/TN/Resources/PDF/building-a-diverse-and-experimentally-curated-fragment-library.pdf> (accessed 2022-09-21).
- (9) R. LaPlante, S.; McCarren, P.; Bilodeau, F.; Lowerson, A.; Serrano-Wu, M. Practical Aspects of Designing the 2nd Generation BIONET Premium Fragment Library. [https://www.namiki-s.co.jp/upload/supplier2/149-supplier\\_content\\_product-001.pdf](https://www.namiki-s.co.jp/upload/supplier2/149-supplier_content_product-001.pdf) (accessed 2022-09-21).
- (10) Shortridge, M. D.; Hage, D. S.; Harbison, G. S.; Powers, R. Estimating Protein-Ligand Binding Affinity Using High-Throughput Screening by NMR. *J. Comb. Chem.* **2008**, *10* (6), 948–958. <https://doi.org/10.1021/cc800122m>.
- (11) Jordan, J. B.; Poppe, L.; Xia, X.; Cheng, A. C.; Sun, Y.; Michelsen, K.; Eastwood, H.; Schnier, P. D.; Nixey, T.; Zhong, W. Fragment Based Drug Discovery: Practical Implementation Based on <sup>19</sup>F NMR Spectroscopy. *J. Med. Chem.* **2012**, *55* (2), 678–687. <https://doi.org/10.1021/jm201441k>.
- (12) Fielding, L. NMR Methods for the Determination of Protein- Ligand Dissociation Constants. *Curr. Top. Med. Chem.* **2003**, *3* (1), 39–53. <https://doi.org/10.2174/1568026033392705>.
- (13) Fischer, J. J.; Jardetzky, O. Nuclear Magnetic Relaxation Study of Intermolecular Complexes. The Mechanism of Penicillin Binding to Serum Albumin1a. *J. Am. Chem. Soc.* **1965**, *87* (14), 3237–3244. <https://doi.org/10.1021/ja01092a040>.
- (14) Akoka, S.; Barantin, L.; Trierweiler, M. Concentration Measurement by Proton NMR Using the ERETIC Method. *Anal. Chem.* **1999**, *71* (13), 2554–2557. <https://doi.org/10.1021/ac981422i>.
- (15) Ayotte, Y.; Marando, V. M.; Vaillancourt, L.; Bouchard, P.; Heffron, G.; Coote, P. W.; Larda, S. T.; LaPlante, S. R. Exposing Small-Molecule Nanoentities by a Nuclear Magnetic Resonance Relaxation Assay. *J. Med. Chem.* **2019**, *62* (17), 7885–7896. <https://doi.org/10.1021/acs.jmedchem.9b00653>.
- (16) LaPlante, S. R.; Roux, V.; Shahout, F.; LaPlante, G.; Woo, S.; Denk, M. M.; Larda, S. T.; Ayotte, Y. Probing the Free-State Solution Behavior of Drugs and Their Tendencies to Self-Aggregate into Nano-Entities. *Nat. Protoc.* **2021**, *16* (11), 5250–5273. <https://doi.org/10.1038/s41596-021-00612-3>.
- (17) LaPlante, S. R.; Carson, R.; Gillard, J.; Aubry, N.; Coulombe, R.; Bordeleau, S.; Bonneau, P.; Little, M.; O'Meara, J.; Beaulieu, P. L. Compound Aggregation in Drug Discovery:

Implementing a Practical NMR Assay for Medicinal Chemists. *J. Med. Chem.* **2013**, *56* (12), 5142–5150. <https://doi.org/10.1021/jm400535b>.

(18) Amin, N.; Chiarparin, E.; Coyle, J.; Nietlispach, D.; Williams, G. 1H, 15N and 13C Backbone Assignments of GDP-Bound Human H-Ras Mutant G12V. *Biomol. NMR Assign.* **2016**, *10* (1), 121–123. <https://doi.org/10.1007/s12104-015-9649-4>.

(19) Buhrman, G.; Kumar, V. S. S.; Cirit, M.; Haugh, J. M.; Mattos, C. Allosteric Modulation of Ras-GTP Is Linked to Signal Transduction through RAF Kinase. *J Biol Chem* **2011**, *286* (5), 3323–3331. <https://doi.org/10.1074/jbc.M110.193854>.

(20) Maurer, T.; Garrenton, L. S.; Oh, A.; Pitts, K.; Anderson, D. J.; Skelton, N. J.; Fauber, B. P.; Pan, B.; Malek, S.; Stokoe, D.; Ludlam, M. J. C.; Bowman, K. K.; Wu, J.; Giannetti, A. M.; Starovasnik, M. A.; Mellman, I.; Jackson, P. K.; Rudolph, J.; Wang, W.; Fang, G. Small-Molecule Ligands Bind to a Distinct Pocket in Ras and Inhibit SOS-Mediated Nucleotide Exchange Activity. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (14), 5299–5304. <https://doi.org/10.1073/pnas.1116510109>.

(21) Sun, Q.; Phan, J.; Friberg, A. R.; Camper, D. V.; Olejniczak, E. T.; Fesik, S. W. A Method for the Second-Site Screening of K-Ras in the Presence of a Covalently Attached First-Site Ligand. *J. Biomol. NMR* **2014**, *60* (1), 11–14. <https://doi.org/10.1007/s10858-014-9849-8>.

(22) Kessler, D.; Gmachl, M.; Mantoulidis, A.; Martin, L. J.; Zoepfel, A.; Mayer, M.; Gollner, A.; Covini, D.; Fischer, S.; Gerstberger, T.; Gmaschitz, T.; Goodwin, C.; Greb, P.; Häring, D.; Hela, W.; Hoffmann, J.; Karolyi-Oezguer, J.; Knesl, P.; Kornigg, S.; Koegl, M.; Kousek, R.; Lamarre, L.; Moser, F.; Munico-Martinez, S.; Peinsipp, C.; Phan, J.; Rinnenthal, J.; Sai, J.; Salamon, C.; Scherbantin, Y.; Schipany, K.; Schnitzer, R.; Schrenk, A.; Sharps, B.; Siszler, G.; Sun, Q.; Waterson, A.; Wolkerstorfer, B.; Zeeb, M.; Pearson, M.; Fesik, S. W.; McConnell, D. B. Drugging an Undruggable Pocket on KRAS. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (32), 15823–15829. <https://doi.org/10.1073/pnas.1904529116>.

(23) Burns, M. C.; Sun, Q.; Daniels, R. N.; Camper, D.; Kennedy, J. P.; Phan, J.; Olejniczak, E. T.; Lee, T.; Waterson, A. G.; Rossanese, O. W.; Fesik, S. W. Approach for Targeting Ras with Small Molecules That Activate SOS-Mediated Nucleotide Exchange. *Proceedings of the National Academy of Sciences* **2014**, *111* (9), 3401–3406. <https://doi.org/10.1073/pnas.1315798111>.

(24) Hennig, M.; Ruf, A.; Huber, W. Combining Biophysical Screening and X-Ray Crystallography for Fragment-Based Drug Discovery. In *Fragment-Based Drug Discovery and X-Ray Crystallography*; Davies, T. G., Hyvönen, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp 115–143. [https://doi.org/10.1007/128\\_2011\\_225](https://doi.org/10.1007/128_2011_225).

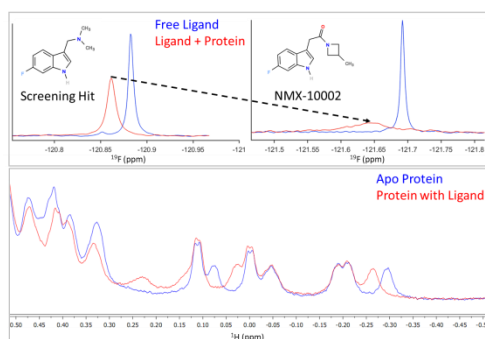
(25) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'Rule of Three' for Fragment-Based Lead Discovery? *Drug Discov. Today* **2003**, *8* (19), 876–877. [https://doi.org/10.1016/S1359-6446\(03\)02831-9](https://doi.org/10.1016/S1359-6446(03)02831-9).

(26) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53* (7), 2719–2740. <https://doi.org/10.1021/jm901137j>.

(27) Khaneja, N.; Reiss, T.; Kehlet, C.; Schulte-Herbrüggen, T.; Glaser, S. J. Optimal Control of Coupled Spin Dynamics: Design of NMR Pulse Sequences by Gradient Ascent Algorithms. *J Magn Reson* **2005**, *172* (2), 296–305. <https://doi.org/10.1016/j.jmr.2004.11.004>.

- (28) Coote, P.; Anklin, C.; Masefski, W.; Wagner, G.; Arthanari, H. Rapid Convergence of Optimal Control in NMR Using Numerically-Constructed Toggling Frames. *J. Magn. Reson.* **2017**, *281*, 94–103. <https://doi.org/10.1016/j.jmr.2017.05.011>.
- (29) Hwang, T. L.; Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson., Series A* **1995**, *112* (2), 275–279. <https://doi.org/10.1006/jmra.1995.1047>.
- (30) Mori, S.; Abeygunawardana, C.; Johnson, M. O.; Vanzijl, P. C. M. Improved Sensitivity of HSQC Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast HSQC (FHSQC) Detection Scheme That Avoids Water Saturation. *J. Magn. Reson., Series B* **1995**, *108* (1), 94–98. <https://doi.org/10.1006/jmrb.1995.1109>.
- (31) Morton, C. J.; Pugh, D. J.; Brown, E. L.; Kahmann, J. D.; Renzoni, D. A.; Campbell, I. D. Solution Structure and Peptide Binding of the SH3 Domain from Human Fyn. *Structure* **1996**, *4* (6), 705–714. [https://doi.org/10.1016/S0969-2126\(96\)00076-7](https://doi.org/10.1016/S0969-2126(96)00076-7).
- (32) Williamson, M. P. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16. <https://doi.org/10.1016/j.pnmrs.2013.02.001>.

#### Table of Contents artwork



## 7 SUPPORTING INFORMATION – ARTICLE 3

---

### A Robust Strategy for Hit-to-Lead Discovery: NMR for SAR

Sacha T. Larda<sup>1</sup>, Yann Ayotte<sup>1,2</sup>, Maria Denk<sup>1,2</sup>, Paul Coote<sup>1,3</sup>, Gregory Heffron<sup>1,3</sup>, David Bendahan<sup>2</sup>, Fatma Shahout<sup>2</sup>, Nicolas Girard<sup>1</sup>, Mustapha Iddir<sup>2</sup>, Patricia Bouchard<sup>1</sup>, Francois Bilodeau<sup>1</sup>, Simon Woo<sup>1,2</sup>, Luc J. Farmer<sup>1,2</sup>, Steven R. LaPlante<sup>1,2\*</sup>

<sup>1</sup>NMX Research and Solutions Inc., Laval, Canada

<sup>2</sup>INRS – Centre Armand-Frappier Santé Biotechnologie, Laval, Canada

<sup>3</sup>Harvard Medical School, Boston, Massachusetts, United States.

\*Corresponding author

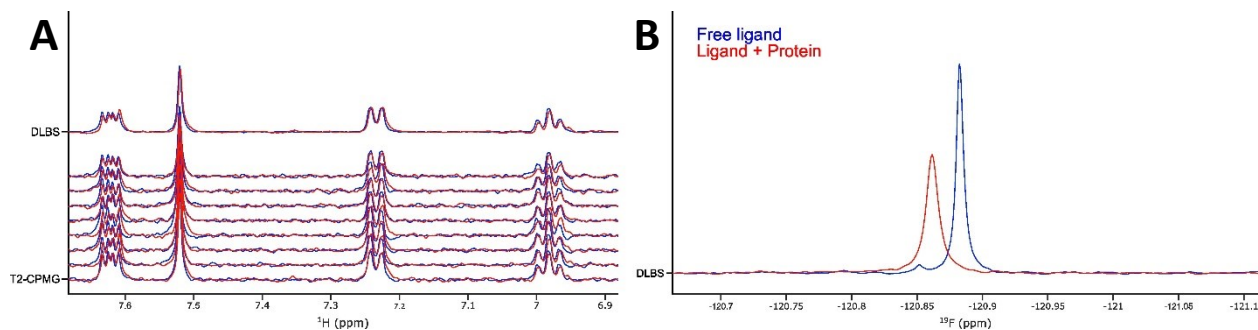
Published in : Journal of Medicinal Chemistry (In press).

DOI : 10.1021/acs.jmedchem.3c00656.

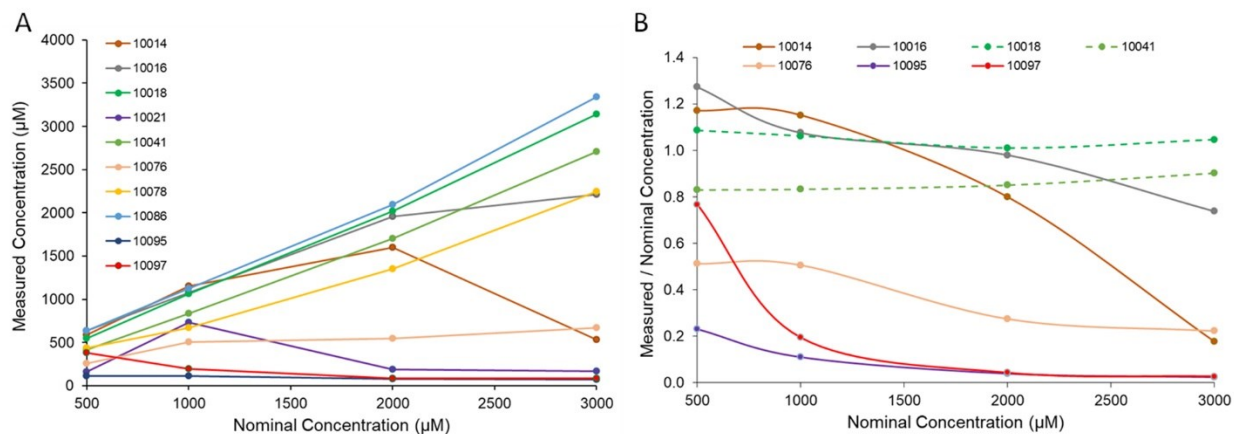
### Contents

- Supplementary figures:
  1. Comparison of Proton- vs Fluorine- Ligand Detected NMR for Series 1 hit FS-1255
  2. Compound behavior assay for monitoring differences between nominal and true solution concentrations of various ligands prior to dose-response assays
  3. Monitoring aggregation properties of related compounds via a high-throughput 1D NMR assay
  4. HRas apo spectrum and  $K_D$  by 1D protein observed NMR
  5.  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC titrations
  6. HRas stability by 1D protein-observed NMR over 96 hrs
  7. Orthogonal biophysical data for selected HRas binders
  8. Binding site mapping based on backbone amide chemical shift perturbations for compound 10097
- Supplementary methods

## Supplementary figures

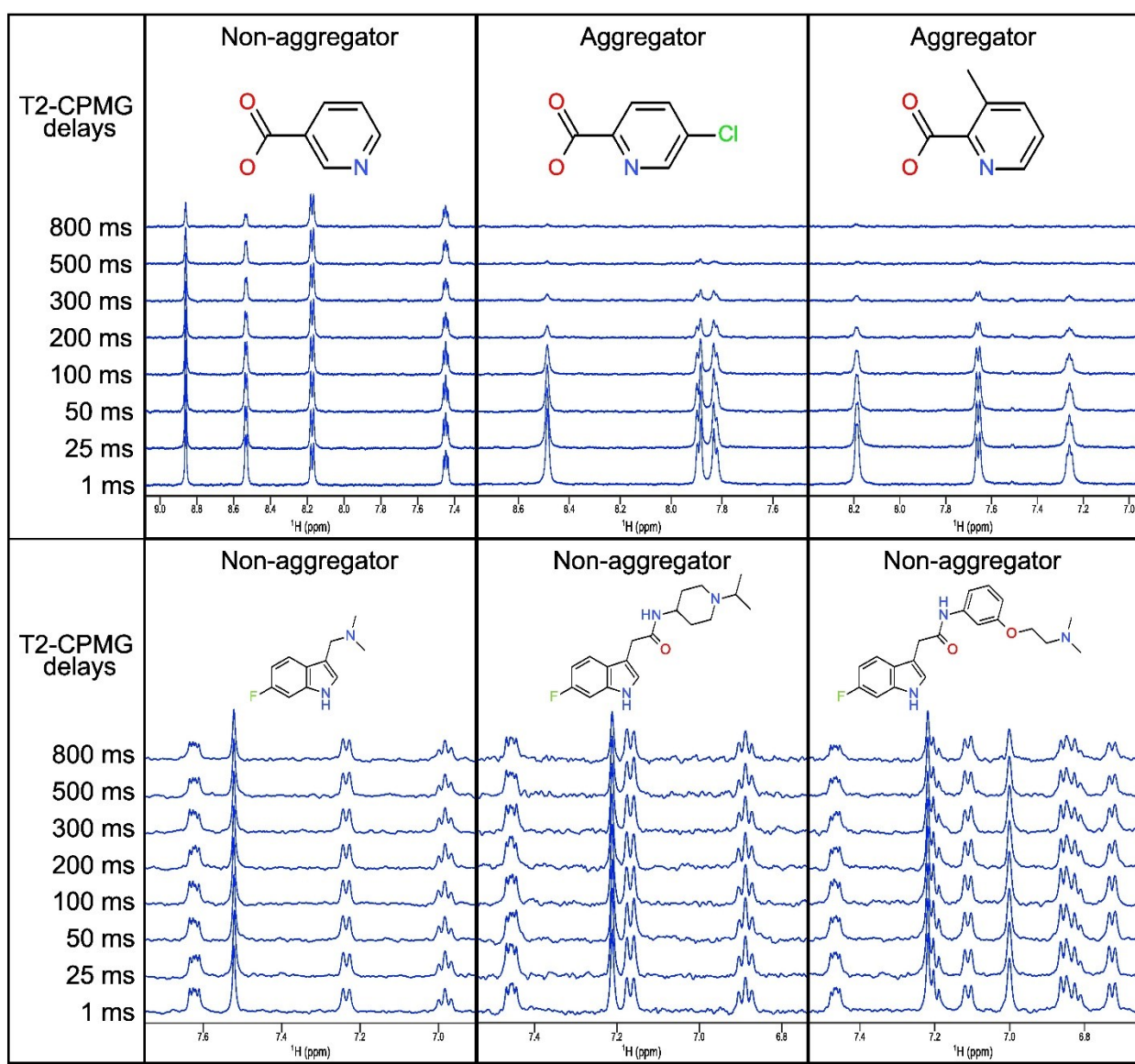


**Figure S1.** Comparison of Proton- vs Fluorine- Ligand Detected NMR for Series 1 hit FS-1255. (A)  $^1\text{H}$  Ligand-detected NMR spectra showing the differential line broadening (DLB) spectra followed by the stack of eight T2-CPMG spectra with different delay times (0, 25, 50, 100, 200, 300, 500, 800 msec). No binding is evident by  $^1\text{H}$  ligand-detected NMR. (B)  $^{19}\text{F}$  1D Ligand-detected NMR spectra demonstrating clear binding for compound FS-1255, highlighting the advantage of  $^{19}\text{F}$  NMR for elucidation of binding of weak fragments.

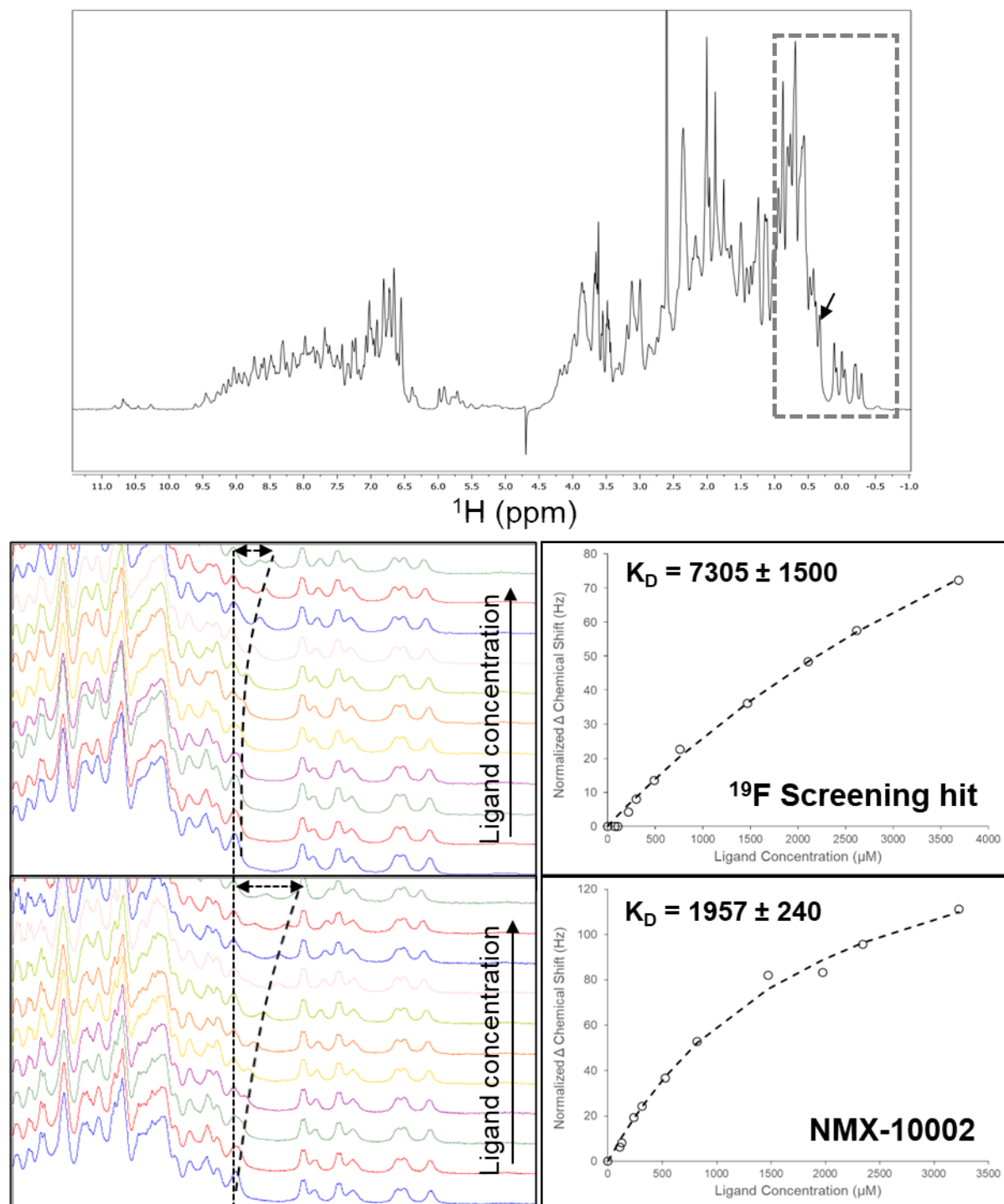


**Figure S2.** Compound behavior assay for monitoring differences between nominal and true solution concentrations of various ligands prior to dose-response assays (A). Compounds for which the ratio between measured and nominal concentration remains constant across the titration are typically well-behaved, whereas large changes in the ratio indicates potential solubility issues (B). In cases where the ratio slightly deviates from 1.0, but remains constant across a titration series (i.e. 10041, B), this is usually attributable to error in the nominal stock concentration (usually prepared by dissolving powder directly with solvent). Discrepancy between molecular

weight and formula weight, along with poor measurement accuracy and precision are common sources of error in assumed nominal stock concentrations. Large initial (low concentration point) discrepancy between nominal and measured concentrations is often simply due to limited solubility of the compound. Dashed lines in (B) correspond to well-behaved compounds across the entire titration. Note the difference between nominal and measured solubility for 10095 and 10097, illustrating the importance of working within the measured solubility limits to obtain accurate binding and affinity data.

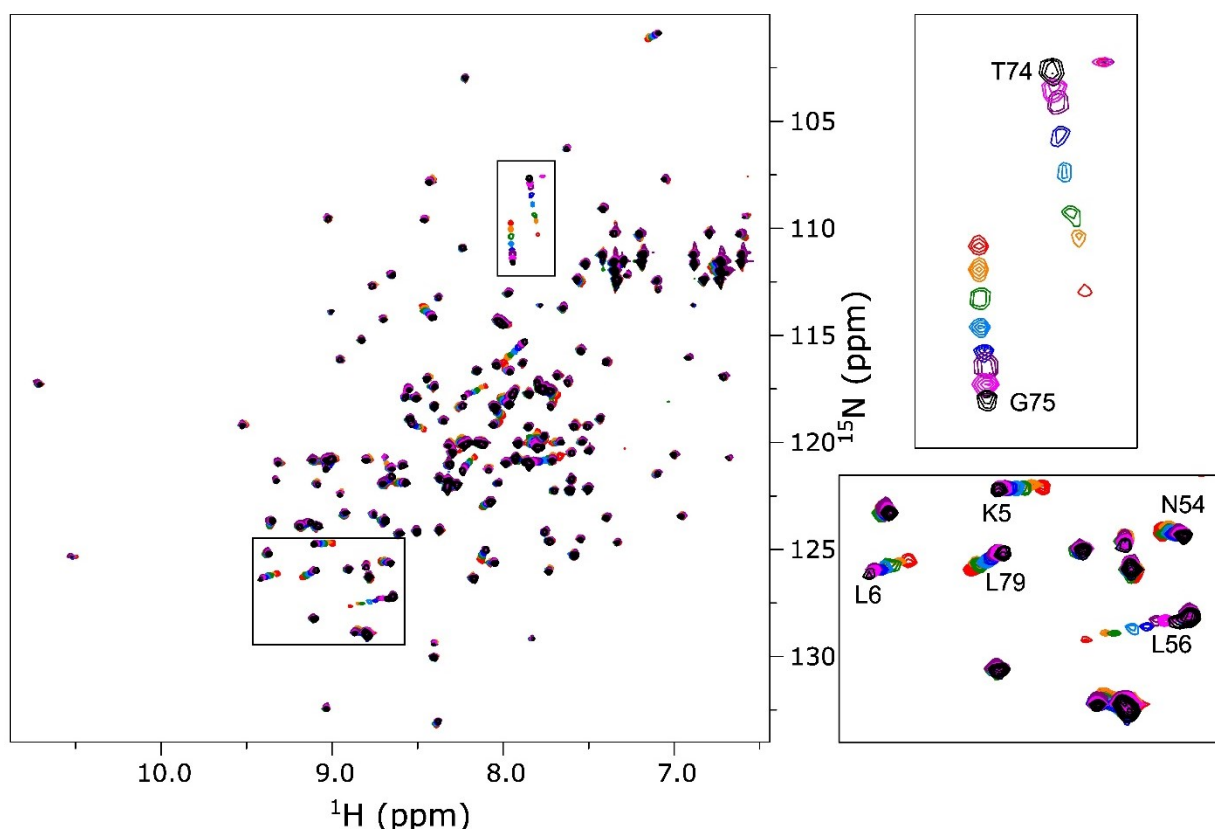


**Figure S3.** (Top three panels) Monitoring aggregation properties of related compounds via a high-throughput 1D NMR assay. Compounds were prepared to 300  $\mu\text{M}$  in 50 mM sodium phosphate, 100 mM NaCl, pH 7.4 with 10%  $\text{D}_2\text{O}$ . Rapid signal decay is indicative of compound aggregation whereas long relaxation times are characteristic of non-aggregating, well-behaved compounds. (Bottom three panels) The initial fragment screening hit and related analogs are shown to be non-aggregating at the concentrations tested (240  $\mu\text{M}$  in 25 mM sodium phosphate, 150 mM NaCl, 5 mM  $\text{MgCl}_2$ , 1 mM TCEP- $\text{d}_{16}$ , pH 7.4 with 10%  $\text{D}_2\text{O}$ ) and were evaluated in binding assays under comparable conditions to mitigate false positives and false negatives.

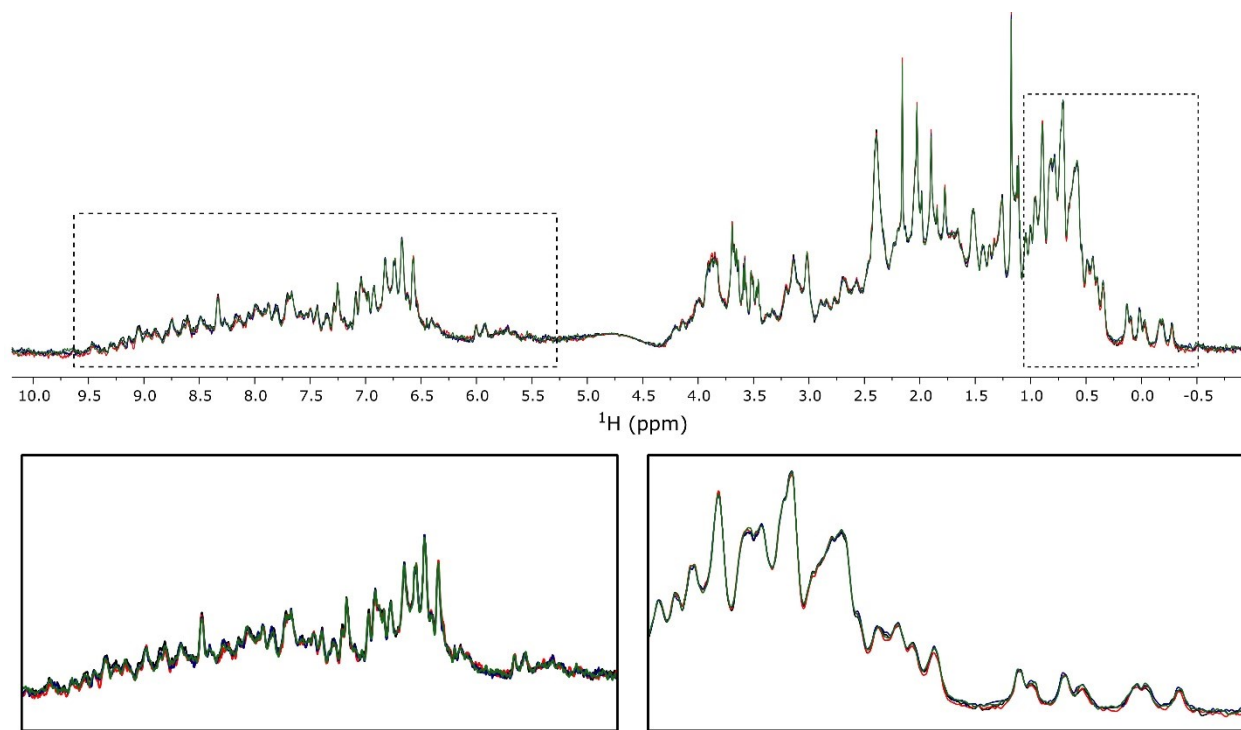


**Figure S4.** Apo HRas protein “fingerprint” spectrum (Top). Monitoring upfield protein methyl shifts with increasing ligand concentration provides direct insight into ligand binding (Bottom). Protein methyl shifts across a range of ligand concentrations provides  $K_D$ . While saturation could not be achieved with either the initial screening hit or NMX-10002, fitted  $K_D$  estimates shown are consistent with orthogonal biophysical data, and most importantly, accurately reflect the relative affinity ranking of the compounds. Note that the ligand concentrations reported are measured rather than assumed nominal concentrations.

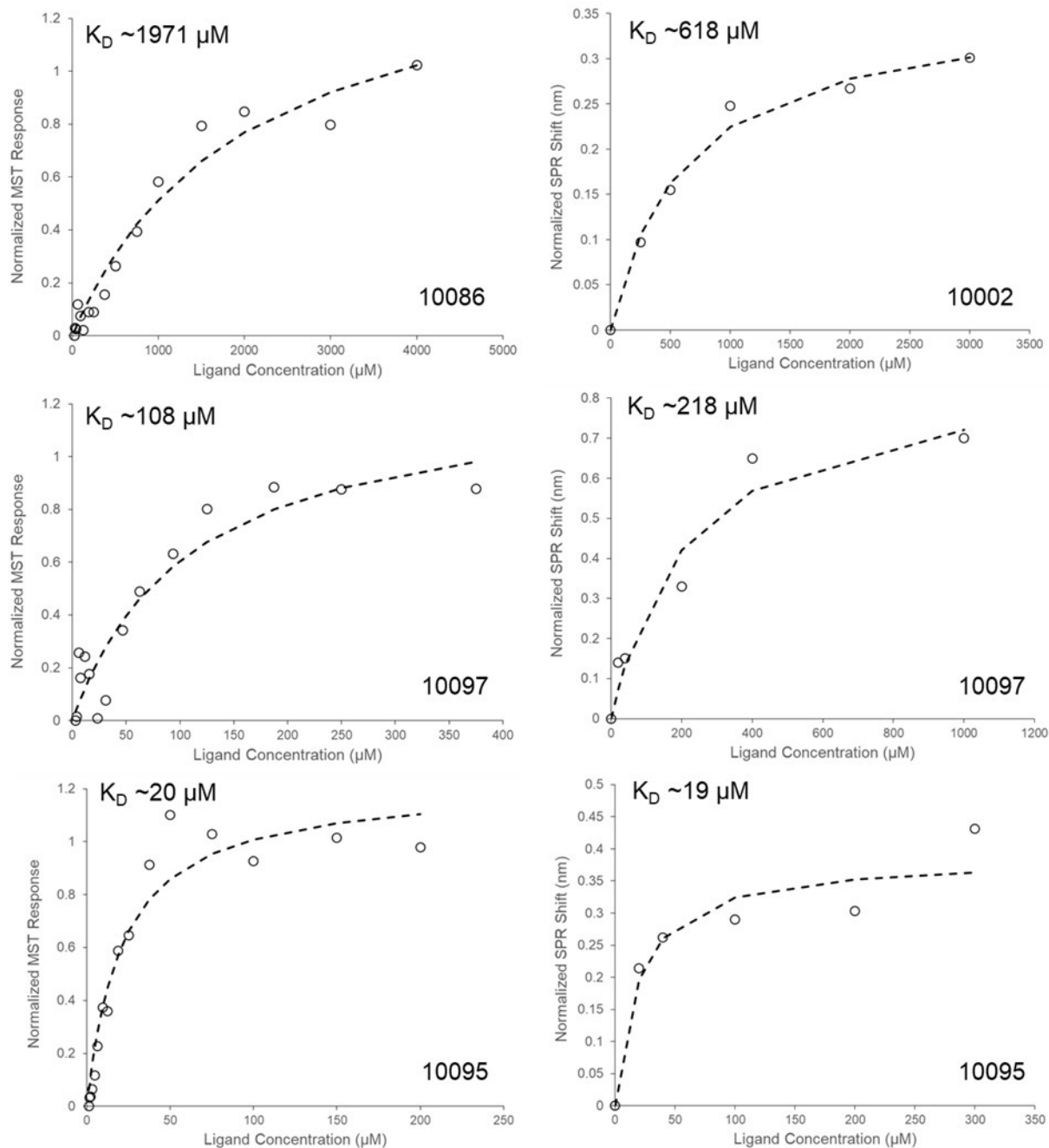




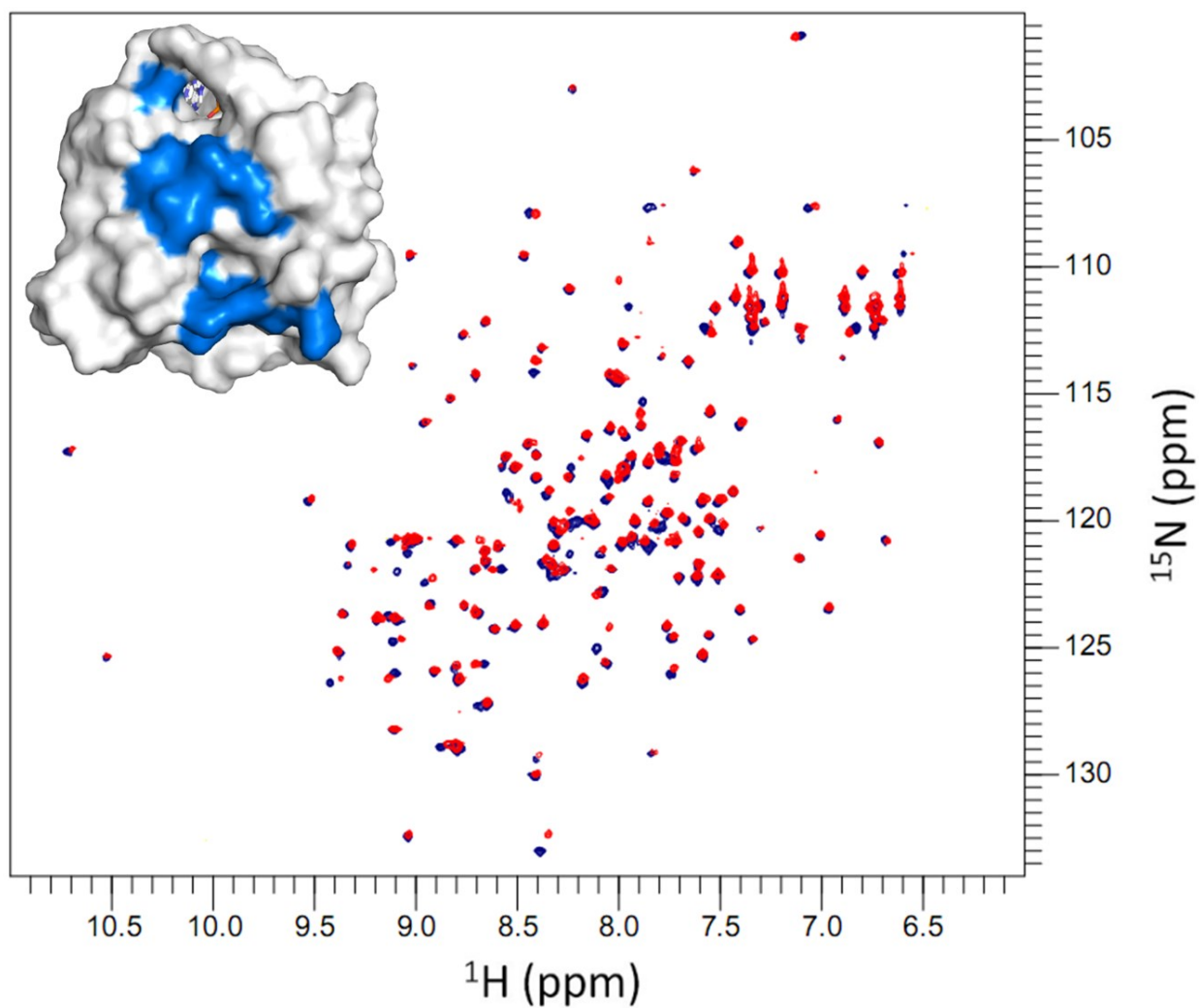
**Figure S5.**  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC titrations were used to estimate binding affinities for well-soluble, weaker affinity compounds. Titration of NMX-10086 is depicted here.



**Figure S6.** HRas stability by 1D protein-observed NMR over 96 hrs. Black, 0h; red, 48h; Blue, 72h; Green, 96h.



**Figure S7.** Orthogonal biophysical data for selected HRAs binders. Left-hand column depicts MST data, whereas the right-hand column shows SPR data. Normalized data were fit in Excel using the solver function (GRG Non-linear method) and simulating the data with a comparable number of points to those measured in the titration.



**Figure S8.** Binding site mapping based on backbone amide chemical shift perturbations between the free protein (blue) and protein in the presence of compound 10097 (red). Only weighted average chemical shift changes  $\geq 0.02$  ppm are mapped onto the structure of HRasG12V (3OIW). Binding of compound 10097 induces changes in the space between switch I and switch II.

## Supplementary methods

### MST

MST experiments were performed with a Monolith NT.115 Pico (NanoTemper Technologies, Munich, Germany). Fluorescence labeling of GDP HRasG12V was achieved according to Nanotemper's protocol for use of the His Tag Labeling Kit RED tris NTA 2nd generation, or Kit RED-NHS 2nd Generation Labeling Kit (NanoTemper Technologies, Munich, Germany). Final protein concentrations were 20 nM fluorescently labeled GDP HRasG12V. Compounds were prepared across a range of concentrations within the measured solubility ranges determined previously by NMR, and loaded into monolith capillaries. Data were acquired and analyzed with the NanoTemper MO.Control and MO. Affinity Analysis software. MST responses that passed the automated QC criteria were exported and normalized before fitting to a 1:1 binding model.

### SPR

SPR was performed using a P4SPR (Affinité Instruments, Montreal, Canada) using His-tagged HRasG12V. Ni-NTA-coated Au SPR chips were first rinsed with DI water. Protein was then immobilized by injecting ~300  $\mu$ L of a 5  $\mu$ M solution of protein in 1X PBS over the chip surface, followed by a 20 min incubation period. Wells were then washed and equilibrated with 1X PBS prior to dose-response experiments with small molecule analyte. Compounds were prepared in DMSO at various concentrations (5 points, dictated by expected affinity ranges given prior biophysical data). Compounds were then dried using a SpeedVac to obtain dry solid powders, which were resolubilized in 1X PBS. Solubilities in 1X PBS were previously known according to NMR data in nearly equivalent buffer conditions. Each ligand solution (~300  $\mu$ L) was injected across the SPR chip surface and left to equilibrate for a minimum of 5 minutes before measuring changes in SPR wavelength, or RU. Subsequent concentrations of the same ligand were likewise tested in series starting from the lowest to highest concentration. Measured shifts in RU were then plotted and fit to a 1:1 binding model.

## 8 ARTICLE 4 - PRACTICAL CONSIDERATIONS AND GUIDELINES FOR SPECTRAL REFERENCING FOR FLUORINE NMR LIGAND SCREENING

---

### Practical Considerations and Guidelines for Spectral Referencing for Fluorine NMR Ligand Screening

Yann Ayotte<sup>a,b</sup>, Simon Woo<sup>a,b</sup>, Steven R. LaPlante<sup>\*a,b</sup>

<sup>a</sup> INRS – Centre Armand-Frappier Santé Biotechnologie, 531 boulevard des Prairies, Laval, Québec, H7V 1B7, Canada.

<sup>b</sup> NMX Research and Solutions Inc., 500 boulevard Cartier Ouest, Suite 6000, Laval, Québec, H7V 5B7, Canada

\* Corresponding author

Published in : ACS Omega 2022, 7, 15, 13155-13163

DOI : 10.1021/acsomega.2c00613

#### Authors contributions :

Yann Ayotte came up with the idea for the project, designed the experiments, performed data analysis, generated the figures, and wrote the manuscript under supervision from Steven Laplante. Experiments were performed by Yann Ayotte with support from an undergraduate intern under supervision from Yann Ayotte. Simon Woo assisted in the writing of the manuscript and provided guidance and support throughout the project. Steven Laplante revised the manuscript.

## Abstract

Fluorine ( $^{19}\text{F}$ ) NMR strategies are increasingly being employed for evaluating ligand binding to macromolecules, among many other uses.  $^{19}\text{F}$  NMR offers many advantages as a result of its sensitive spin 1/2 nucleus, 100% natural abundance, and wide chemical shift range. Moreover, because of its absence from biological samples, one can directly monitor ligand binding without background interference from the macromolecule. Therefore, all these aforementioned features make it an attractive approach for screening compounds. However, the detection of ligand binding, especially those with weak affinities, can require interpretations of minor changes in chemical shifts. Thus, chemical shift referencing is critical for accurate measurements and interpretations. Unfortunately, one cannot rely on spectrometer indirect referencing alone, and internal chemical references have sample-dependent issues. Here, we evaluated 10 potential candidate compounds that could serve as  $^{19}\text{F}$  NMR chemical references. Multiple factors were systematically evaluated for each candidate to monitor the suitability for  $^{19}\text{F}$  NMR screening purposes. These factors include aqueous solubility, buffer compatibility, salt compatibility, aqueous stability, tolerability to pH changes, temperature changes, and compound pooling. It was concluded that there was no ideal candidate, but five compounds had properties that met the screening requirements.

## Introduction

Fragment-based lead discovery (FBLD) involves the screening of low-molecular-weight compounds to identify binders to essential disease target proteins or nucleic acids. Nuclear magnetic resonance (NMR) is a very useful tool for FBLD due to its ability to detect weak binding events in a label- and immobilization-free environment. Traditional ligand-detected experiments have mostly been performed using proton ( $^1\text{H}$ ) NMR, but fluorine ( $^{19}\text{F}$ ) experiments have increasingly gained in popularity in recent years. The large, background-free chemical shift dispersion of the fluorine moiety, combined with its 100% natural abundance and high sensitivity to molecular interactions, has made it an attractive tool in the field of drug discovery.<sup>1-17</sup>

Binding events are usually detected by monitoring changes in chemical shifts, signal width, and/or peak intensities. Variation in the chemical shift is expected to occur if a difference between the bound and free states is experienced by the  $^{19}\text{F}$  nucleus.<sup>18</sup> However, these observations can be skewed, as minor but significant changes in chemical shifts are common due to the wide spectral

dispersion of the  $^{19}\text{F}$  nucleus coupled with spectrometer instabilities and sample-dependent shift changes. Thus, one cannot always rely on spectrometer indirect referencing alone.

In general, the IUPAC recommendations favor internal referencing or substitution methods,<sup>19</sup> with internal referencing being generally more practical for a drug screening context. The IUPAC also recommends the use of  $\text{CCl}_3\text{F}$  as a reference compound for  $^{19}\text{F}$  NMR, but this molecule presents several practical limitations: it has limited aqueous solubility, is highly volatile at ambient temperature, and possesses ozone-depleting properties, which restricts its commercial availability.<sup>20</sup> Analogously, currently recommended  $^{19}\text{F}$  quantitative NMR references are intended to be used in organic solvents and are less optimal for screening, as their aqueous solubility is also limited.<sup>21</sup> Therefore, there are no definitive guidelines for  $^{19}\text{F}$  NMR screening, and the choice of reference depends on lab-specific preferences or arbitrary reasons.<sup>1</sup>

Some essential characteristics should be considered in the choice of a  $^{19}\text{F}$  NMR shift reference, keeping in mind that requirements can be project-dependent. These characteristics include solubility and stability in aqueous media, compatibility with common buffer components, absence of promiscuous binding to protein systems, chemical shift compatibility with standard NMR experimental screening parameters, and minimal chemical shift changes from variations in sample conditions (e.g., pH, temperature, dimethyl sulfoxide (DMSO) content, mixtures of compounds). Other desirable (but not essential) features include the presence of a polyfluoro moiety (e.g.,  $\text{CF}_3$ ), allowing for a sufficient  $^{19}\text{F}$  NMR signal-to-noise ratio even at low concentrations of the reference compound in the sample to limit potential artifacts, as well as commercial availability, lack of safety concerns, and ease of handling. Furthermore, reference compounds that lack or have minimal nonexchangeable aromatic hydrogens would allow concurrent  $^1\text{H}$  NMR experiments to be acquired on the same sample.

Herein, we evaluated some of the most commonly used fluorine shift references for  $^{19}\text{F}$  NMR under a variety of conditions and environments to assess their suitability for drug discovery studies. We also provide some guidelines that may help users choose the most appropriate reference for their project.

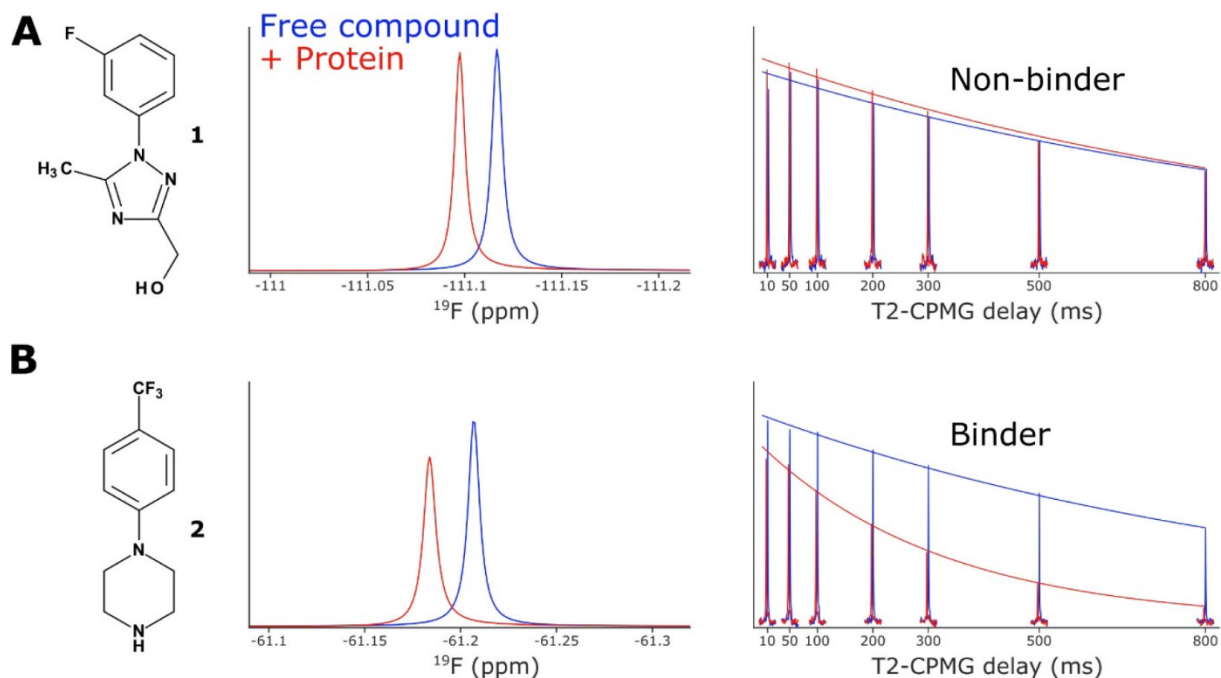
## **Results and Discussion**

### **Example: Inconclusive Results in the Absence of an Internal Shift Reference**



As with routine  $^1\text{H}$  NMR applications, spectrometer indirect referencing is often preferred for  $^{19}\text{F}$  NMR shift calibration. However, inconsistencies can arise from one sample to the next due to the reasons already described above.

One ramification is that interpretations can become ambiguous. An example is illustrated in Figure 1, which shows a  $^{19}\text{F}$  NMR screen aimed at determining whether two compounds bind to a target protein. In panel A, the differential line broadening/shifting (DLBS) method would suggest that compound **1** could bind to the target protein given that the  $^{19}\text{F}$  spectrum of free compound **1** (blue spectrum) experiences a distinct change in chemical shift upon the addition of target protein (red spectrum). However, a confirmational T2-CPMG (Carr–Purcell–Meiboom–Gill) experiment shows that compound **1** does not appear to bind the target protein.<sup>22</sup> On the one hand, no significant changes in relaxation rates (i.e., slopes as a function of delay periods) are observed between the samples containing free **1** (blue) and **1** with added protein (red). On the other hand, in panel B the changes in DLBS and T2-CPMG data (red vs blue spectra) support that compound **2** indeed binds to the target protein. Thus, perhaps the chemical shift changes in panel A were due to either spectrometer drift or local chemical environment changes rather than a binding event. The lack of a significant difference in the T2-CPMG spectra in the presence of protein supports this hypothesis; therefore, an analysis based on shift alone could result in the misinterpretation that compound **1** bound to the target protein. Hence, it is crucial that a  $^{19}\text{F}$  chemical shift reference be employed in such a context.



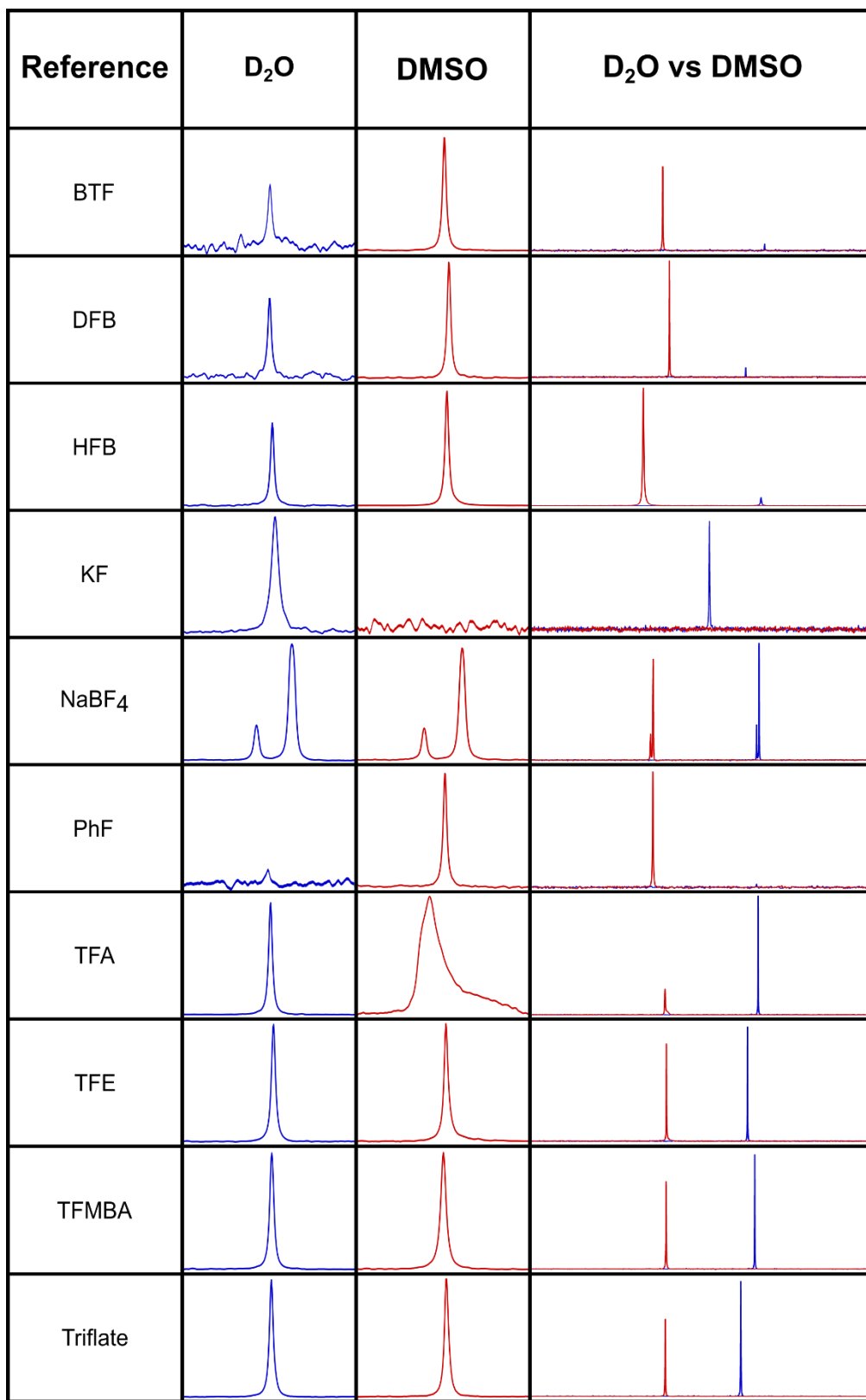
**Figure 1.**  $^{19}\text{F}$  NMR screen using DLBS and T2-CPMG methods. (A) Compound **1** exhibits changes in chemical shift in the presence of protein (red spectrum), as compared to the free compound (blue spectrum). Further evaluation of the binding using T2-CPMG suggests that this compound is not a binder and that the chemical shift difference is likely explained by the sensitivity to small changes in conditions. In comparison, (B) shows compound **2**, which appears to be a real binder to the protein based on the significant differences in line broadening and T2-CPMG decay rate observed

### Selection of $^{19}\text{F}$ Reference Compounds

A set of commonly used  $^{19}\text{F}$  references was chosen for this study. Their names and abbreviations are as follows: **BTF**: benzotrifluoride; **DFB**: 1,2-difluorobenzene; **HFB**: hexafluorobenzene; **KF**: potassium fluoride; **NaBF<sub>4</sub>**: sodium tetrafluoroborate; **PhF**: fluorobenzene; **TFA**: trifluoroacetic acid; **TFE**: trifluoroethanol; **TFMBA**: 2-(trifluoromethyl)benzoic acid; **Triflate**: sodium trifluoromethanesulfonate.

### Solubility Tests

A chemical shift reference should exhibit sufficient solubility under aqueous conditions to be practical for NMR screening. The solubility of the 10 reference candidates was therefore assessed at 200  $\mu\text{M}$  nominal concentration in 100%  $\text{D}_2\text{O}$ . Data were also acquired on samples at 200  $\mu\text{M}$  nominal concentration in 100%  $\text{DMSO-}d_6$  for comparisons of signal intensities. Figure 2 illustrates the respective signal intensities, as well as the overlaid spectra (far-right column), in  $\text{D}_2\text{O}$  and  $\text{DMSO-}d_6$  for each molecule. PhF was completely insoluble in  $\text{D}_2\text{O}$ , while BTF, DFB, and HFB exhibited poor signal-to-noise in aqueous solution. These four candidates were therefore eliminated from further testing due to their insufficient solubility and, thus, incompatibility as a  $^{19}\text{F}$  NMR screening reference. Even though it appeared to be insoluble in DMSO, KF showed sufficient signal-to-noise under aqueous conditions and was therefore advanced to further evaluations. Similarly, TFA showed a lower signal intensity in DMSO as well as a broader, nonsymmetrical peak line shape, which could perhaps be due to an exchange phenomenon. Given an acceptable intensity profile in  $\text{D}_2\text{O}$ , TFA was also considered for additional studies. Finally,  $\text{NaBF}_4$ , TFE, TFMBA, and Triflate all showed adequate solubility profiles and were also selected for the next rounds of testing.

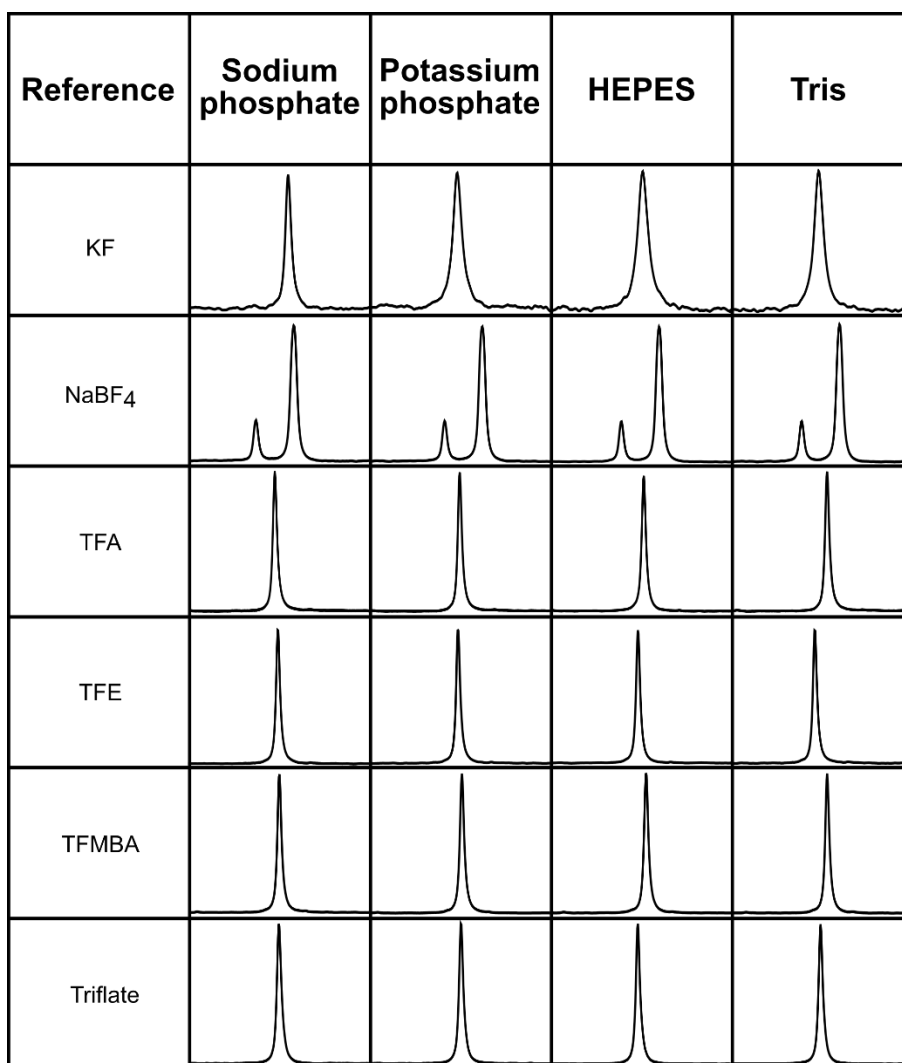


**Figure 2.** <sup>19</sup>F spectra of the candidate molecules were acquired in both D<sub>2</sub>O (blue spectra) and DMSO-*d*<sub>6</sub> (red spectra) to evaluate their aqueous solubility. Overlaid spectra of both conditions (last column) allow

for a better evaluation of the relative signal intensities. Compounds were tested at 200  $\mu\text{M}$  nominal concentration.

### Buffer Compatibility Tests

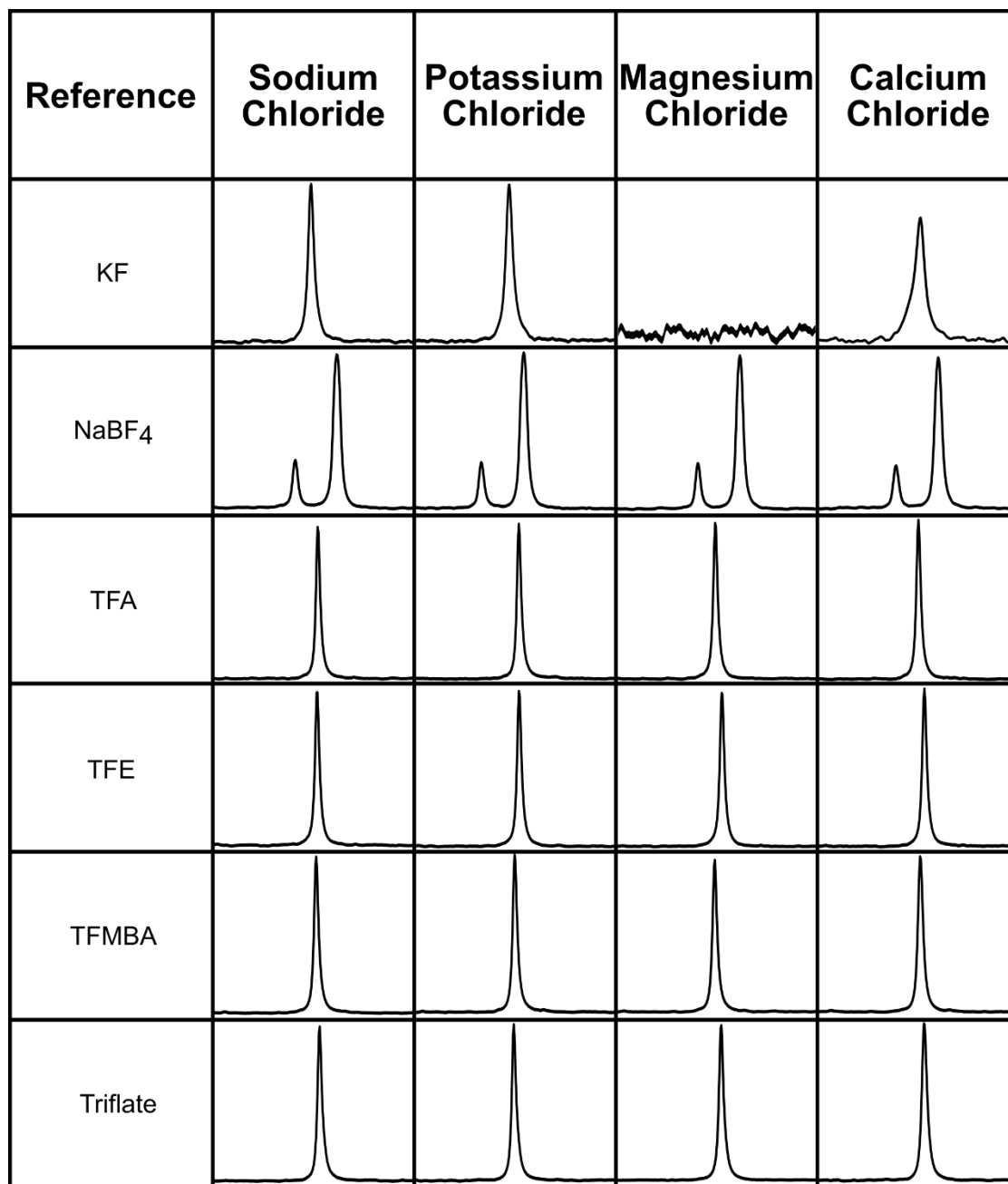
Another important attribute for a fluorine reference compound should be compatibility with various buffer conditions. The remaining six reference candidates were then evaluated for their compatibility with four common buffer components: sodium phosphate, potassium phosphate, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), and tris(hydroxymethyl)aminomethane (tris). An analysis of the  $^{19}\text{F}$  NMR spectra of the six candidates in these buffers is shown in Figure 3. Given the reasonable peak intensities in all conditions, it is apparent that all six are compatible with these buffers.



**Figure 3.** Compatibility of six candidates with common buffer components. 1D  $^{19}\text{F}$  spectra are shown for compounds in each condition. Buffers were prepared at 50 mM, pH 7.0 with 10%  $\text{D}_2\text{O}$ , and compounds were tested at 200  $\mu\text{M}$  nominal concentration.

### Salt Compatibility Tests

For screening purposes, a fluorine reference should also be compatible with various salts. The compatibility with four commonly used salts (sodium chloride, potassium chloride, magnesium chloride, and calcium chloride) was then assessed as shown in Figure 4. With the exception of KF, all compounds were compatible with the four salts investigated. However, KF exhibited insolubility in the presence of magnesium chloride. This effect was further confirmed by titrating KF against a lower concentration (1 mM) of  $\text{MgCl}_2$  (Figure S1), and significant loss of signal intensity and broad line widths could be observed, suggesting the formation of larger unknown entities in the sample. In a similar fashion, reduced signal intensity with a wider line shape was observed for KF in the presence of calcium chloride (Figure 4), which suggested limited compatibility with this salt as well. This is in line with the relatively limited solubility expected for  $\text{MgF}_2$  and  $\text{CaF}_2$ .<sup>23a,b</sup> KF was therefore eliminated as a potential reference candidate. The remaining five molecules were then tested against a wide range of buffer components and additives (Figures S2–S5), and no significant incompatibility could be observed in any of the tested conditions.

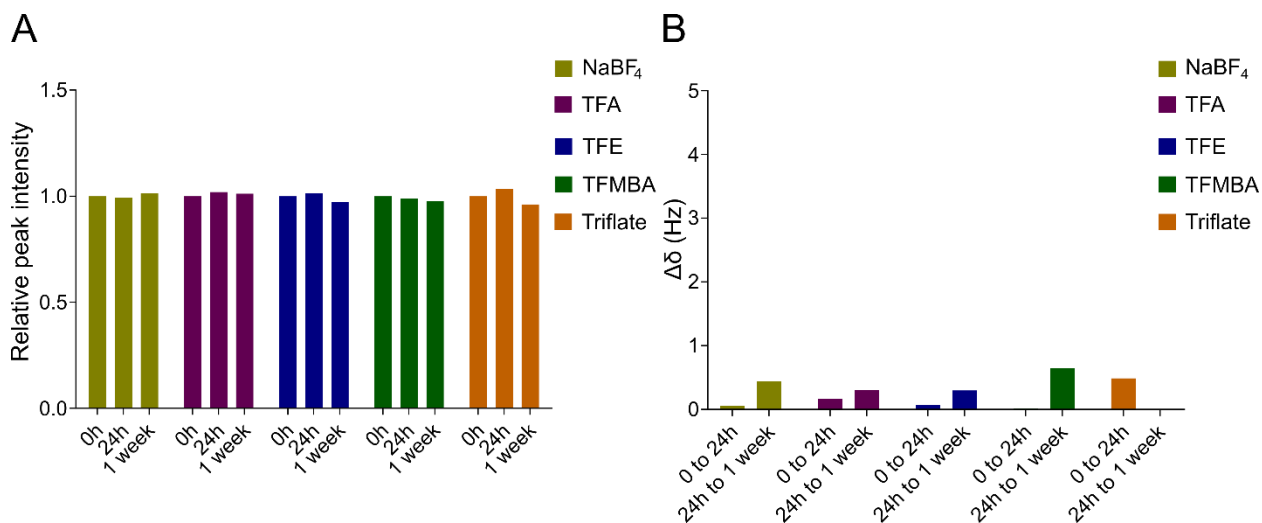


**Figure 4.** Compatibility of the six candidates with four common salts. 1D  $^{19}\text{F}$  spectra are shown for compounds in each condition. Sodium and potassium chloride were tested at 200 mM, while magnesium and calcium chloride were tested at 50 mM. Compounds were tested at a nominal concentration of 200  $\mu\text{M}$ , and solutions were prepared with 10%  $\text{D}_2\text{O}$ .

### Aqueous Stability Tests

Given that some NMR screening studies can last for several days, the aqueous stability of each remaining candidate was then assessed at three different time points: 0 h, 24 h, and one

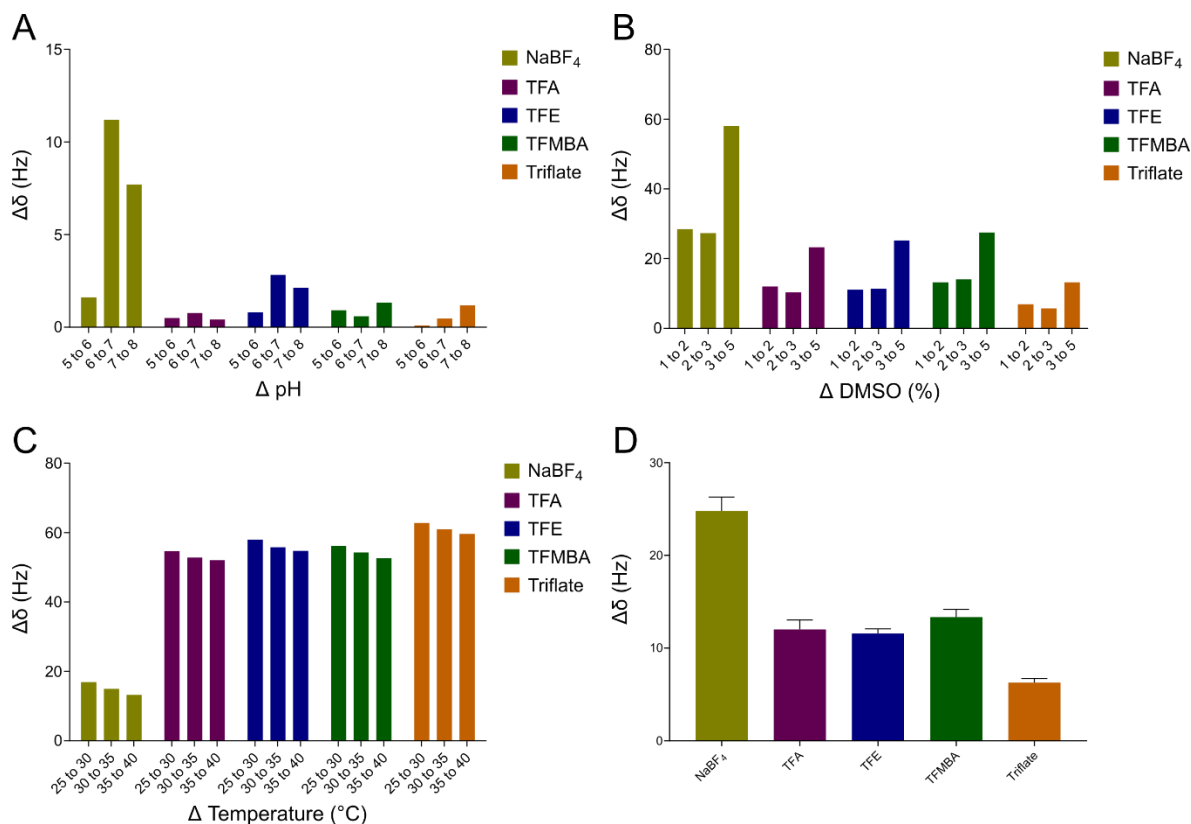
week. Figure 5A shows that signal intensities remain relatively stable across all time points for the five compounds. Similarly, minor variations in chemical shift of less than 1 Hz were noted for periods up to a week (Figure 5B).



**Figure 5.** Evaluation of compound stability under aqueous conditions. 1D <sup>19</sup>F spectra were acquired at three time points: 0 h, 24 h, and one week. Signal intensities were measured at each time point (A) as well as variation in the <sup>19</sup>F chemical shift between each time point (B). Signal intensities in (A) were normalized to *t* = 0 h for each molecule. Samples were measured at 200 μM in 50 mM sodium phosphate pH 7.4, 100 mM NaCl, 10% D<sub>2</sub>O.

### pH, %DMSO, Temperature, and Compound Pooling Tests

We then evaluated the “sensitivity” of the compounds’ chemical shifts given variations in pH, amount of DMSO, temperature, and in the presence of pools of other compounds. Figure 6A shows that the chemical shift of NaBF<sub>4</sub> seems to be relatively sensitive to pH variations above pH values of 6 but that all the other molecules were rather stable across the pH range tested. Therefore, small variations in pH upon additions of various components, such as a protein or other compound, would only be expected to potentially cause more significant variations in the chemical shift of NaBF<sub>4</sub>.



**Figure 6.** Fluorine chemical shift variations of five compounds upon variation of pH (A), DMSO-*d*<sub>6</sub> content (B), and temperature (C) and in the presence of pools of compounds (D). Samples for B and C were run in 50 mM sodium phosphate pH 7, 10% D<sub>2</sub>O. For A, the same buffer was used as in B & C, at various pH values (5, 6, 7, and 8). In D, each molecule was tested against 20 pools of fragments for an average of ~13 fragments/pool. Average change in chemical shift is plotted with the standard deviation. Samples in D were run in 20 mM Tris-*d*<sub>11</sub> pH 7.5, 150 mM NaCl, 10% D<sub>2</sub>O.

Similarly, because compounds are usually added into a buffer from DMSO solvent stocks, there was concern that chemical shifts of the potential reference compounds could be sensitive to different concentrations of DMSO (e.g., for drug titration purposes, from different stock concentrations, or even variations in pipetting). Thus, we were interested in evaluating if the reference candidates experienced changes in chemical shifts as the percentage of DMSO was altered. Figure 6B shows the changes in chemical shifts upon addition of 1, 2, 3, or 5% (v/v) DMSO-*d*<sub>6</sub>. NaBF<sub>4</sub> was once again more influenced by variations in the amounts of DMSO.

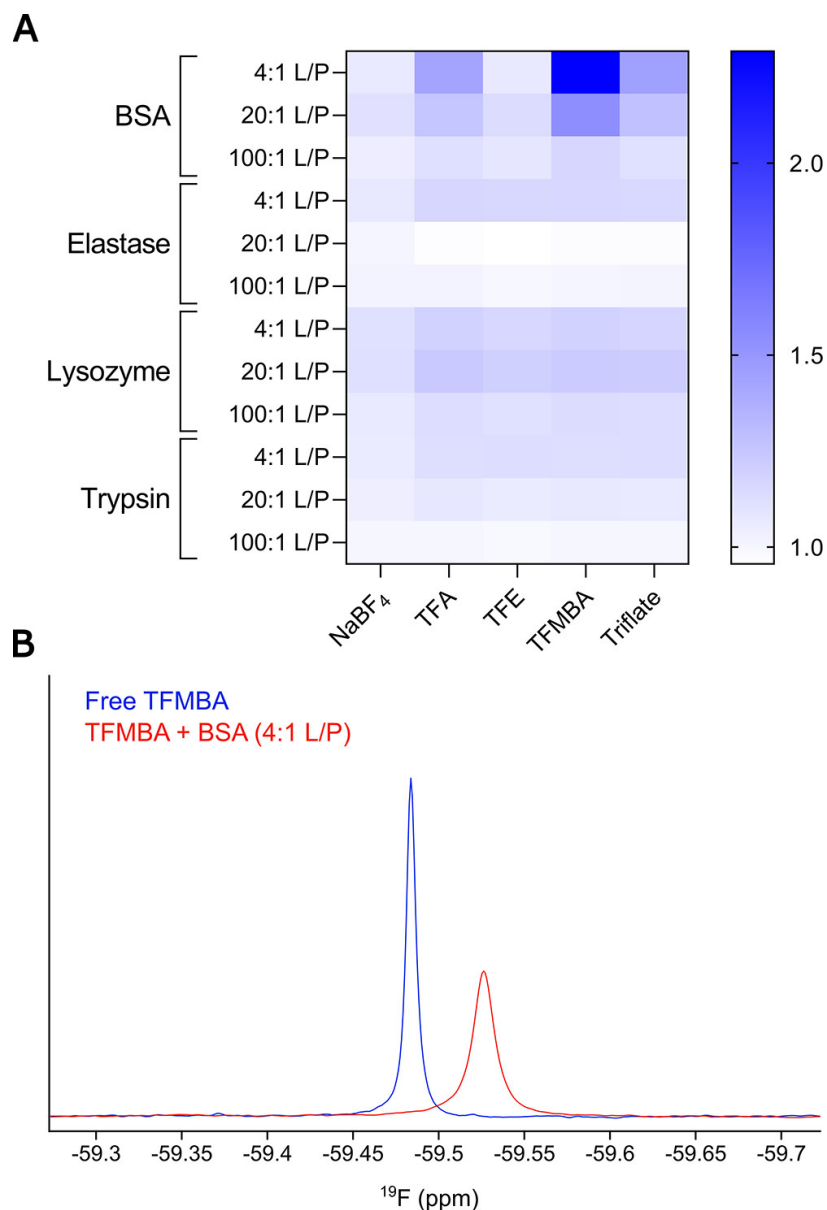
Although temperature is usually well-controlled during NMR experiments, some NMR pulse sequences can result in some sample heating. Moreover, insufficient equilibration of the sample temperature before acquisition can also result in variation during the experiment. The influence of temperature fluctuation on the reference's chemical shifts was therefore probed by varying the temperature from 25 to 40 °C by 5 °C increments (Figure 6C). Most of the compounds show



comparable changes in chemical shifts with varying temperature, but, interestingly, the chemical shift of NaBF<sub>4</sub> was the least affected by changes in temperature.

Fragment-based NMR screens are often performed using mixtures of compounds (pools) in order to increase throughput,<sup>4,8,18,24-27</sup> so an internal reference used in these screens should experience minimal effects on its chemical shift in the presence of other compounds. Each reference candidate was therefore tested in 20 different pools of small-molecule fragments containing 11–15 fragments per pool. Figure 6D shows the average change in chemical shift for each reference candidate when placed in the pools compared to the free compounds. It is noteworthy that NaBF<sub>4</sub>'s chemical shift is, on average, more affected, while triflate's is the least affected. TFE, TFMA, and TFA exhibit comparable intermediate profiles. In light of the results observed in Figure 6A, the presence of multiple compounds in solution may induce slight changes in pH within the samples, and these changes could explain some of the observed effects. These results suggest that NaBF<sub>4</sub> would be a less-favorable option in the context of screening mixtures.

Ideally, an internal reference should not bind to the target macromolecule (usually a protein) that is being screened. Therefore, we screened the five remaining candidates against four different commercially available protein systems: bovine serum albumin (BSA), elastase, lysozyme, and trypsin. Binding was assessed at three different ligand-to-protein ratios (L/P) by monitoring changes in the <sup>19</sup>F differential line broadening (DLB) of each reference's signal upon addition of protein. Figure 7A shows that the addition of elastase, lysozyme, or trypsin results in little to no DLB effects under the various L/P tested. However, the addition of BSA results in relatively large DLB effects for TFA, triflate, and especially TFMA (illustrated in Figure 7B) as the ligand-to-protein ratio approaches equimolar. Note that this is not unexpected considering that albumins are known to bind a wide variety of molecules.<sup>28</sup> However, this observation suggests that these reference candidates may be more prone to binding target macromolecules. Interestingly, NaBF<sub>4</sub> exhibited minimal DLB across all the conditions tested, suggesting that it might be less susceptible to protein binding.

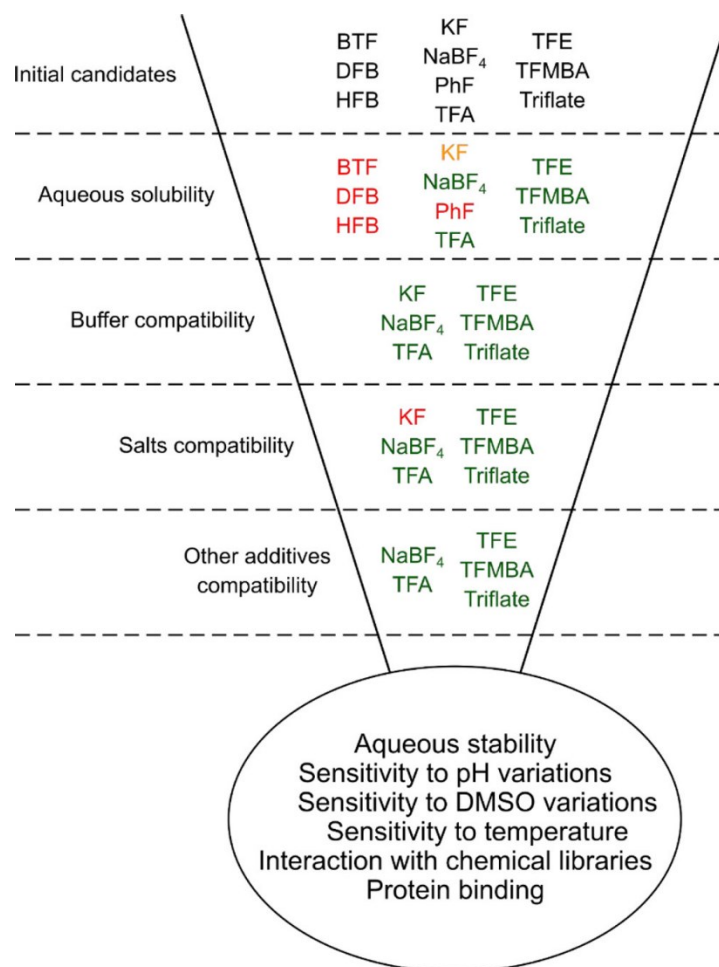


**Figure 7.** (A) Evaluation of binding to various proteins by fluorine DLB. Proteins were added at three different concentrations to 200  $\mu$ M compound. Samples were tested in 50 mM sodium phosphate pH 7.0, 100 mM NaCl, 10% D<sub>2</sub>O. Broadening was measured as a ratio of fluorine peak line width in the presence/absence of protein. A line broadening ratio of 1 represents the absence of any evidence of binding. (B) Example for TFMBa binding in the presence of BSA at 4:1 L/P.

### Summary of the Main Pros and Cons for the Five <sup>19</sup>F NMR References

This study has shown that the evaluation of fluorine reference compounds for NMR screening must consider multiple parameters (Figure 8) and as a result is challenging. Overall, five reasonable candidate references emanated from this study. Table 1 summarizes the main pros and cons for each of these five references. Interestingly, all five contained a polyfluoro moiety

(either a  $\text{CF}_3$  group or  $\text{BF}_4^-$ ) giving them sufficient signal-to-noise even if used at lower concentrations for referencing in NMR screens. Moreover, all five are commercially available and affordable.



**Figure 8.** Funnel-like view of the reference evaluation steps. Compounds eliminated at each step are colored in red. KF was colored orange in the solubility assessment since it appeared soluble in aqueous conditions but insoluble in DMSO.

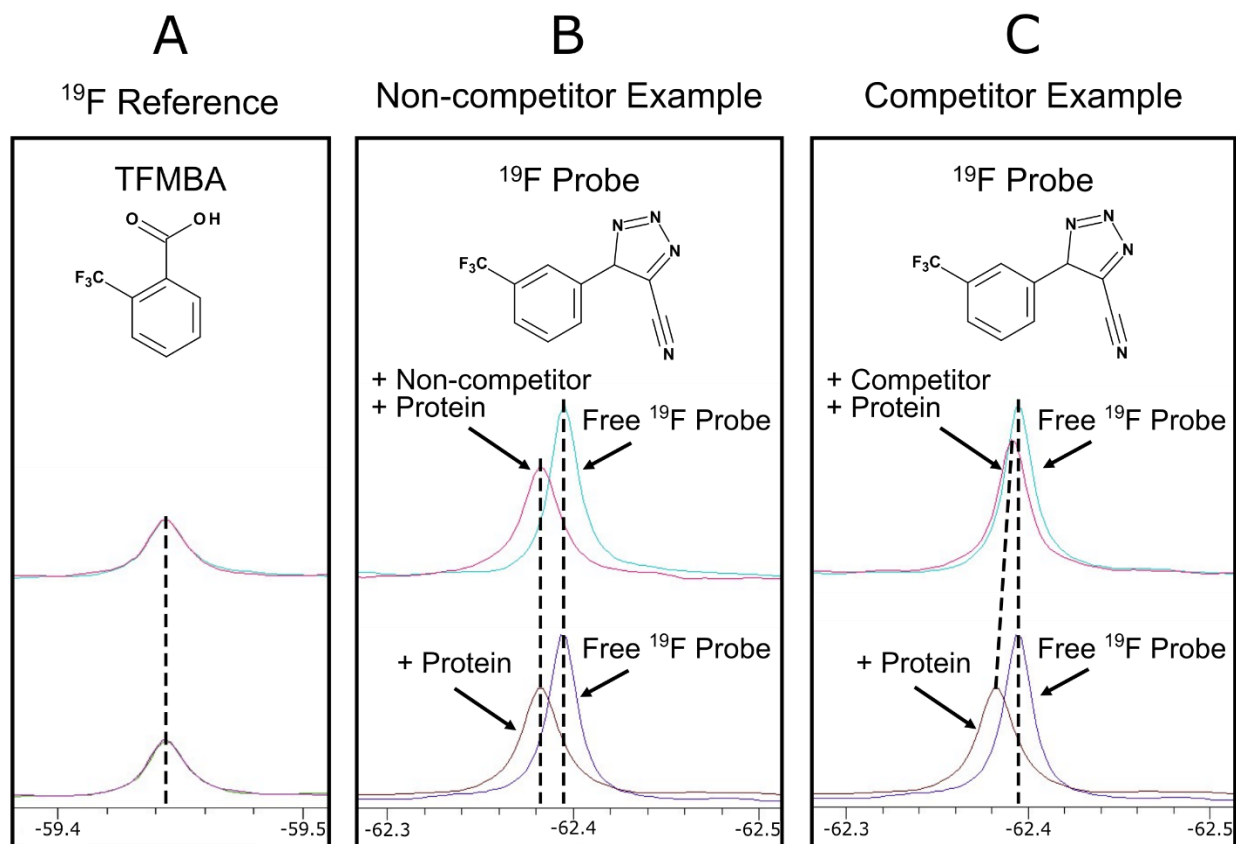
**Table 1**

Reference Compound	Advantages	Disadvantages
NaBF <sub>4</sub>	<ul style="list-style-type: none"><li>• Minimal chance of resonance overlapping with common fragments</li><li>• Absence of <sup>1</sup>H</li></ul>	<ul style="list-style-type: none"><li>• Chemical shift might fall outside the spectral width of some NMR sequences</li><li>• 2 species observed due to natural abundance of boron isotopes</li><li>• Appears to be more sensitive to variations in DMSO or pH</li></ul>
TFA	<ul style="list-style-type: none"><li>• Absence of <sup>1</sup>H</li></ul>	<ul style="list-style-type: none"><li>• Can already be a residual from chemical synthesis</li></ul>
TFE	<ul style="list-style-type: none"><li>• Only aliphatic <sup>1</sup>H</li></ul>	<ul style="list-style-type: none"><li>• Volatile and flammable</li><li>• Aliphatic <sup>1</sup>H</li><li>• Known to promote protein changes at higher concentrations<sup>29</sup></li></ul>
TFMBA	<ul style="list-style-type: none"><li>• Decent all-around performance</li></ul>	<ul style="list-style-type: none"><li>• Aromatic <sup>1</sup>H</li><li>• More fragment-like molecule, could be more prone to binding proteins</li></ul>
Triflate	<ul style="list-style-type: none"><li>• Absence of <sup>1</sup>H</li><li>• Appears to be slightly less sensitive to variations in DMSO or pH than other candidates</li></ul>	<ul style="list-style-type: none"><li>• Appears to be slightly more sensitive to variations in temperature than other candidates</li></ul>

### Example Demonstration of the Utility of TFMBA for <sup>19</sup>F Chemical Shift Referencing in a Competition Study

Given the ensemble of data described herein, our laboratory frequently uses TFMBA as an internal reference for <sup>19</sup>F NMR screening studies. Figure 9 demonstrates an example of such a study for which the aim was to evaluate two compounds to determine whether they bound to a specific pocket (P-1) of the multipocketed target human rhinovirus polymerase (HRV). To do so, we had access to a known P-1 binder (<sup>19</sup>F probe—see Figure 9B,C) that could serve as a displacement probe for other binders to pocket P-1. This probe also contained a CF<sub>3</sub> group, which increased the <sup>19</sup>F NMR signal-to-noise while also exhibiting relatively sharp linewidths due to a fast rotation along the CF<sub>3</sub>–phenyl bond. Given this, samples were prepared with the internal

TFMBA  $^{19}\text{F}$  reference at a lower concentration of  $5\ \mu\text{M}$ . The low concentration of both the reference and the probe ( $^{19}\text{F}$  probe at  $22\ \mu\text{M}$ ) would minimize any potential intermolecular association and interference. As a precaution, separate control experiments were run and showed that the TFMBA reference did not bind to the probe, HRV protein, or to the two test compounds (data not shown). Samples were then prepared with the three compounds present simultaneously (TFMBA at  $5\ \mu\text{M}$ ,  $^{19}\text{F}$  probe at  $22\ \mu\text{M}$ , test compound at  $100\ \mu\text{M}$ ), both with and without HRV protein.



**Figure 9.** Demonstration of the utility of  $^{19}\text{F}$  chemical shift referencing in a competition study. (A)  $^{19}\text{F}$  NMR spectra of TFMBA ( $5\ \mu\text{M}$ ). (B)  $^{19}\text{F}$  NMR spectra of  $^{19}\text{F}$  probe at  $22\ \mu\text{M}$ , HRV at  $22\ \mu\text{M}$ , noncompetitor test compound at  $100\ \mu\text{M}$ . (C)  $^{19}\text{F}$  NMR spectra of  $^{19}\text{F}$  probe at  $22\ \mu\text{M}$ , HRV at  $22\ \mu\text{M}$ , competitor test compound at  $100\ \mu\text{M}$ .

Key to the proper analyses of the data shown in Figure 9 was that all  $^{19}\text{F}$  NMR spectra were chemical shift referenced to the TFMBA peak as shown in Figure 9A, which then facilitated accurate interpretations. Given that the  $^{19}\text{F}$  probe and the test compounds were relatively weak

binders, resonance averaging was expected due to fast exchange between the free and bound states (on the NMR time scale). Therefore, minor chemical shift changes would be expected upon binding. Upon addition of the HRV protein, the  $^{19}\text{F}$  NMR resonance of the  $^{19}\text{F}$  probe would shift downfield to that of the free  $^{19}\text{F}$  probe, indicating a binding event. In the presence of a competitor test compound, however, as shown in the upper spectrum of Figure 9C, the resonance of the  $^{19}\text{F}$  probe returned to its free-state chemical shift, indicating that it was competed out of pocket P-1 by the competitor test compound. In contrast, in the presence of a noncompetitor test compound, as shown in the upper spectrum of Figure 9B, the chemical shift of the  $^{19}\text{F}$  probe remained the same as that of the sample without the noncompetitor compound. This result indicated that the  $^{19}\text{F}$  probe remained bound, and the noncompetitor compound did not compete for pocket P-1. Without the chemical shift referencing to TFMBA, these results would have been much less convincing.

### **Additional Considerations**

Several factors make identification of a universal  $^{19}\text{F}$  internal reference very challenging. The main one being the large chemical shift distribution of  $^{19}\text{F}$  compounds, which may require the use of a reference with a chemical shift close to the compound(s) of interest. For example, different references might be desired if mixtures are designed based on the presence of CF or  $\text{CF}_3$ , since they may require different carrier frequencies and spectral width depending on the experimental setup used.<sup>4</sup> Alternatively, the use of broadband  $^{19}\text{F}$  pulse sequences can help circumvent this difficulty.

If users also plan to include  $^1\text{H}$  NMR experiments for follow-up steps, then some  $^1\text{H}$ -containing references will be less appropriate due to the potential for signal overlap with the compounds of interest. For example, Figure S6 shows the one-dimensional (1D)  $^1\text{H}$  spectra of TFMBA and TFE. On the one hand, because of the absence of any aromatic protons on TFE, this mitigates potential problems of resonance overlap considering that aromatic protons are often favored during an NMR binding analysis due to the simplicity and the usual lack of any signal overlap with common buffer components. On the other hand, TFMBA possesses four aromatic protons, which is likely to overlap various compounds of interest, as depicted in Figure S6B.

### **Conclusions**

In summary, we have evaluated various fluorine reference candidates under various conditions to assess their suitability for NMR drug discovery experiments. We also highlighted the complexity of choosing an appropriate reference molecule and provided some recommendations to guide

these choices. Evidently, there are likely other interesting reference candidates out there, and therefore, a similar testing sequence could be extended to other compounds to identify additional candidates. After a screen, a molecule defined as nonbinder can also be used as project-specific internal reference for follow-up steps.<sup>1</sup> However, care must be taken when changing experimental conditions to ensure compatibility.

Regardless of the references to be used, they should always be assessed against each screening target to rule out binding of the reference to the target or even destabilization of the latter. To streamline the process, a selected subset of references can be pooled together and tested against the protein/nucleic acid of interest before launching screening efforts.

## **Materials and Methods**

### **Compounds and Libraries**

All compounds investigated in this work were ordered from external vendors. The suppliers and catalog numbers are provided in the Supporting Information. The fragment library used to assess pool effects was provided by NMX Research and Solutions Inc. (Fast-Screen <sup>19</sup>F Fragment Library).

### **NMR Sample Preparation**

Each compound was prepared as a 30 mM stock solution from the purchased powder or liquid in either dimethyl sulfoxide-*d*<sub>6</sub> or deuterium oxide (D<sub>2</sub>O). This solution was then diluted to give the desired final compound concentrations. NMR samples were stored at 4 °C in a SampleJet sample handler, and data were acquired at 25 °C unless otherwise specified.

### **NMR Experiments**

All experiments were run on a 600 MHz Bruker Avance III spectrometer equipped with a helium HFCN cryoprobe. 1D <sup>1</sup>H-decoupled <sup>19</sup>F experiments were acquired using the standard Bruker sequence zgfhgqn. Spectra were acquired with 32 scans and a relaxation delay of 10 s. 1D <sup>1</sup>H spectra were acquired using standard Bruker 1D <sup>1</sup>H sequence with excitation sculpting (zgesgp) and a relaxation delay of 10 s. Spectra were acquired with 16 scans.

To avoid potential interference with another internal reference, a deuterium lock on the magnet was used for <sup>19</sup>F referencing. To ensure sufficient robustness of this method, repeated measurements were performed on three samples containing 200 μM trifluoroethanol over three

time points (0, 12, and 24 h). Very good consistency was observed across time points, with average variations in chemical shifts below 0.1 Hz (Table S1).

### **Data Interpretations**

Data visualization and analysis were done in Bruker's TopSpin software (<https://www.bruker.com/en/products-and-solutions/mr/nmr-software/topspin.html>).

### **Proteins**

All proteins used in this report were purchased from external vendors. The suppliers and catalog numbers are provided in the Supporting Information.

### **Supporting Information**

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.2c00613>.

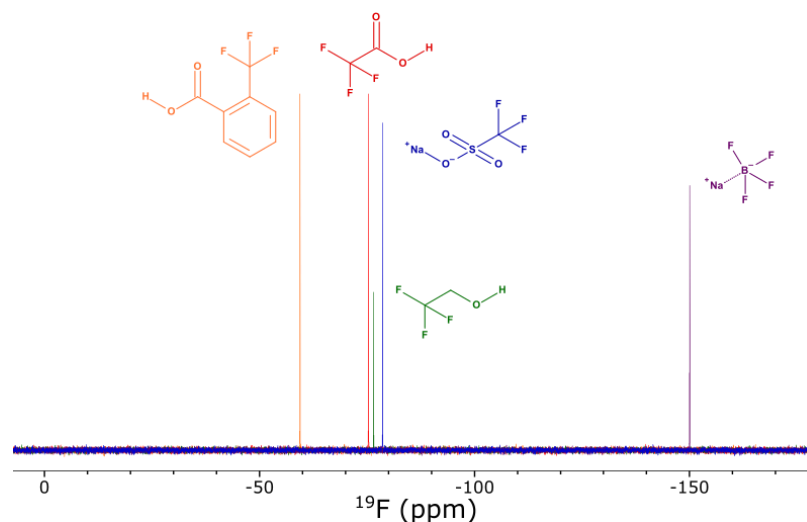
- Additional NMR figures; spectrometer stability of indirect referencing; information on compounds and proteins used in this report.

### **Acknowledgments**

We acknowledge the following for their financial support: CQDM SynergiQc grant, Mitacs, NSERC, CFI, PROTEO, and NMX Research and Solutions Inc. We also thank colleagues at NMX for sharing their expertise and use of their software, as well as E. K. Nounke for her help in initiating the project.



## Table of content figure



## References

- (1) Dalvit, C.; Vulpetti, A., Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **2019**, *62* (5), 2218-2244.
- (2) Cobb, S. L.; Murphy, C. D.,  $^{19}\text{F}$  NMR applications in chemical biology. *J. Fluorine Chem.* **2009**, *130* (2), 132-143.
- (3) Dalvit, C.; Ardini, E.; Flocco, M.; Fogliatto, G. P.; Mongelli, N.; Veronesi, M., A General NMR Method for Rapid, Efficient, and Reliable Biochemical Screening. *J. Am. Chem. Soc.* **2003**, *125* (47), 14620-14625.
- (4) Vulpetti, A.; Hommel, U.; Landrum, G.; Lewis, R.; Dalvit, C., Design and NMR-Based Screening of LEF, a Library of Chemical Fragments with Different Local Environment of Fluorine. *J. Am. Chem. Soc.* **2009**, *131* (36), 12949-12959.
- (5) Rüdisser, S. H.; Goldberg, N.; Ebert, M.-O.; Kovacs, H.; Gossert, A. D., Efficient affinity ranking of fluorinated ligands by  $^{19}\text{F}$  NMR: CSAR and FastCSAR. *J. Biomol. NMR* **2020**, *74* (10), 579-594.
- (6) Norton, R. S.; Leung, E. W. W.; Chandrashekar, I. R.; MacRaid, C. A., Applications of  $^{19}\text{F}$ -NMR in Fragment-Based Drug Discovery. *Molecules* **2016**, *21* (7), 860.
- (7) Lingel, A.; Vulpetti, A.; Reinsperger, T.; Proudfoot, A.; Denay, R.; Frommlet, A.; Henry, C.; Hommel, U.; Gossert, A. D.; Luy, B.; Frank, A. O., Comprehensive and High-

Throughput Exploration of Chemical Space Using Broadband <sup>19</sup>F NMR-Based Screening. *Angew. Chem. Int. Ed.* **2020**, *59* (35), 14809-14817.

(8) Jordan, J. B.; Poppe, L.; Xia, X.; Cheng, A. C.; Sun, Y.; Michelsen, K.; Eastwood, H.; Schnier, P. D.; Nixey, T.; Zhong, W., Fragment Based Drug Discovery: Practical Implementation Based on <sup>19</sup>F NMR Spectroscopy. *J. Med. Chem.* **2012**, *55* (2), 678-687.

(9) Divakaran, A.; Kirberger, S. E.; Pomerantz, W. C. K., SAR by (Protein-Observed) <sup>19</sup>F NMR. *Acc. Chem. Res.* **2019**, *52* (12), 3407-3418.

(10) Troelsen, N. S.; Shanina, E.; Gonzalez-Romero, D.; Danková, D.; Jensen, I. S. A.; Śniady, K. J.; Nami, F.; Zhang, H.; Rademacher, C.; Cuenda, A.; Gottfredsen, C. H.; Clausen, M. H., The 3F Library: Fluorinated Fsp<sup>3</sup>-Rich Fragments for Expeditious <sup>19</sup>F NMR Based Screening. *Angew. Chem. Int. Ed.* **2020**, *59* (6), 2204-2210.

(11) Zech, S. G.; Kohlmann, A.; Zhou, T.; Li, F.; Squillace, R. M.; Parillon, L. E.; Greenfield, M. T.; Miller, D. P.; Qi, J.; Thomas, R. M.; Wang, Y.; Xu, Y.; Miret, J. J.; Shakespeare, W. C.; Zhu, X.; Dalgarno, D. C., Novel Small Molecule Inhibitors of Choline Kinase Identified by Fragment-Based Drug Discovery. *J. Med. Chem.* **2016**, *59* (2), 671-686.

(12) Jordan, J. B.; Whittington, D. A.; Bartberger, M. D.; Sickmier, E. A.; Chen, K.; Cheng, Y.; Judd, T., Fragment-Linking Approach Using <sup>19</sup>F NMR Spectroscopy To Obtain Highly Potent and Selective Inhibitors of  $\beta$ -Secretase. *J. Med. Chem.* **2016**, *59* (8), 3732-3749.

(13) Fang, C.; D'Souza, B.; Thompson, C. F.; Clifton, M. C.; Fairman, J. W.; Fulroth, B.; Leed, A.; McCarren, P.; Wang, L.; Wang, Y.; Feau, C.; Kaushik, V. K.; Palmer, M.; Wei, G.; Golub, T. R.; Hubbard, B. K.; Serrano-Wu, M. H., Single Diastereomer of a Macrolactam Core Binds Specifically to Myeloid Cell Leukemia 1 (MCL1). *ACS Med. Chem. Lett.* **2014**, *5* (12), 1308-1312.

(14) Assemat, O.; Antoine, M.; Fourquez, J. M.; Wierzbicki, M.; Charton, Y.; Hennig, P.; Perron-Sierra, F.; Ferry, G.; Boutin, J. A.; Delsuc, M. A., <sup>19</sup>F nuclear magnetic resonance screening of glucokinase activators. *Anal. Biochem.* **2015**, *477*, 62-68.

(15) Goudreau, N.; Coulombe, R.; Faucher, A.-M.; Grand-Maitre, C.; Lacoste, J.-E.; Lemke, C. T.; Malenfant, E.; Bousquet, Y.; Fader, L.; Simoneau, B.; Mercier, J.-F.; Titolo, S.; Mason, S. W., Monitoring Binding of HIV-1 Capsid Assembly Inhibitors Using <sup>19</sup>F Ligand-and <sup>15</sup>N Protein-Based NMR and X-ray Crystallography: Early Hit Validation of a Benzodiazepine Series. *ChemMedChem* **2013**, *8* (3), 405-414.

- (16) Brasca, M. G.; Mantegani, S.; Amboldi, N.; Bindi, S.; Caronni, D.; Casale, E.; Ceccarelli, W.; Colombo, N.; De Ponti, A.; Donati, D.; Ermoli, A.; Fachin, G.; Felder, E. R.; Ferguson, R. D.; Fiorelli, C.; Guanci, M.; Isacchi, A.; Pesenti, E.; Polucci, P.; Riceputi, L.; Sola, F.; Visco, C.; Zuccotto, F.; Fogliatto, G., Discovery of NMS-E973 as novel, selective and potent inhibitor of heat shock protein 90 (Hsp90). *Biorg. Med. Chem.* **2013**, *21* (22), 7047-7063.
- (17) Casale, E.; Amboldi, N.; Brasca, M. G.; Caronni, D.; Colombo, N.; Dalvit, C.; Felder, E. R.; Fogliatto, G.; Galvani, A.; Isacchi, A.; Polucci, P.; Riceputi, L.; Sola, F.; Visco, C.; Zuccotto, F.; Casuscelli, F., Fragment-based hit discovery and structure-based optimization of aminotriazoloquinazolines as novel Hsp90 inhibitors. *Biorg. Med. Chem.* **2014**, *22* (15), 4135-4150.
- (18) Tengel, T.; Fex, T.; Emtenas, H.; Almqvist, F.; Sethson, I.; Kihlberg, J., Use of <sup>19</sup>F NMR spectroscopy to screen chemical libraries for ligands that bind to proteins. *Org. Biomol. Chem.* **2004**, *2* (5), 725-731.
- (19) Harris, R. K.; Becker, E. D.; Cabral De Menezes, S. M.; Granger, P.; Hoffman, R. E.; Zilm, K. W., Further conventions for NMR shielding and chemical shifts IUPAC recommendations 2008. *Solid State Nucl. Magn. Reson.* **2008**, *33* (3), 41-56.
- (20) Rosenau, C. P.; Jelier, B. J.; Gossert, A. D.; Togni, A., Exposing the Origins of Irreproducibility in Fluorine NMR Spectroscopy. *Angew. Chem. Int. Ed.* **2018**, *57* (30), 9528-9533.
- (21) Rigger, R.; Rück, A.; Hellriegel, C.; Sauermoser, R.; Morf, F.; Breittruck, K.; Obkircher, M., Certified Reference Material for Use in <sup>1</sup>H, <sup>31</sup>P, and <sup>19</sup>F Quantitative NMR, Ensuring Traceability to the International System of Units. *J. AOAC Int.* **2019**, *100* (5), 1365-1375.
- (22) Meiboom, S.; Gill, D., Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **1958**, *29* (8), 688-691.
- (23) CRC Handbook of Chemistry and Physics, 87th ed Editor-in-Chief: David R. Lide (National Institute of Standards and Technology). CRC Press/Taylor and Francis Group: Boca Raton, FL. 2006. 2608 pp. \$139.95. ISBN 0-8493-0487-3. *J. Am. Chem. Soc.* **2007**, *129* (3), 724-724.
- (24) Peng, C.; Frommlet, A.; Perez, M.; Cobas, C.; Blechschmidt, A.; Dominguez, S.; Lingel, A., Fast and Efficient Fragment-Based Lead Generation by Fully Automated

Processing and Analysis of Ligand-Observed NMR Binding Data. *J. Med. Chem.* **2016**, *59* (7), 3303-3310.

(25) Dalvit, C.; Fagerness, P. E.; Hadden, D. T. A.; Sarver, R. W.; Stockman, B. J., Fluorine-NMR Experiments for High-Throughput Screening: Theoretical Aspects, Practical Considerations, and Range of Applicability. *J. Am. Chem. Soc.* **2003**, *125* (25), 7696-7703.

(26) Dalvit, C.; Flocco, M.; Veronesi, M.; Stockman, B.J., Fluorine-NMR Competition Binding Experiments for High-Throughput Screening of Large Compound Mixtures. *Combinatorial Chem. High Throughput Screening* **2002**, *5* (8), 605-611.

(27) Stockman, B. J., 2-Fluoro-ATP as a Versatile Tool for <sup>19</sup>F NMR-Based Activity Screening. *J. Am. Chem. Soc.* **2008**, *130* (18), 5870-5871.

(28) Peters, T., 3 - Ligand Binding by Albumin. In *All About Albumin*, Peters, T., Ed. Academic Press: San Diego, 1995; pp 76-132.

(29) Povey, J. F.; Smales, C. M.; Hassard, S. J.; Howard, M. J., Comparison of the effects of 2,2,2-trifluoroethanol on peptide and protein structure and function. *J. Struct. Biol.* **2007**, *157* (2), 329-338.

## 9 SUPPORTING INFORMATION – ARTICLE 4

---

### **Practical Considerations and Guidelines for Spectral Referencing for Fluorine NMR Ligand Screening**

Yann Ayotte<sup>a,b</sup>, Simon Woo<sup>a,b</sup>, Steven R. LaPlante<sup>\*a,b</sup>

<sup>a</sup> *INRS – Centre Armand-Frappier Santé Biotechnologie, 531 boulevard des Prairies, Laval, Québec, H7V 1B7, Canada.*

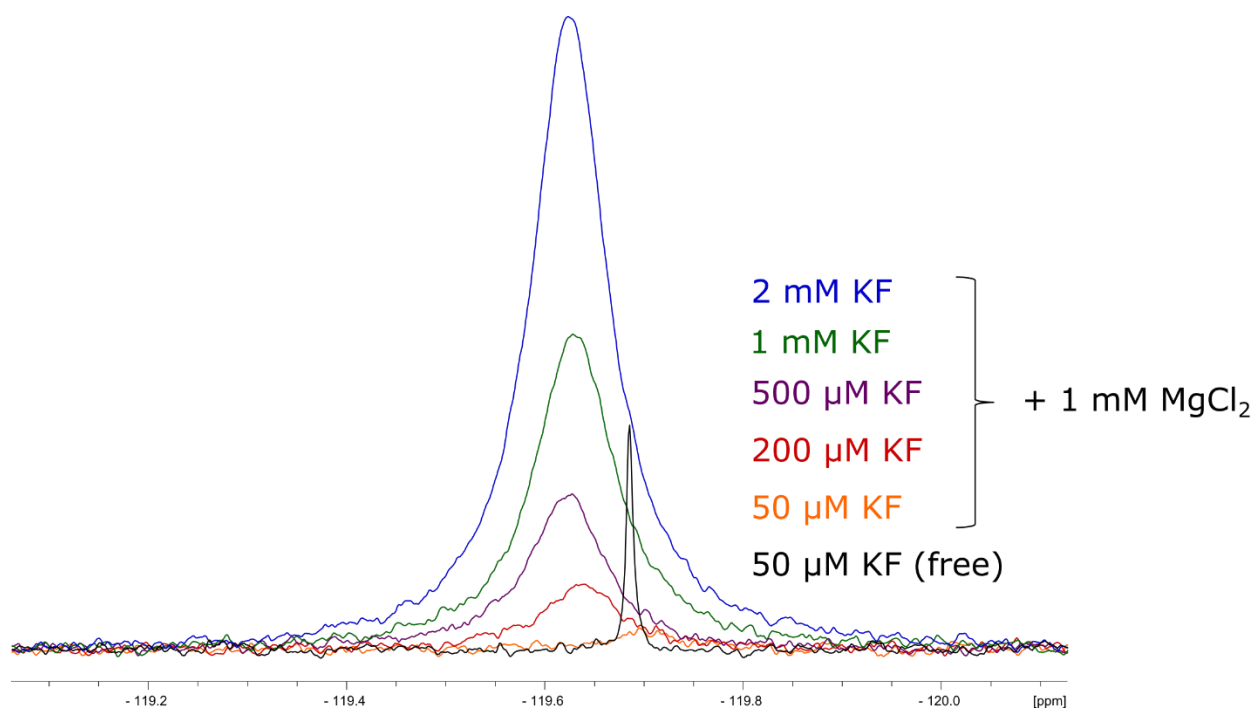
<sup>b</sup> *NMX Research and Solutions Inc., 500 boulevard Cartier Ouest, Suite 6000, Laval, Québec, H7V 5B7, Canada*

\* Corresponding author

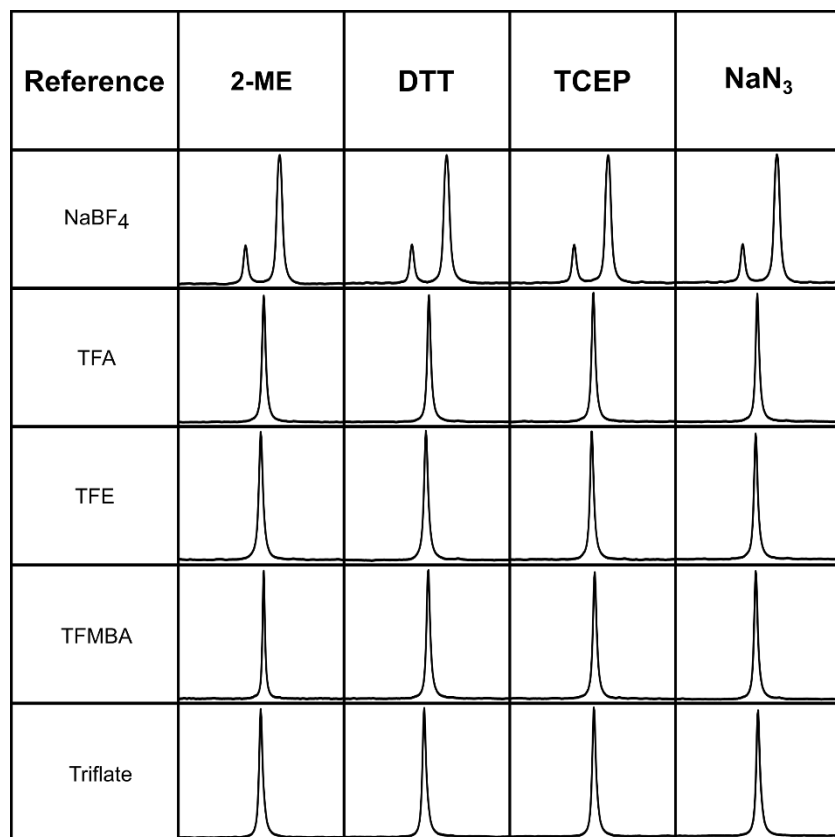
Published in : ACS Omega 2022, 7, 15, 13155-13163

DOI : 10.1021/acsomega.2c00613

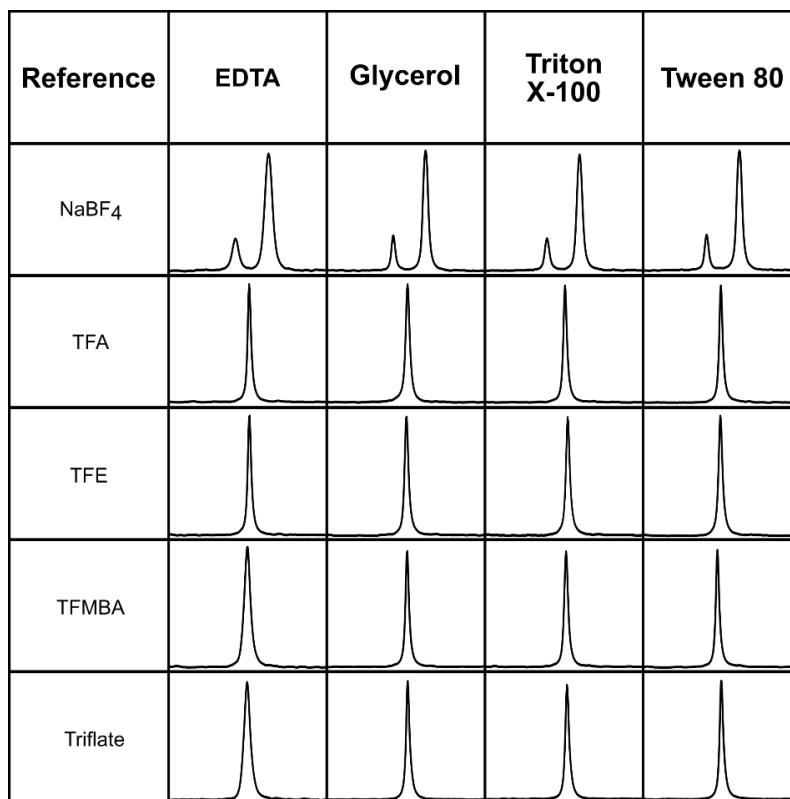
**Supplementary figures**



**Figure S1.** Titration of potassium fluoride (KF) against 1 mM  $\text{MgCl}_2$  results in broadening of the KF  $^{19}\text{F}$  resonance.

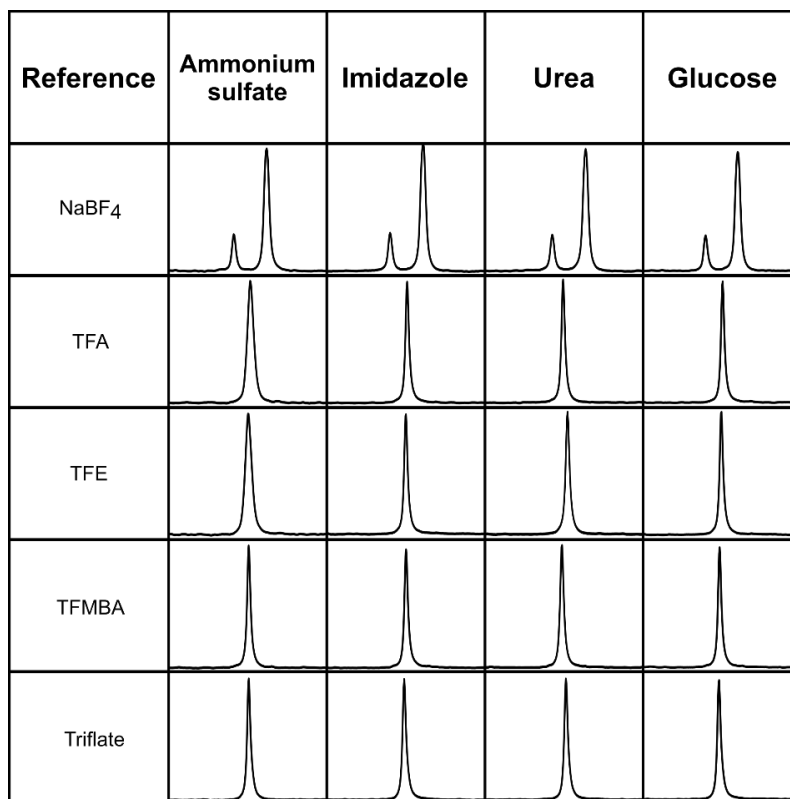


**Figure S2.** Evaluating the compatibility of five fluorine reference candidates with 0.5% 2-Mercaptoethanol (2-ME), 10 mM Dithiothreitol (DTT), 10 mM tris(2-carboxyethyl)phosphine (TCEP), and 0.1% sodium azide (NaN<sub>3</sub>).

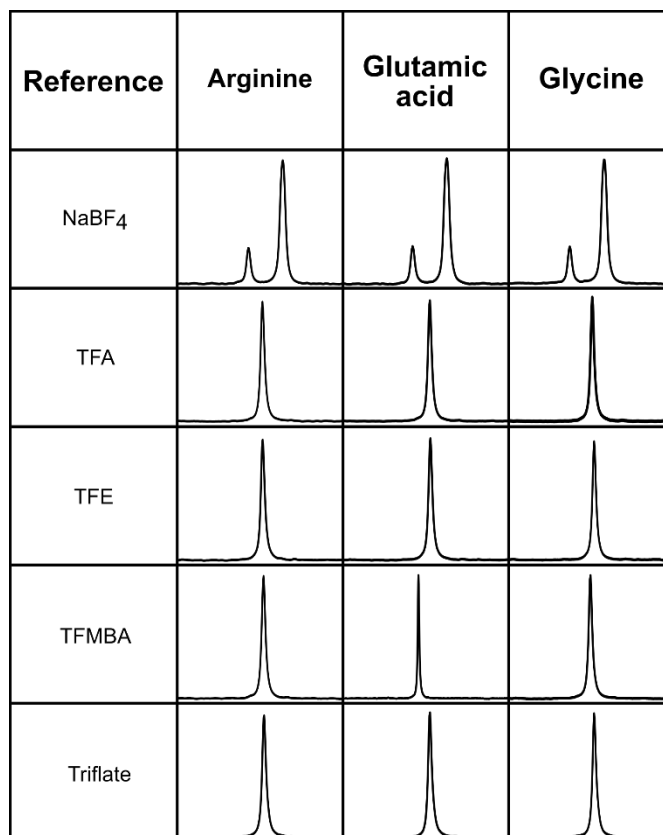


**Figure S3.** Evaluating the compatibility of five fluorine reference candidates with 100 mM Ethylenediaminetetraacetic acid (EDTA), 5% glycerol, 0.5% Triton X-100, and 0.1 % Tween 80.

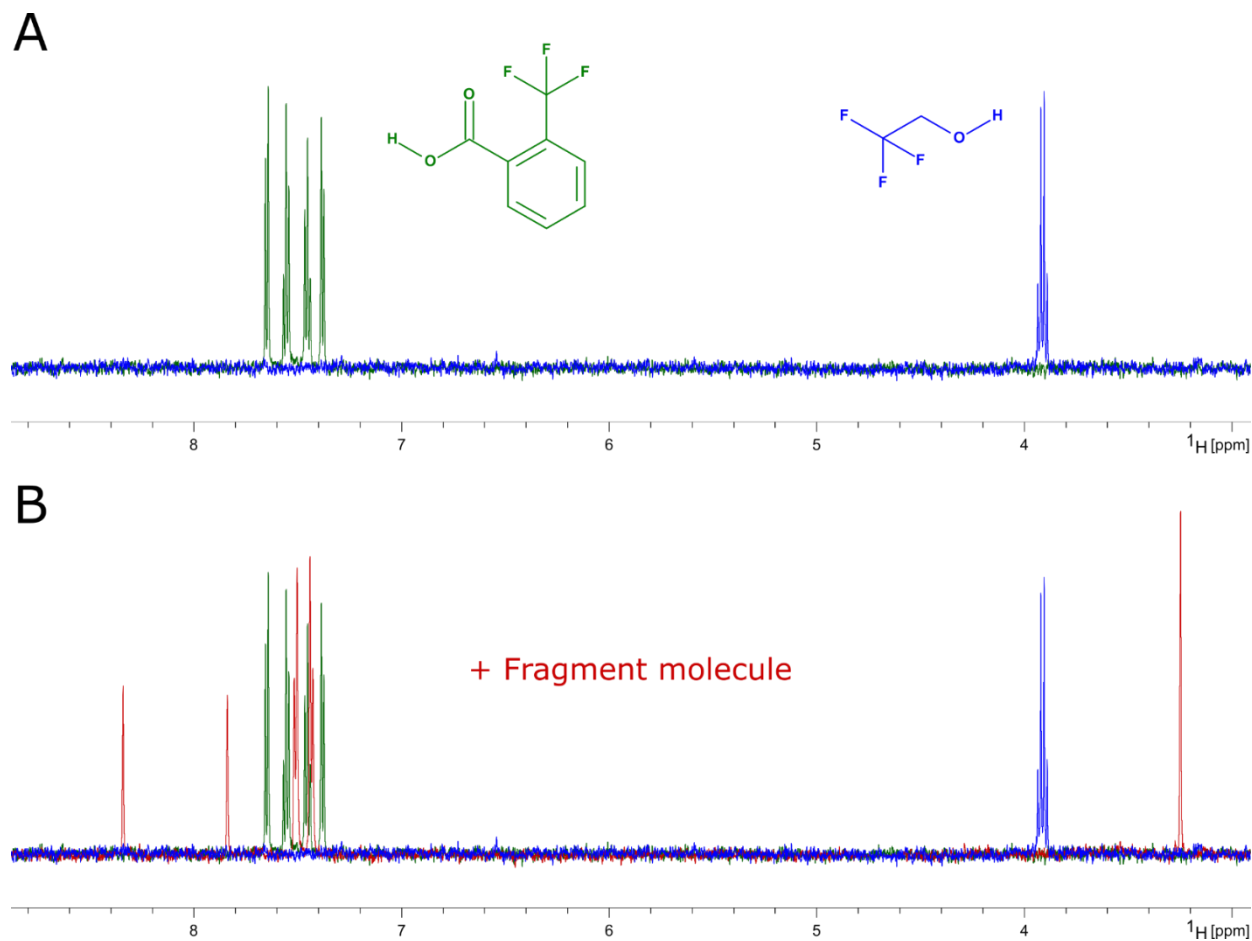




**Figure S4.** Evaluating the compatibility of five fluorine reference candidates with 250 mM ammonium sulfate, 100 mM imidazole, 1 M urea, and 5 mM glucose.



**Figure S5.** Evaluating the compatibility of five fluorine reference candidates with 100 mM Arginine, 100 mM glutamic acid, and 100 mM glycine.



**Figure S6.** Among the final candidates, TFMBA (green spectra) and TFE (blue spectra) exhibit  $^1\text{H}$  resonances (A). However, only TFMBA possesses aromatic  $^1\text{H}$  resonances which can hinder analysis of traditional molecules which oftentimes contain aromatic  $^1\text{H}$ . An example of such aromatic overlap is evident upon addition of a randomly chosen aromatic fragment molecule (B, red spectrum).

### Spectrometer stability

**Table S1.** Stability of the measurements using spectrometer indirect referencing was evaluated by acquiring 1D  $^{19}\text{F}$  NMR spectra of 3 trifluoroethanol samples at three different time points. The average variation in chemical shift between time points was calculated, along with the standard deviation. The nominal concentration used was 200  $\mu\text{M}$ .

---

**Time points**

**Average variation in chemical shift**

**Standard deviation**

---

0 h to 12 h	0.0758553 Hz (0.0001343 ppm)	0.0369001 Hz (0.0000653 ppm)
12 h to 24 h	0.0948662 Hz (0.0001680 ppm)	0.0036885 Hz (0.0000065 ppm)

### Compound information

Reference	Supplier	Catalog #	CAS	Formula
Hexafluorobenzene	Sigma	326720	392-56-3	C <sub>6</sub> F <sub>6</sub>
Fluorobenzene	TCI America	F0034	462-06-6	C <sub>6</sub> H <sub>5</sub> F
2,2,2-Trifluoroethanol	TCI America	T0435	75-89-8	CF <sub>3</sub> CH <sub>2</sub> OH
Trifluoroacetic acid-d	Sigma	152005	599-00-8	CF <sub>3</sub> CO <sub>2</sub> D
Benzotrifluoride (Trifluorotoluene)	TCI America	T0439	98-08-8	C <sub>6</sub> H <sub>5</sub> CF <sub>3</sub>
1,2-Difluorobenzene	TCI America	D1628	367-11-3	C <sub>6</sub> H <sub>4</sub> F <sub>2</sub>
Sodium tetrafluoroborate	Sigma	202215	13755-29-8	NaBF <sub>4</sub>
2-(Trifluoromethyl)benzoic acid	Sigma	196886	433-97-6	C <sub>8</sub> H <sub>5</sub> F <sub>3</sub> O <sub>2</sub>
Potassium fluoride	Sigma	221872	13455-21-5	KF
Sodium trifluoromethanesulfonate	Sigma	367907	2926-30-9	CF <sub>3</sub> SO <sub>3</sub> Na

### Protein information

Protein	Supplier	Catalog #
Bovine serum albumin	Sigma	A2153

Elastase	Sigma	E7885
Lysozyme	Sigma	L6876
Trypsin	Sigma	T1426

## 10 ARTICLE 5 – PROOF-OF-CONCEPT STUDY OF A FRAGMENT-BASED PHENOTYPIC SCREENING APPROACH ON LEISHMANIA PARASITES

---

### Fragment-based Phenotypic Lead Discovery: Cell-based Assay to Target Leishmaniasis

Yann Ayotte,<sup>[a]</sup> François Bilodeau,<sup>[b]</sup> Albert Descoteaux,<sup>[a]\*</sup> and Steven R. LaPlante<sup>[a]\*</sup>

[a] INRS – Institut Armand-Frappier  
531, boulevard des Prairies, Laval, Québec, Canada, H7V 1B7

[b] NMX Research and Solutions Inc.  
500 boulevard Cartier, Laval, Québec, Canada, H7V 5B7

Published in : ChemMedChem 2018, 13(14), 1377-1386

DOI : 10.1002/cmdc.201800161

#### Authors contributions:

Yann Ayotte designed and performed the experiments and data analysis, generated the figures, and wrote the manuscript under supervision from Steven Laplante, Albert Descoteaux and François Bilodeau. François Bilodeau provided support for the medicinal chemistry. All authors were involved in revising the manuscript.

## Abstract

A rapid and practical approach for the discovery of new chemical matter for targeting pathogens and diseases is described. Fragment-based phenotypic lead discovery (FPLD) combines aspects of traditional fragment-based lead discovery (FBLD), which involves the screening of small-molecule fragment libraries to target specific proteins, with phenotypic lead discovery (PLD), which typically involves the screening of drug-like compounds in cell-based assays. To enable FPLD, a diverse library of fragments was first designed, assembled, and curated. This library of soluble, low-molecular-weight compounds was then pooled to expedite screening. Axenic cultures of *Leishmania* promastigotes were screened, and single hits were then tested for leishmanicidal activity against intracellular amastigote forms in infected murine bone-marrow-derived macrophages without evidence of toxicity toward mammalian cells. These studies demonstrate that FPLD can be a rapid and effective means to discover hits that can serve as leads for further medicinal chemistry purposes or as tool compounds for identifying known or novel targets.

## Introduction

### Current lead discovery strategies

Currently there are only a handful of established strategies available for discovering new chemical matter or leads that can then be developed into future medicaments to combat diseases. Of course, each strategy has its pros and cons. Protein-based biologics are effective, specific, and relatively straightforward to develop, but they generally must be injected for delivery purposes and they are most appropriate for extracellular targets. Substrate-based mimics have been successful as a quick start for designing drugs from peptides, nucleosides and sugars, but often are difficult to make and have specificity issues, especially those that attach to the target via covalent “warheads”. High-throughput screens (HTS) involving large collections of drug-like molecules (>1 million) are widely used by large pharma, but design cycles are lengthy and successes have been limited. Moreover, medium and small institutions simply cannot run such screens due to high costs, and they do not have easy access to such large compound collections. Phenotypic or cell-based screening can be a rapid method for discovering new chemical inhibitors, as it usually relies on direct measurement of the effect on intact cells or organisms. This approach has the advantage of not being biased toward a known target and therefore has the potential to identify new molecular targets.<sup>1</sup> For example, phenotypic screening approaches have

been applied to discover chemical starting material for neglected diseases,<sup>2</sup> and successes have been reported by Medicine for Malaria Venture for the advancement of compounds to preclinical and clinical studies.<sup>3</sup> However, phenotype screens have historically implicated the use of large collections of drug-like molecules which is impractical for medium to small institutions. Furthermore, subsequent identification of the exact target of hits can be arduous, but innovative elucidation methods have been developed, for example, those using mass spectrometry and covalent labeling strategies.

Fragment-based lead discovery (FBLD) was introduced a couple of decades ago as a means to screen small collections of low-molecular-weight compounds (fragments) to identify binders to target proteins from which larger drug-like compounds can be designed. Drugs designed via this method have entered the clinic and market, and many more are in the pipeline. FBLD only requires small libraries (1000–2000 compounds),<sup>4</sup> which makes it amenable for a wide range of institutions to target many diseases.<sup>5</sup> However, there are difficulties in initiating FBLD programs, in that milligram quantities of target protein must be expressed in order to use biophysical methods such as NMR spectroscopy.<sup>6</sup> Also, molecular artifacts are widespread, medicinal chemistry efforts are difficult at the initial stage, and appropriate technologies need to be developed. Nonetheless, FBLD is an excellent and promising strategy especially given the current environment where biotech companies and universities are increasingly bearing the responsibility for discovering the leads for drugs.

### **Hybrid phenotypic and fragment screening**

Here, we describe how we developed a hybrid strategy called fragment-based phenotypic lead discovery (FPLD) that borrows concepts from phenotypic-based lead discovery (PLD) and from target-based FBLD. The overall idea is to fast-track lead discovery by exploiting the advantages of FBLD where a small library collection is rapidly screened, with the advantage of PLD where many potential targets can be inhibited.<sup>7</sup> Moreover, PLD projects tend to have stronger chains of translatability in the area of infectious diseases<sup>8</sup> and small fragment molecules may cross cell membranes more easily. However, due to the small size of these compounds, there was a concern about the potential impact of promiscuous inhibition, which would translate to a large number of hits in primary screens. Fortunately, this was not the case in this exercise. A recent report used a similar approach to identify small-molecule protein interactions using photoaffinity probes containing fragments common to several drug structures to globally map small-molecule fragment–protein interactions in human cells. They found clear evidence of structure–activity relationships (SARs) for most interactions, which supports the fact that these interactions reflect



real recognition events, even at such low binding affinity.<sup>9</sup> However, such compounds are not readily available commercially and must be synthesized, which can limit the size of a library. This work therefore focuses on the use of a curated fragment library with PLD which is should be practical and accessible. This hybrid strategy was applied and tested to identify fragment compounds that target the parasite *Leishmania*. The remaining section of this Introduction provides a brief overview of leishmaniasis.

## **Leishmaniasis**

Parasites of the genus *Leishmania* are responsible for a spectrum of pathologies in humans collectively termed leishmaniasis. These parasites are the second highest cause of death by parasitic infections, right behind malaria, and figure among the 17 most important neglected tropical diseases listed by the World Health Organization. According to their estimation, 12 million people are currently affected and a tenth of the world population could be at risk of contracting the disease.<sup>10</sup> Even though leishmaniasis is still considered as a disease affecting mainly the poorest countries, it has been detected in the United States among soldiers coming back from the Middle East, as well as in dogs and cats.<sup>11</sup> It is also considered as a threat in southwestern Europe and it is expected that climate change will promote propagation in Europe in the near future.<sup>12</sup>

Leishmaniasis is a vector-borne disease caused by at least 20 different species of the intracellular parasite *Leishmania* transmitted by more than 30 species of sandflies. *Leishmania* exists under two developmental forms: an extracellular flagellated promastigote form that replicates in the sandfly midgut where the parasite undergoes differentiation from a noninfectious to an infectious form and an intracellular amastigote form that strives within the endolysosomal compartment of mammalian host macrophages.<sup>10, 11</sup> There are four major possible clinical manifestations of the disease (cutaneous, mucocutaneous, visceral, and post kala-azar dermal leishmaniasis) depending on the *Leishmania* species involved and most importantly, the host immune response.<sup>13</sup>

## **Current treatments and need for new alternatives**

Considering the lack of an effective vaccine, control of leishmaniasis relies mainly on chemotherapy and vector control.<sup>14</sup> For more than half a century, antimonies like sodium stibogluconate have been the first line of treatment for the disease. However, these drugs can cause severe side effects and treatment failure is increasingly high, mostly due to emergence of resistance. Since then, some other drugs like miltefosine, amphotericin B, and paromomycin have

been developed, but all those treatments have different flaws such as toxicity, high cost, need for hospitalization, teratogenicity, or resistance.<sup>10, 12b, 14</sup>

Clearly, there is an urgent need to discover other treatment options, especially in light of the emergence of resistance worldwide. Ideally, these new treatments should be effective, orally administered, nontoxic and affordable. Here, we explore the potential of FPLD to discover new chemical matter that may satisfy such stringent requirements.

## Results and Discussion

### Overview

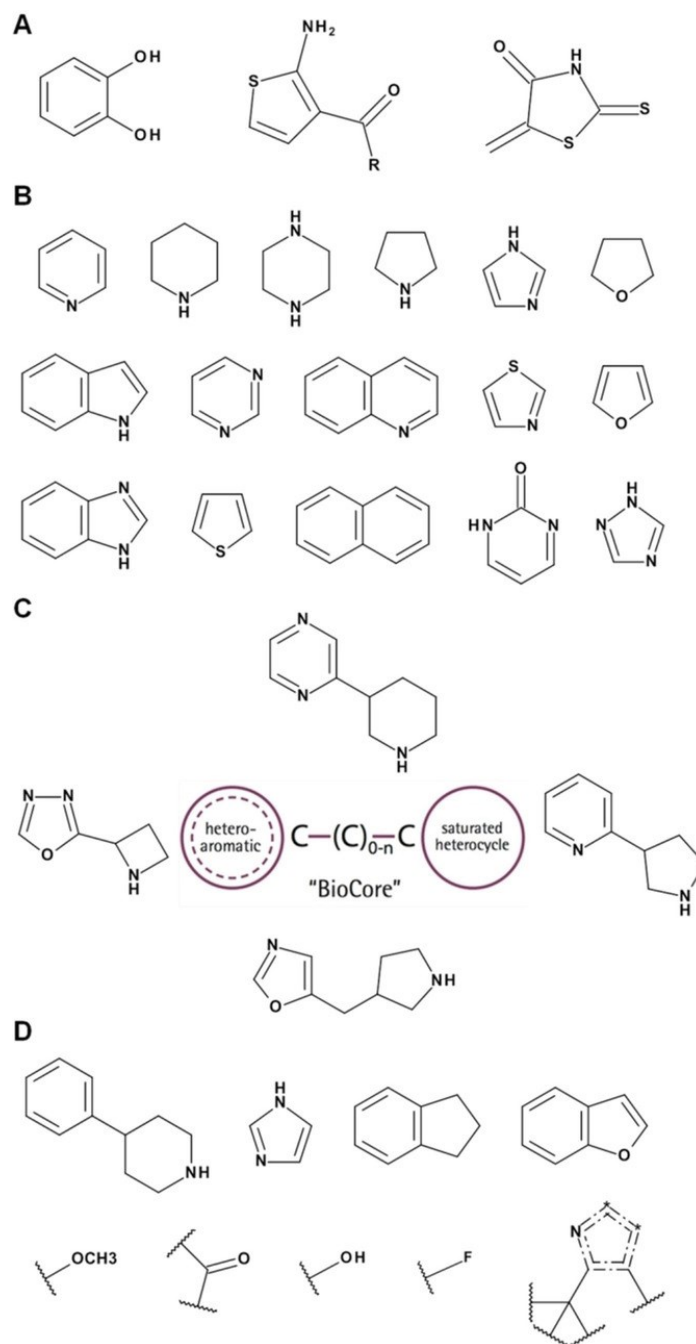
One of the main challenges of leishmaniasis drug development relies on the fact that protozoan parasites like *Leishmania* are eukaryotic cells which means that they share common characteristics with the host cells, making the development of effective and selective treatment a more difficult task. Thus, cell-based phenotypic screening can have advantages given that hit compounds can easily be counterscreened against other types of cells or in the case of parasitic infection, screened against the parasite within host cells.<sup>2</sup>

In this study, we applied FPLD to screen a library of 1604 fragments on *Leishmania* promastigotes and counterscreened against bone marrow-derived macrophages (BMM) to eliminate any compound that would be cytotoxic for the macrophages. The positive hits were then tested on *L. amazonensis*-infected macrophages to select for the molecules that remain active against intracellular parasites. These fragments of interest were next tested to determine their dose-response curves and quantitatively estimate their activities via EC<sub>50</sub> values. SARs were then evaluated by testing fragments analogues with similar chemical scaffolds as the active compounds.

### Fragment library design

A collection of over 8000 low-molecular-weight compounds were evaluated and triaged using cheminformatics, then curated by NMR spectroscopy to select a final library of 1604 compounds with the desired characteristics (Figure 1). The cheminformatics strategy employed stringent filtering methods to remove undesirable compounds based on published rules (PAINS, BMS filters, FAFDrugs, Kazius and Bursi toxicophores, Lilly MedChem Rules).<sup>15</sup> Fragments were systematically prioritized based on desired calculated physicochemical properties, substructure matching, and counts based on 2D and 3D molecule structure. This resulted in an enriched subset

of compounds that had desirable substructures found in current drugs and natural products, functional group diversity, three-dimensionality, and span of distinct chemical space from existing commercial collections. Experimental curation then followed by preparing compound samples at 300  $\mu\text{m}$  nominal concentration and recording of  $^1\text{H}$  NMR spectra of individual compounds. The NMR-based experimental profiling allowed the careful selection of highly soluble and well-behaved fragments with desirable physicochemical (such as high water solubility) and stability characteristics. NMR triaging therefore resulted in the removal of compounds exhibiting undesirable characteristics such as insufficient solubility, degradation, potential structural liabilities, poor resolution of NMR resonances, or signs of aggregation. It was expected that this experimental curation would minimize sources of false positive hits (due to aggregation) and attrition for fragment screening. To expedite screening, compounds were pooled based on the distribution of NMR chemical shifts to minimize overlap of compounds resonances and therefore facilitate deconvolution steps for target-based screening. This resulted in 7–12 compatible compounds per pool; chemically incompatible compounds were not pooled together. As a result, the 1604 compounds were divided into 169 pools which was a manageable size for low-throughput screening purposes by NMR. A more detailed report about the library design and attributes will be published elsewhere. This library was kindly provided by NMX Research and Solutions Inc. (<https://www.nmxresearch.com/>), and a subset of this library is available as Key Organics' BIONET library (<https://www.keyorganics.net/services/bionet-products/fragment-libraries/>).

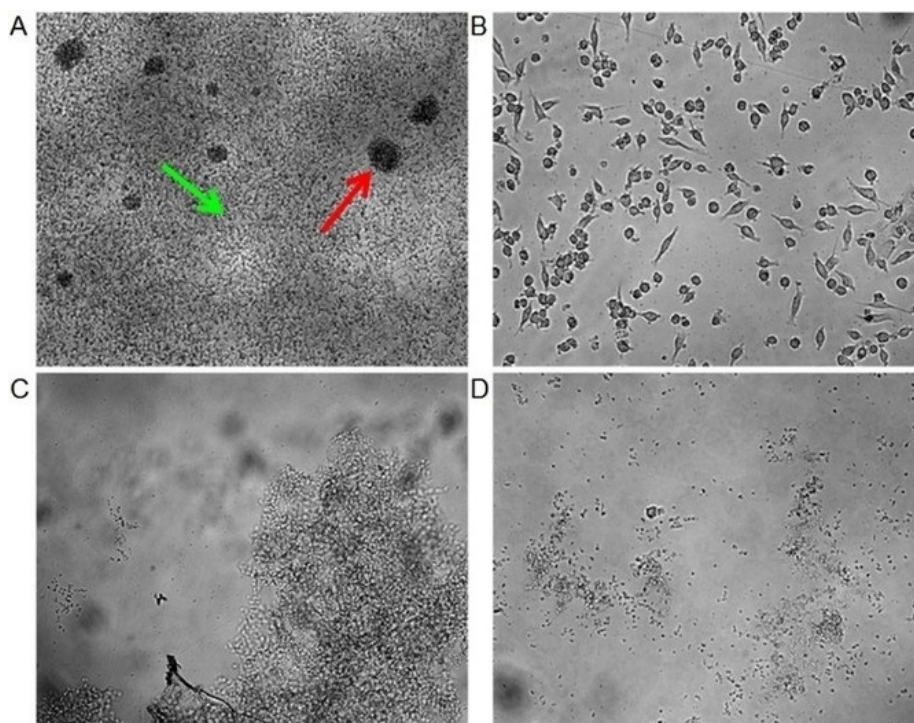


**Figure 1.** Overview of library curation. A) Undesirable compounds were removed using cheminformatics (PAINS, BMS, FAFDrugs, Lilly MedChem Rules) and NMR filters. The library was enriched by adding B) privileged scaffold and C) BioCores. D) Diversity was then maximized.

### Testing the activities of the fragment pools

To expedite testing, 169 pools were screened against two different *Leishmania* species (*L. amazonensis* and *L. donovani*) promastigotes at a concentration of 166  $\mu\text{M}$  in axenic cultures.

Since a leishmanicidal drug needs to inhibit the parasites without harming the macrophages, the same pools were also counterscreened on macrophages to test for cytotoxicity. The samples were visualized by microscopy at 2, 24, 48 and 72 h post-treatment, and then attributed a value ranging from 0 (complete cellular death) to 5 (equivalent to control) (see Supporting Information). The reference time for macrophage viability was set at 48 h due to the fact that after 72 h, even the controls were starting to decay due to saturation of the culture. Among the pools, 50 showed a high leishmanicidal effect (score of 0) on both parasite species. However, only 16 of these had little or no cytotoxic effect on macrophages. Figure 2 is an example of a pool that showed inhibitory effect on promastigotes while having no notable effect on macrophages. Figure 2 A shows control parasites for *L. amazonensis* only because there was no discernable difference between the controls for *L. amazonensis* and *L. donovani*, whereas Figure 2 B is the counterscreen on macrophages. On Figure 2 C and Figure 2 D, the compound pool is tested of *L. donovani* and *L. amazonensis*, respectively and both samples show very high parasite death.



**Figure 2.** Example of a leishmanicidal pool (P1F9) at a concentration of 166  $\mu\text{M}$  ( $t=48$  h) A) Control *L. amazonensis* promastigotes. Green arrow: high parasitic density. Red arrow: rosettes (clusters) of parasites which are a sign of growth. B) Macrophages toxicity counterscreen. C) *L. donovani*. D) *L. amazonensis*. Visual inspection of the samples shows that the P1F9 pool possesses a leishmanicidal effect on two *Leishmania* species (C–D) without any clear cytotoxic effect on the macrophages (B).

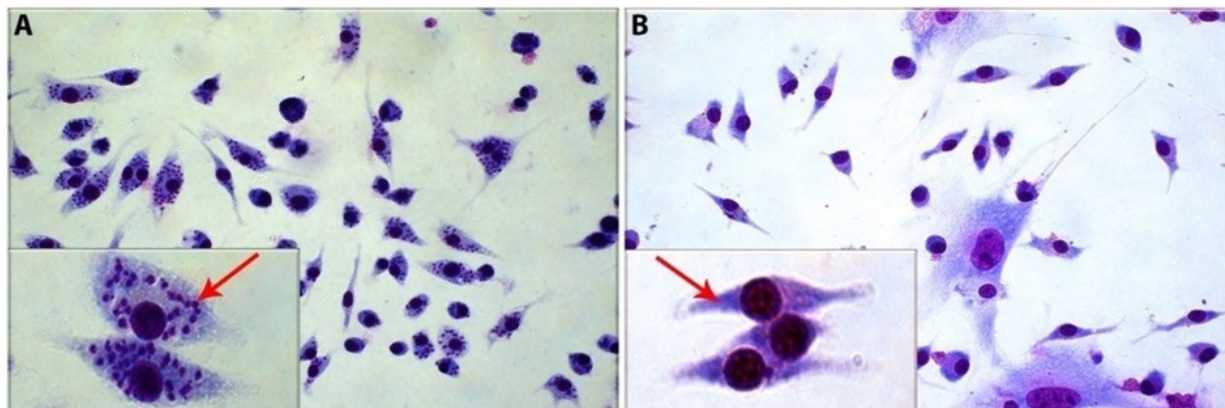
## Testing and verifying the activities of singleton fragments in promastigotes and macrophages

Most pools (with an exception of two) displaying medium to high inhibitory effect on *L. amazonensis* promastigotes (54 pools) were also effective against *L. donovani* promastigotes (86 pools), but not all the pools effective against *L. donovani* were effective against *L. amazonensis*. Due to the low throughput of the assays, we therefore assumed that the latter was more resistant to the compounds and that focus could be put on *L. amazonensis* in the following experiments under the working assumption that an efficient compound on *L. amazonensis* will likely also be efficient on *L. donovani*. A deconvolution of the five most active nontoxic pools, along with an intermediary active and an inactive resulted in 65 individual compounds. These singletons were then screened once again at 166  $\mu\text{m}$  against *L. amazonensis* promastigotes and counterscreened on macrophages. Of the 65 initial singletons, only **1 (MAGAN-1719)** and **2 (MAGAN-2421)** had an inhibitory effect on the parasites without killing the macrophages. Such effect is likely due to some pool effect caused by the interaction of multiple molecules. A higher-throughput assay could allow for the study of the effect of various combination of compounds, as there is growing evidence for polypharmacology where a clinical effect is the result of interaction between one or multiple drugs against different targets.<sup>16</sup>

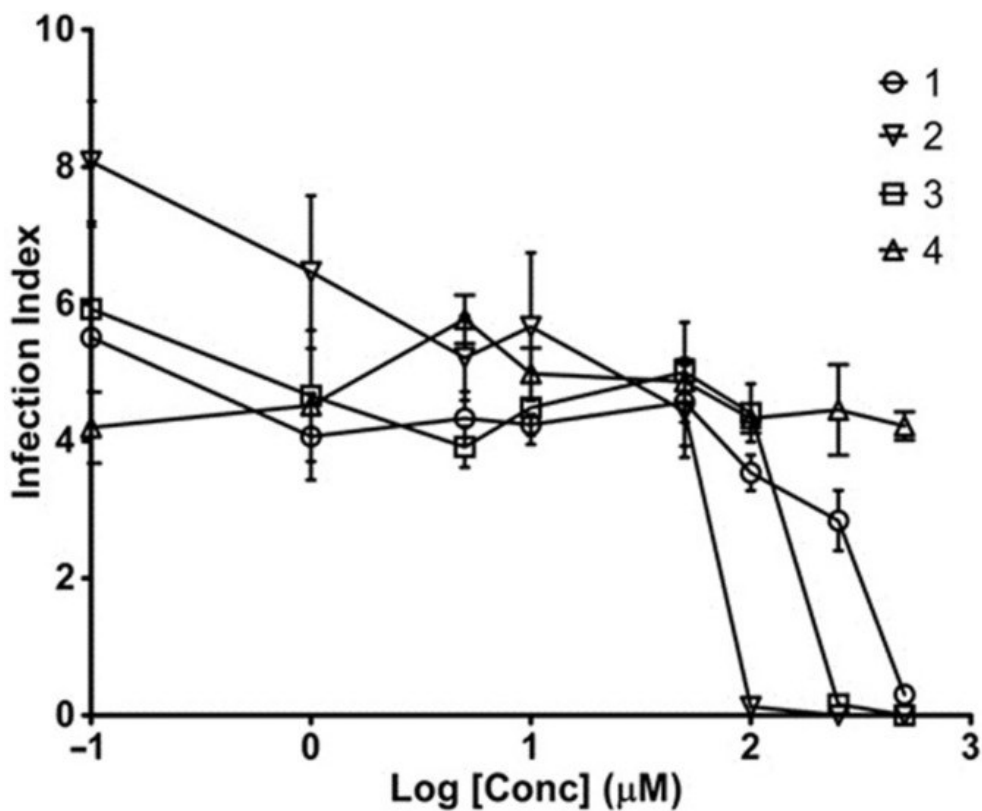
## Testing singleton compounds in infected macrophages

For a leishmanicidal drug to be clinically relevant, it should selectively kill the amastigote forms that replicate within host macrophages. Therefore, compounds **1** and **2** were tested at 166  $\mu\text{m}$  on *L. amazonensis*-infected macrophages (Figure 3). The two compounds caused a decrease in infection index and were therefore tested at eight different concentrations ranging from 0.1 to 500  $\mu\text{m}$  to determine if the effect was dose-dependent (Figure 4). Because the chemical scaffolds of compounds **1** and **2** were very similar, two other related compounds **3 (MAGAN-2242)** and **4 (MAGAN-1107)** that showed no leishmanicidal effect on promastigotes were used as negative controls. Interestingly, compound **3** turned out to have an inhibitory effect on the parasite in infected macrophages. The fact that **3** did not show up on primary screens can be explained by the fact that although screening on axenic promastigotes is a lot more efficient and cheap, there is only a modest confirmation rate of promastigote active compounds in intracellular assays that could results from the difference between the assays or parasite stage-specific characteristics.<sup>17</sup> Previous works have shown that around 50 % of hits found to be active against

intracellular amastigote assays were not active against promastigotes.<sup>18</sup> In fact, having a higher-throughput assay could allow primary screens to be done directly on intracellular amastigotes.



**Figure 3.** Macrophages infected with *L. amazonensis* parasites at 48 h post-treatment. Some cells have been zoomed and put at the bottom left in order to better observe the parasites inside the macrophages. The red arrows indicate the presence or absence of parasites inside the host cells. A) DMSO vehicle control. B) Compound 2 at a concentration of 166 μM.



**Figure 4.** Infection index (number of parasites per 100 macrophages) for the three hits and an inactive compound ( $t=48$  h). Each sample was tested in triplicate, experiments were repeated twice, and the data are presented as means  $\pm$ SD.

#### Effect on TNF and nitric oxide secretion

Because TNF and nitric oxide (NO) secretion by the macrophage are closely linked to the capacity of the host cell to control the parasite infection,<sup>19</sup> their secretion were measured in the previous screens to determine if the compounds could have an indirect inhibitory effect on the parasite by stimulating microbicidal activity of the macrophages. Some compounds were found to induce the production of TNF or NO in non-infected macrophages (see Supporting Information) but were ineffective on *Leishmania*-infected macrophages, which can be expected considering that *Leishmania* impairs host cell signaling pathways.<sup>20</sup> These results combined with the fact that compounds 1 and 2 had an effect on axenic promastigotes alone support the hypothesis that the inhibitory effect likely comes from a direct interaction of the compounds with the parasite.

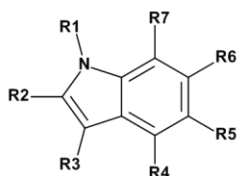
#### Establishing early structure–activity relationships

To get some idea of the structure–activity relationship (SAR) of these compounds, readily accessible fragments with similar scaffolds were also tested against infected macrophages at multiple concentrations for 96 h to maximize the effect of the sodium stibogluconate control, as trivalent antimony has been shown to take up to 96 h to reach full leishmanicidal potential.<sup>21</sup> To increase macrophage viability during this period, medium was renewed after 48 h. Tables 1 and 2 show the EC<sub>50</sub> values for two compound series.

The first series of indole fragments illustrated in Table 1 shows that moving the hydroxy group from R6 (2) to R7 (12) resulted in a modest 2.8-fold increase in potency whereas putting the OH group on R5 (11) resulted in a complete loss of activity. Combining a hydroxy group on R2 with a methyl on R1 also ensued into an inactive compound in the case of 4. Replacing the R6 hydroxy (2) for an amine (10) did not produce any significant change in potency.



**Table 1.** SAR for series 1 fragments.

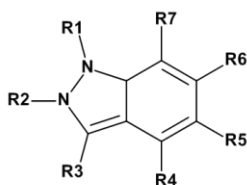


Compound	Groups	EC <sub>50</sub> (μM)
2	R6 = OH	30.9
4	R1 = Me R2 = OH	> 500
10	R6 = NH <sub>2</sub>	28.9
11	R5 = OH	> 500
12	R7 = OH	11.2

Unless defined otherwise, R=H. Each sample was tested in triplicate, experiments were repeated twice.

For the indazole compounds in Table 2, moving the hydroxy group from R4 (**3**) to R6 (**8**) decreased the potency by 8.7-fold. Switching the hydroxy for a hydroxymethyl on R6 led to an inactive compound **7**. Changing R6 for a nitrile also induced a loss of activity for **6**. Adding a methyl group on R1 (**5**) had no significant effect relative to **3**, but moving this methyl to R3 led to a gain of 3.7-fold activity (**1**). Putting a methoxy on R5 led to an inactive compound (**9**).

**Table 2.** SAR for series 2 fragments.



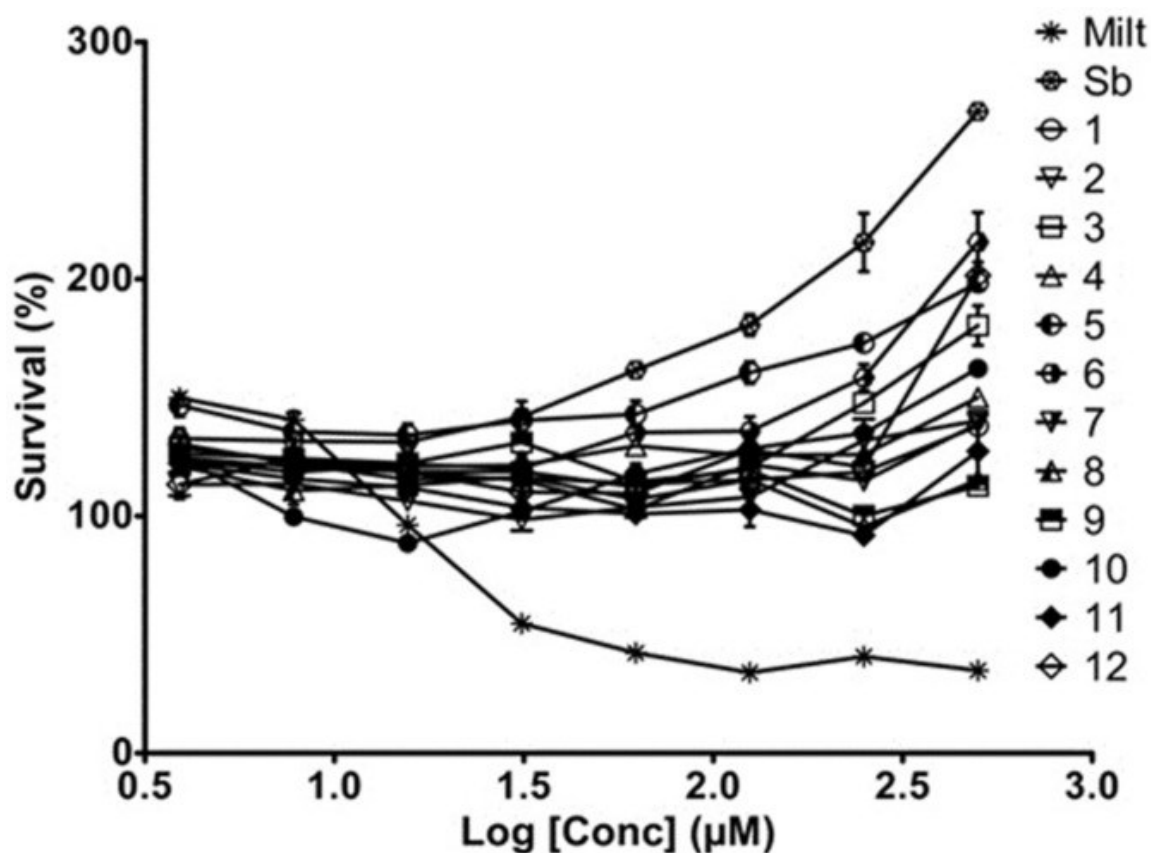
Compound	Groups	EC <sub>50</sub> (μM)
1	R3 = Me R4 = OH	4
3	R4 = OH	14.8
5	R1 = Me R4 = OH	13.4
6	R6 = CN	> 500
7	R6 = CH <sub>2</sub> OH	> 500
8	R6 = OH	128.9
9	R5 = O-Me	> 500

Unless defined otherwise, R=H. Each sample was tested in triplicate, experiments were repeated twice.

The fact that such important variations of activity is observed means that one could do systematic SAR and better understand the contribution of chemical groups. Moreover, compounds **1**, **3**, **5** and **12** showed leishmanicidal activity similar to the reference drug sodium stibogluconate (Supporting Information), and such activities in the low micromolar range meet one of the hit and lead criteria from the Global Health Innovative Technology (GHIT) postulating that a lead should demonstrate an EC<sub>50</sub> below 10 μm for leishmaniasis,<sup>22</sup> which is further discussed below.

Cytotoxicity of the compounds against BMMs was measured by MTT assay (Figure 5). Interestingly, at the concentrations tested, none of those fragments had a negative effect on macrophages viability and the only compound that had a negative impact was the reference drug miltefosine (Milt). The fact that most compounds had a tendency toward an increase survival could be explained by the fact that primary macrophage cells have a propensity to undergo apoptosis

when left in culture unstimulated. Therefore, compound **1** would possess a selectivity index (SI) ( $CC_{50}/EC_{50}$ ) higher than 125 for BMMs which is above the minimum 100-fold selectivity window previously suggested for drugs against leishmaniasis.<sup>22</sup>



**Figure 5.** Macrophage cytotoxicity assay using MTT ( $t=48$  h). Data are reported as survival percentage of DMSO-treated cells. The reference drug miltefosine (Milt) is the only compound having a negative effect on macrophage viability. At the concentrations tested, the fragments as well as the second reference drug sodium stibogluconate (Sb) had no negative effect on the BMMs. Each sample was tested in triplicate, experiments were repeated two times, and the data are presented as the means  $\pm$ SD.

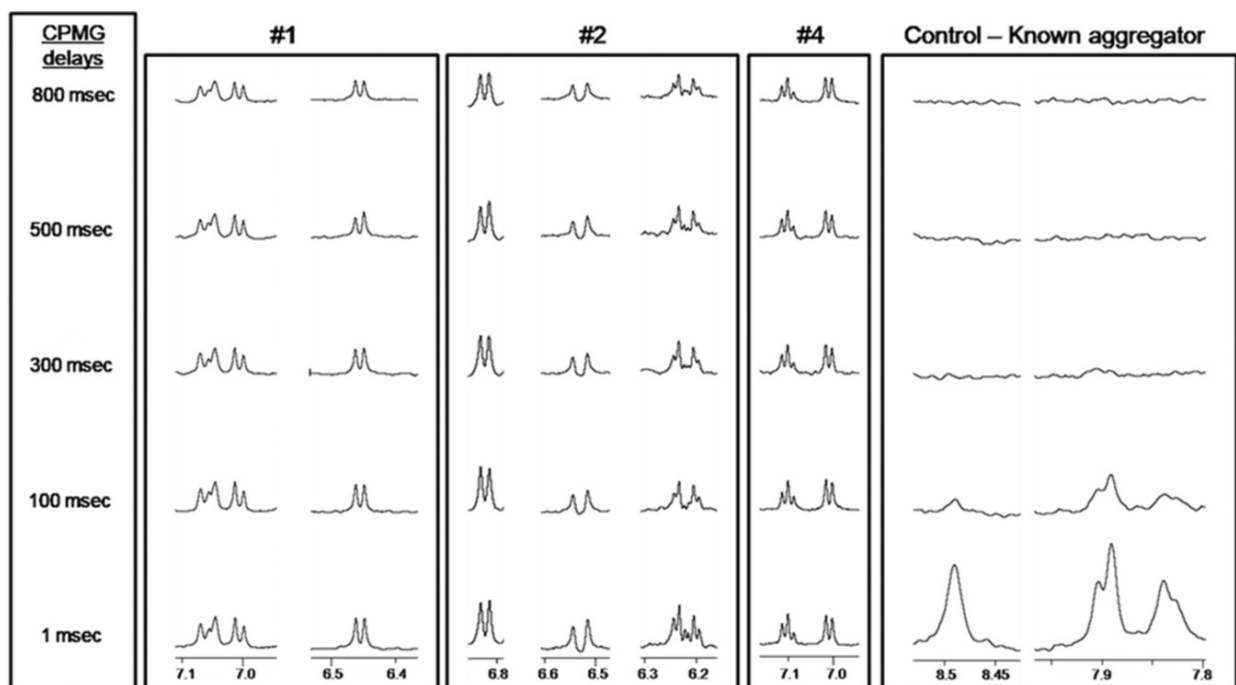
However, more optimization must be done to further determine SAR; in vivo experiments in models such as BALB/c mice or hamsters<sup>23</sup> will need to be done when satisfying leads will be obtained. Furthermore, we are in the process of generating drug-resistant parasites for the most promising compounds to: 1) determine if the parasites easily develop resistance to these

compounds, and 2) identify the target(s) of those compounds through whole genome sequencing.<sup>24</sup>

### **Ruling out compound aggregation as a potential mechanism**

A compound's physical properties in aqueous media can influence its behavior and compound self-aggregation has been associated with promiscuous hits in drug screens,<sup>25</sup> even though in the case of cell-based assays, aggregation tend to have the opposite effect and generally results in false negatives due to lower membrane diffusion.<sup>26</sup> Considering the fact that some dose-response curves had fairly high Hill coefficient, it was important to determine if slope steepness could be related to a compound aggregation phenomenon. NMR spectra of aggregating compounds are expected to exhibit a rapid loss of peak intensity in a  $T_2$  variation of Carr–Purcell–Meiboom–Gill sequence<sup>27</sup> ( $T_2$ -CPMG) NMR experiments due to their slower tumbling rate, while soluble compounds that behave as fast-tumbling single molecules show slow signal decay (the use of  $T_2$ -CPMG NMR to monitor aggregation will be addressed in an upcoming publication).

Consequently, NMR spectra of all indoles and indazoles analogues were analyzed in potassium phosphate buffer; NMR data acquisition for compounds in cell culture media is unfeasible. Figure 6 shows some examples of spectra for compounds **1**, **2** and **4** and  $T_2$ -CPMG NMR data shows no substantial loss of peak intensity as delays increase. A known small-molecule fragment aggregator is also included as a comparison. Therefore, none of these compounds exhibit any evidence of aggregating behavior in aqueous buffer conditions, which provides a certain degree of confidence that the effect observed is unlikely due to this phenomenon.

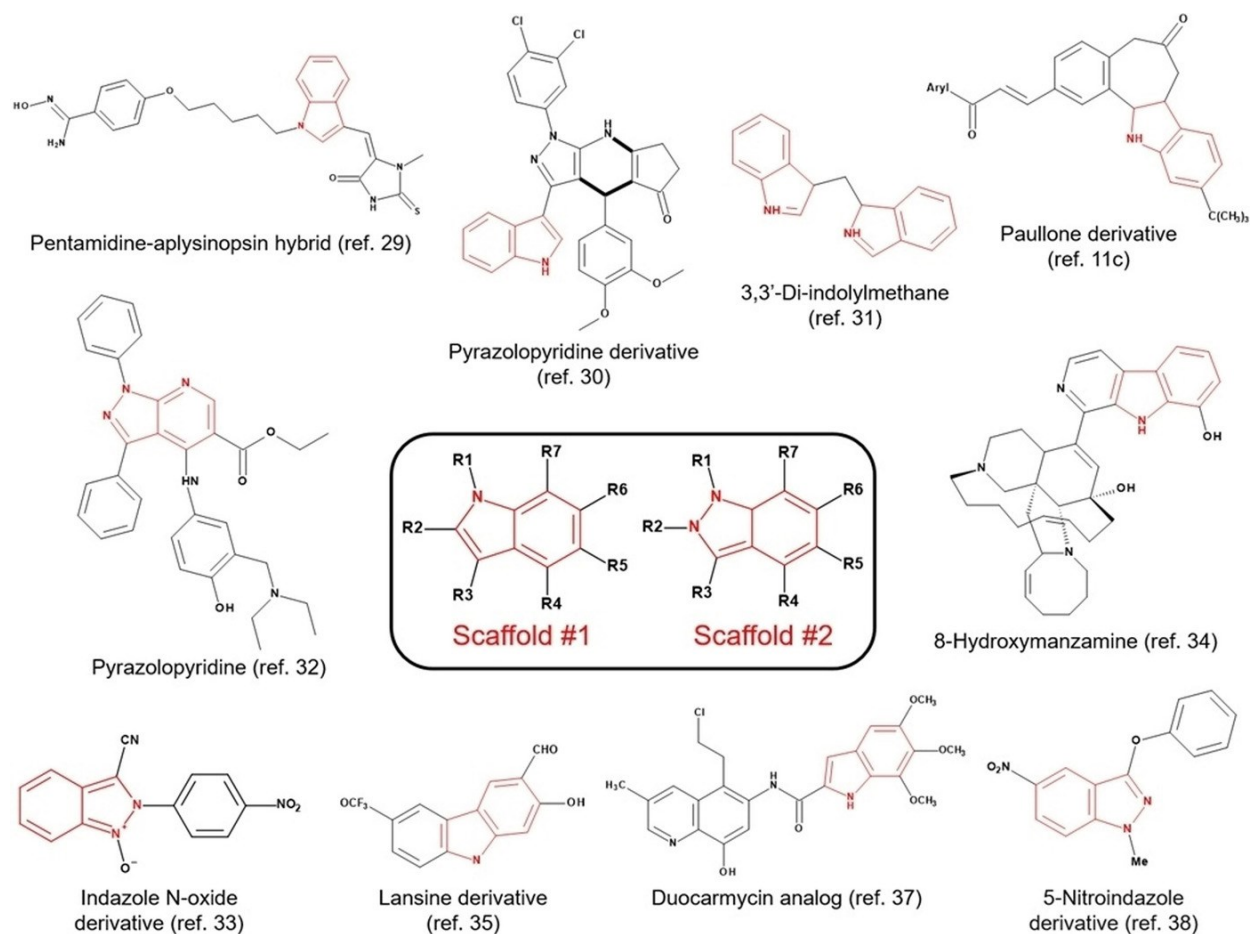


**Figure 6.** Representative  $^1\text{H}$  NMR  $T_2$ -CPMG spectra of compounds 1, 2 (active compounds) and 4 (inactive compound) compared with a known small-molecule aggregator control. Concentration of compounds was  $300\ \mu\text{M}$  in 20 mM potassium phosphate, pH 7, and 10 %  $\text{D}_2\text{O}$  for water suppression. Samples were run at 300 K in a 600 MHz magnet.

### Published compounds with indole/indazole-like scaffolds

Several drugs on the market are indoles and indazoles derivatives. Some of their very diverse applications include anti-inflammatory, antipsychotic, antiproliferative and anti-infectious effects. Indole derivatives are frequently found in plant species and are produced by various bacteria. Its occurrence is most likely due to its presence in the amino acid tryptophan and electrophilicity of indole nucleus is well amenable to synthetic manipulation by medicinal chemists. Indazoles on the other hand are less frequently found in nature, which may explain why they received more limited chemistry research efforts even though they both hold interesting biological properties.<sup>28</sup> The literature on leishmanicidal lead compounds reveals that such nitrogen-containing heterocycles hold promising potential as drug building blocks. Figure 7 shows some compounds that have been reported to be active against *Leishmania* parasite: A pentamidine-aplysinopsin hybrid was effective against *L. donovani* promastigotes with an  $\text{EC}_{50}$  of  $2\ \mu\text{M}$  and a selectivity index of 53.<sup>29</sup> A combination of low doses of miltefosine and pyrazolopyridine derivatives were reported to decrease liver and spleen parasite burden by more than 89 % in rodent model, with  $\text{EC}_{50}$  values of  $7\ \mu\text{M}$  and  $4\ \mu\text{M}$  against promastigotes and amastigotes, respectively.<sup>30</sup> Further immunological and molecular studies showed that their lead compound

exhibited a dual effect by acting as an immunostimulant and by directly killing the parasites.<sup>30</sup> Di-indolylmethane has been reported to be an inhibitor of *L. donovani* topoisomerase I with an EC<sub>50</sub> of 1.2 μm,<sup>31</sup> while paullone derivatives have shown EC<sub>50</sub> values <1 μm on axenic amastigotes, and high inhibition of *L. donovani* in infected macrophages at 5 μm.<sup>11c</sup> Pyrazolopyridine derivatives had EC<sub>50</sub> values <1 μm on *L. amazonensis* promastigotes<sup>32</sup> and indazole *N*-oxide derivative induced high parasite lysis against *L. amazonensis*, *L. infantum* and *L. braziliensis* promastigotes at 100 μg mL<sup>-1</sup>, possibly by inducing damages into parasite cells due to the nucleophilic properties of the compounds.<sup>33</sup> 8-Hydroxymanzamine had an EC<sub>50</sub> <3 mg mL<sup>-1</sup> for *L. donovani* promastigotes,<sup>34</sup> whereas a lansine derivative exhibited EC<sub>50</sub> values of 13 μm and 7 μm against *L. donovani* promastigotes and amastigotes, respectively,<sup>35</sup> possibly due to intercalation of the compound into DNA,<sup>36</sup> while having SIs of 21 and 4 for LU-1 and J-774 cells. A duocarmycin analogue had EC<sub>50</sub> values of 340 nm and 119 nm against *L. donovani* and *L. mexicana*, probably due a DNA sequence specificity of the compound.<sup>37</sup> C5-Nitroindazole derivative showed EC<sub>50</sub> values between 12 and 17 μm against *L. infantum* and *braziliensis* promastigotes as well as intracellular amastigotes, with SIs ranging between 10–15 and reports observed damage at the glycosomal and mitochondrial level.<sup>38</sup> Therefore, some of our compounds are already within activity ranges similar to these molecules. Moreover, many of these reports lack any explanation as to the mechanisms of action but one issue that emerges from the few that do have some idea about the processes involved is that such scaffold does not seem to be limited to a single mechanism. Indeed, it seems that one could create different specificities by applying various systematic chemical substitutions.



**Figure 7. Scaffold comparison of both fragment series with published leishmanicidal compounds.**

Therefore, it is very conceivable that the two series tested in this study could be optimized into effective leads due to their promising scaffold properties that can be amenable to various chemical modifications. One only has to look at the closely related benzimidazole scaffold that has been associated with several antiparasitic compounds such as albendazole and mebendazole which are already on the market for the treatment of various worm infections. The effect of some benzimidazoles have been attributed to inhibition of tubulin (particularly the  $\beta$  subunit) and therefore disruption of the microtubule structure and function in the parasite.<sup>39</sup> However, some derivatives showed no inhibition of tubulin polymerization<sup>40</sup> which again shows the potential of such scaffold to be designed with different mechanisms of action in mind. Another possible avenue would be to target the parasite proteasome activity. A previous report has even targeted kinetoplastid proteasome using azabenzoxazole derivatives resulting in low nanomolar range inhibition of parasite growth, while maintaining a high selectivity relative to mammalian cell lines.<sup>41</sup>

## Criteria for hits and leads against leishmaniasis

Previous report by Slingsby defines some guidelines convened by the GHIT regarding characteristics for hit and lead discovery in the case of infectious diseases of the developing world.<sup>22</sup> In regard to the choice of assay for primary screens, phenotypic screening involving intact pathogens like the method used in this work is a suggested approach as it has been more productive in the case of infectious diseases. This work satisfies several other general hit selection criteria listed in this report such as: most hits have Hill coefficients between 0.5 and 1.8, some basic SAR knowledge is available, because of the library curation, compounds have no highly reactive or unstable moieties in the pharmacophore and pass basic drug-like filters, the compounds are amenable to chemical modification and no major synthesis or formulation issues are anticipated, there are no known intellectual property conflicts and the most potent compound has a selectivity index higher than the suggested 100-fold (as there were no measurable negative effects at the highest dose of 500  $\mu\text{m}$ , it is very likely that other compounds would show >100 SI if they were to be tested at much higher concentration).

Some general criteria that would need to be fulfilled in the future include: obtaining good oral bioavailability and efficacy in animal models, determining the pharmacokinetic profiles and further early safety assessment. Other criteria specific to leishmaniasis that would need to be addressed include: 1) to obtain an  $\text{EC}_{50} < 10 \mu\text{m}$  against intracellular *L. donovani* (current efforts are in obtaining parasites expressing luciferase in order to test interspecies selectivity) since this work has focused on *L. amazonensis* following the primary screens, and 2) to obtain a >70 % decrease in liver parasite burden using a maximum of five oral doses of 50  $\text{mg kg}^{-1}$  once or twice a day in a rodent model.

## Conclusions

In this study, we screened *Leishmania* parasites with a fragment library and obtained hits from indole and indazole families with activities similar to sodium stibogluconate while having no observable cytotoxicity on macrophages at concentrations up to 500  $\mu\text{m}$ . Moreover, no evidence of compound aggregation was observed. This study therefore serves as a proof of concept demonstrating that FPLD is a rapid and practical strategy. Although this method does not allow target identification, metabolomics analysis of the cells treated with the compounds could address such aspect. This approach can enable institutions of all sizes and types to launch new drug discovery projects where FPLD can identify new chemical matter that can serve as new seeds to targets diseases that include neglected diseases. These compounds could now serve as leads for more advanced medicinal chemistry or as tool compounds to identify known or novel targets.



## Experimental Section

**Ethics statement:** Experiments involving mice were done as prescribed by protocols approved by the Comité Institutionnel de Protection des Animaux of the INRS—Institut Armand-Frappier (1302-03). These protocols respect procedures on good animal practices provided by the Canadian Council on Animal Care.

**Cell culture:** Bone marrow-derived macrophages (BMM) were differentiated from the bone marrow of 6- to 8-week-old female 129B6 (Charles River Laboratories) as described.<sup>20a</sup> Cells were cultured for seven days in complete medium (DMEM [Life Technologies] supplemented with l-glutamine [Life Technologies]), 10 % heat-inactivated FBS [PAA Laboratories], 10 mM HEPES at pH 7.4, and antibiotics) containing 15 % v/v L929 cell-conditioned medium as a source for M-CSF. Macrophages were kept at 37 °C in a humidified incubator with 5 % CO<sub>2</sub>. To render BMM quiescent prior to experiments, cells were transferred to 24- or 96-well tissue culture microplates (TrueLine) and kept for 16 h in complete DMEM without L929 cell-conditioned medium. Promastigotes of *L. amazonensis* LV79 and *L. donovani* LV9 were grown at 26 °C in *Leishmania* medium (Medium 199 supplemented with 10 % heat-inactivated FBS, 40 mM HEPES pH 7.4, 100 µM hypoxanthine, 5 µM hemin, 3 µM biopterin, 1 µM biotin, and antibiotics).

**Infections:** BMM were seeded in 24-well plates (1.6×10<sup>5</sup> per well) containing microscope coverslips (Fisher Scientific) and were incubated with serum-opsonized *L. amazonensis* LV79 promastigotes for 2 h at 34 °C in a humidified incubator with 5 % CO<sub>2</sub>. Non-internalized parasites were removed by several washes with PBS and were further incubated. Compounds were prepared from 8.3 mM stock solutions in [D<sub>6</sub>]DMSO for pool screens and from 100 mM stocks for subsequent experiments. Compounds were added 24 h post-infection and data acquisition was done at several time-points.

**Microscopy:** 96-well plates containing *Leishmania* promastigotes or BMM were photographed using the 15× objective of a Nikon Eclipse TE2000-U microscope. Coverslips were fixed with methanol and stained by Giemsa staining (Protocols). Fluoromount-G (Southern Biotechnology Associates) was used to mount coverslips on glass slides (Fisher), and coverslips were sealed with nail polish (Sally Hansen). Pictures were taken using the 40× objective of a Nikon Eclipse E800 microscope.

**Quantification of nitric oxide and TNF:** The amounts of TNF in cell supernatants were determined by enzyme-linked immunosorbent assays, as previously described.<sup>42</sup> The amounts of nitrite released in cell supernatants were determined with the Griess reagent as described.<sup>43</sup>

**MTT cytotoxicity assay:** Culture medium was replaced in each well by 100  $\mu$ L of fresh culture medium containing MTT (Thermo Fisher Scientific) prepared as per retailer's recommendations and incubated for 3 h at 37 °C. Half of the medium was then removed and 200  $\mu$ L of 1:1 DMSO and isopropanol solution was added to solubilize the formazan crystals. Cell viability was then determined by measuring optical density at 570 nm and subtracting the 650 nm background absorbance. Cell survival was reported as percentage of the highest concentration of vehicle control ([D<sub>6</sub>]DMSO), which was 2 % in primary screens and 0.5 % in compound titration experiments.

**NMR experiments:** The NMR curation method employed the acquisition of <sup>1</sup>H NMR data for each compound in aqueous buffer to allow the rapid evaluation of structural integrity, purity, solubility, stability, aggregation tendencies and chemical shift encoding positions. Each sample consisted of 300  $\mu$ m nominal compound concentration which was prepared from 100 mm stock solutions in [D<sub>6</sub>]DMSO and diluted in buffer (50 mm sodium phosphate pH 7.4, 100 mm NaCl) containing 10 % D<sub>2</sub>O. <sup>1</sup>H NMR spectra were acquired on a 600 MHz NMR spectrometer equipped with a helium cryoprobe which significantly increases signal to noise. Simple 1D <sup>1</sup>H NMR spectra were acquired along with 1D <sup>1</sup>H T<sub>2</sub>-CPMG for size-filtering aggregation analyses. Samples for the CPMG experiments were run in 50 mm potassium phosphate, 100 mm NaCl, 10 % D<sub>2</sub>O, pH 7.4, and were stored and measured at 300 K. CPMG experiments employed a sweep width (SW)=20.0263 ppm, recycle delay (D1)=3 s, number of points (TD)=32 000, and four scans for each spectrum. Water suppression was applied using excitation sculpting with gradients.<sup>44</sup> Data were visualized using Bruker's TopSpin software and peak alignment was done using DMSO peak set at 2.49 ppm.

**Statistical analysis:** Initial scoring system for primary screens consisted of visual examination under microscope and EC<sub>50</sub> values were obtained by nonlinear regression and curves were produced using GraphPad Prism 5.03 (GraphPad, San Diego). Infection indexes for infected BMM were obtained by calculating the average number of parasites per macrophages.<sup>21</sup> At least 100 macrophages were counted per triplicate coverslip.

## **Acknowledgements**

The authors thank M. Serrano-Wu, P. Coote, A. Lowerson (Key Organics) and P. McCarren for their help in preparing the fragment library. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada to S.R.L. (RGPIN-2016-06747), the Canadian Institutes of Health Research to A.D. (MOP-125990), and NMX Research and Solutions Inc. A.D. is the holder of the Canada Research Chair on the Biology of intracellular parasitism. Y.A. was supported in part by an Undergraduate Student Research Award from the Natural Sciences and Engineering Research Council of Canada and by a Graduate Student Scholarship from the Quebec Network for Research on Protein Function, Engineering, and Applications.

**Conflict of interest**

*The authors declare no conflict of interest.*

## 11 SUPPORTING INFORMATION - ARTICLE 5

---

### **Fragment-based Phenotypic Lead Discovery: Cell-based Assay to Target Leishmaniasis**

Yann Ayotte,<sup>1</sup> François Bilodeau,<sup>2</sup> Albert Descoteaux,<sup>1\*</sup> Steven R. LaPlante,<sup>1,2\*</sup>

<sup>1</sup> *INRS-Institut Armand-Frappier. 531, boulevard des Prairies, Laval, Québec, Canada, H7V 1B7*

<sup>2</sup> *NMX Research and Solutions Inc. 500 boulevard Cartier, Laval, Québec, Canada, H7V 5B7*

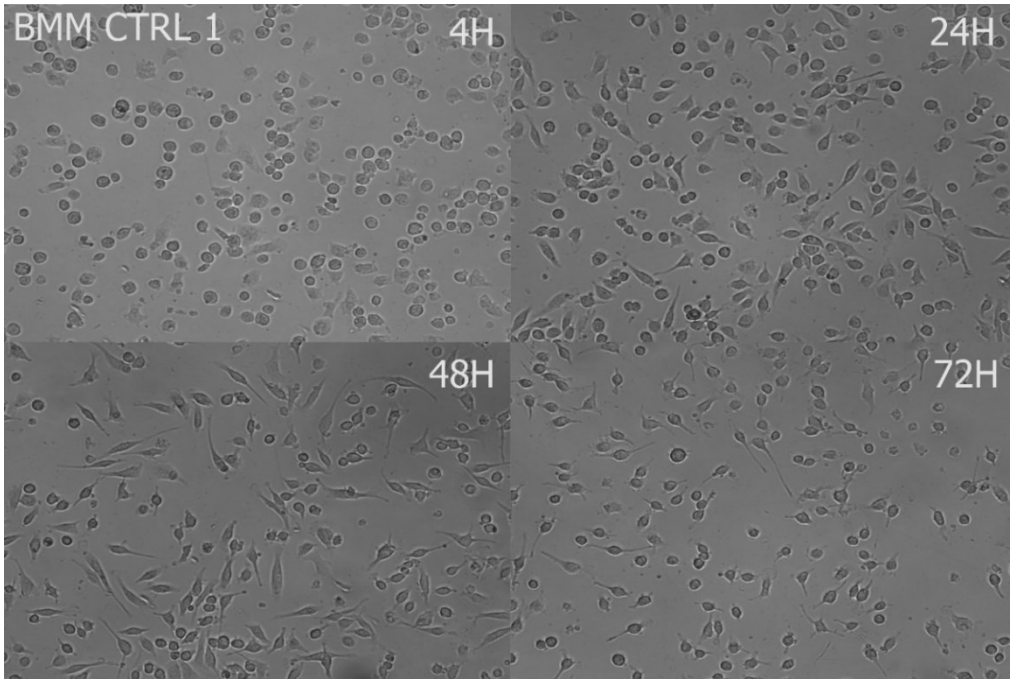
Corresponding Authors

\*Steven R. LaPlante: [steven.laplante@iaf.inrs.ca](mailto:steven.laplante@iaf.inrs.ca)

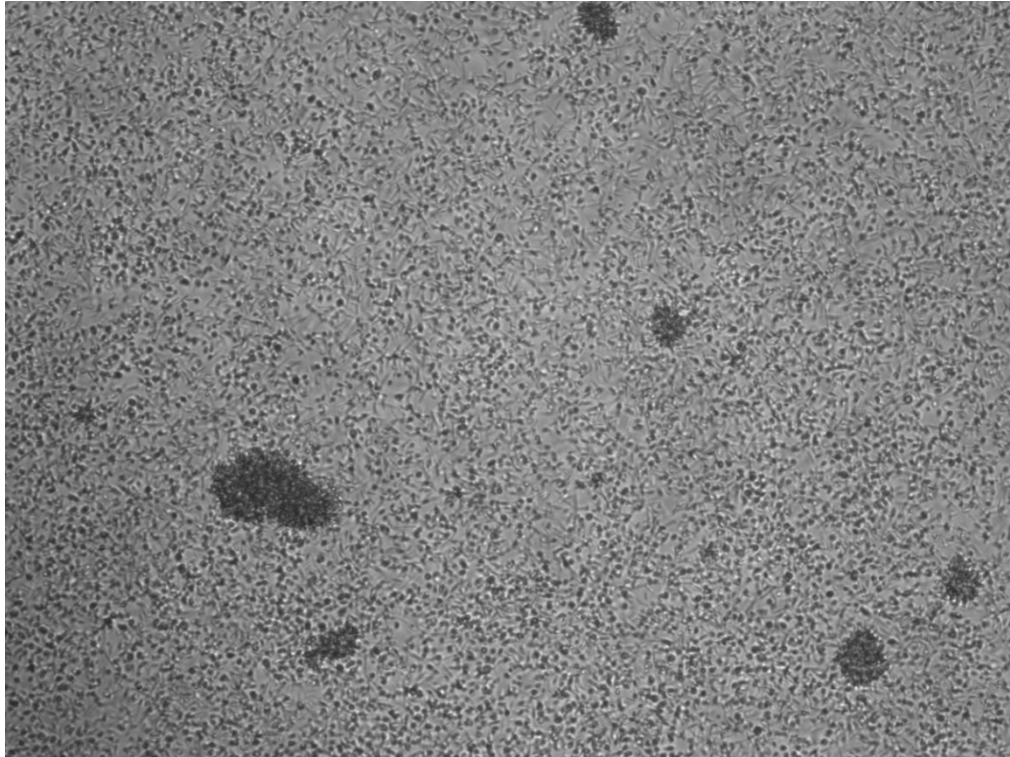
\*Albert Descoteaux: [albert.descoteaux@iaf.inrs.ca](mailto:albert.descoteaux@iaf.inrs.ca)

Published in : ChemMedChem 2018, 13(14), 1377-1386

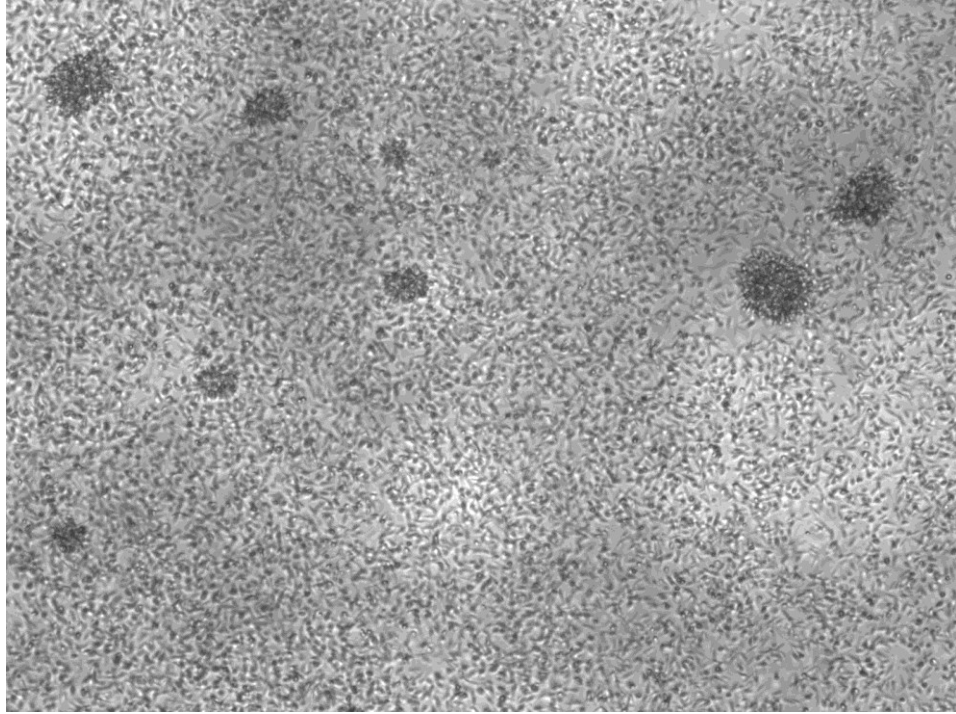
DOI : 10.1002/cmdc.201800161



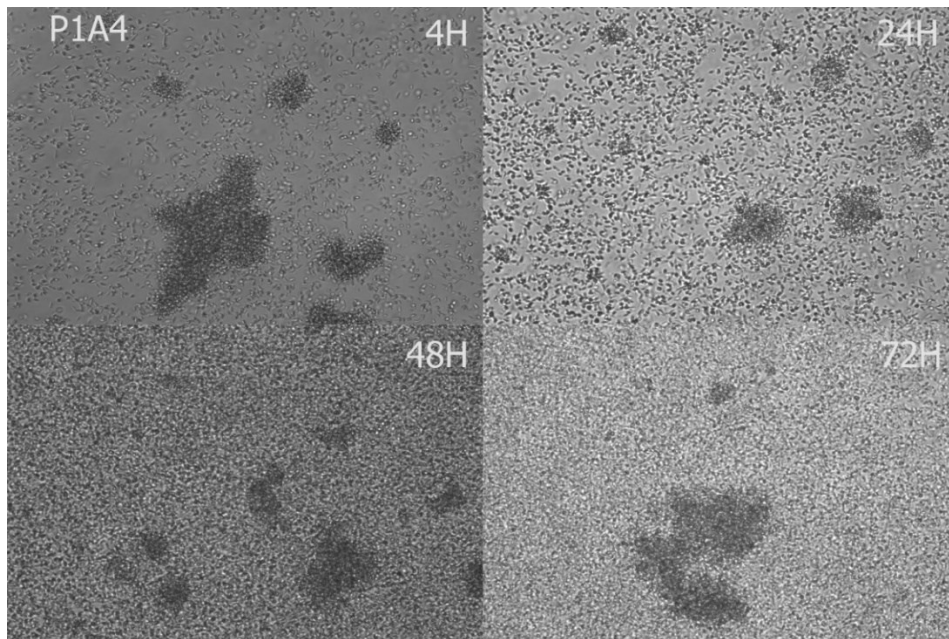
**Figure S1.** Representative bone marrow-derived macrophages controls with 2% DMSO.



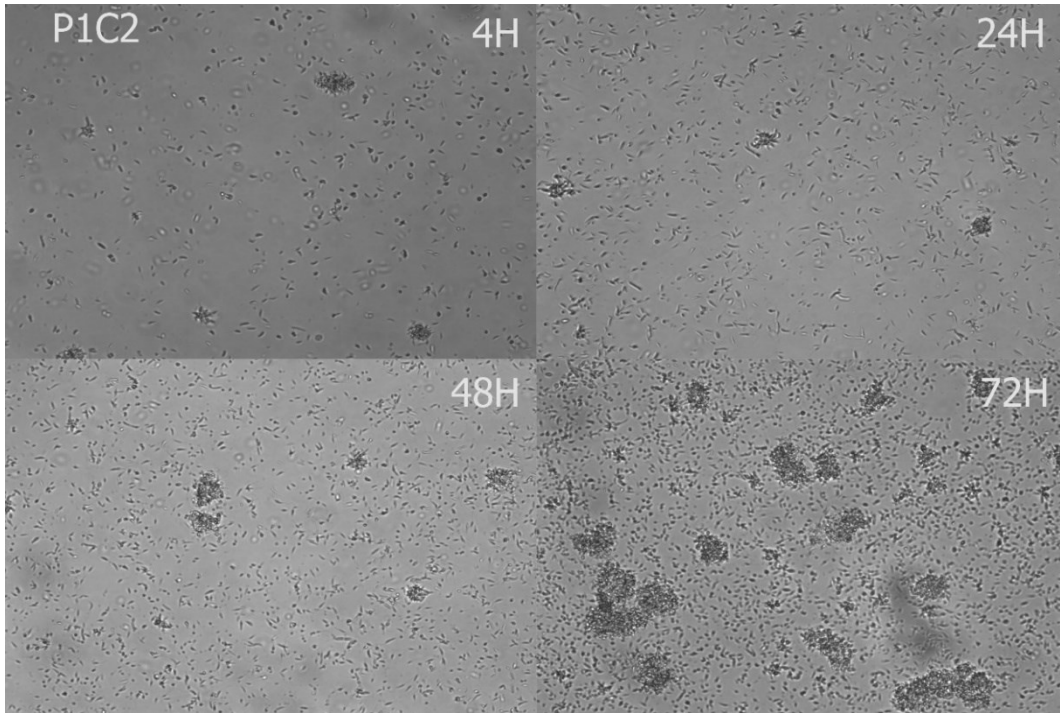
**Figure S2.** Representative *L. donovani* control with 2% DMSO after 72 h.



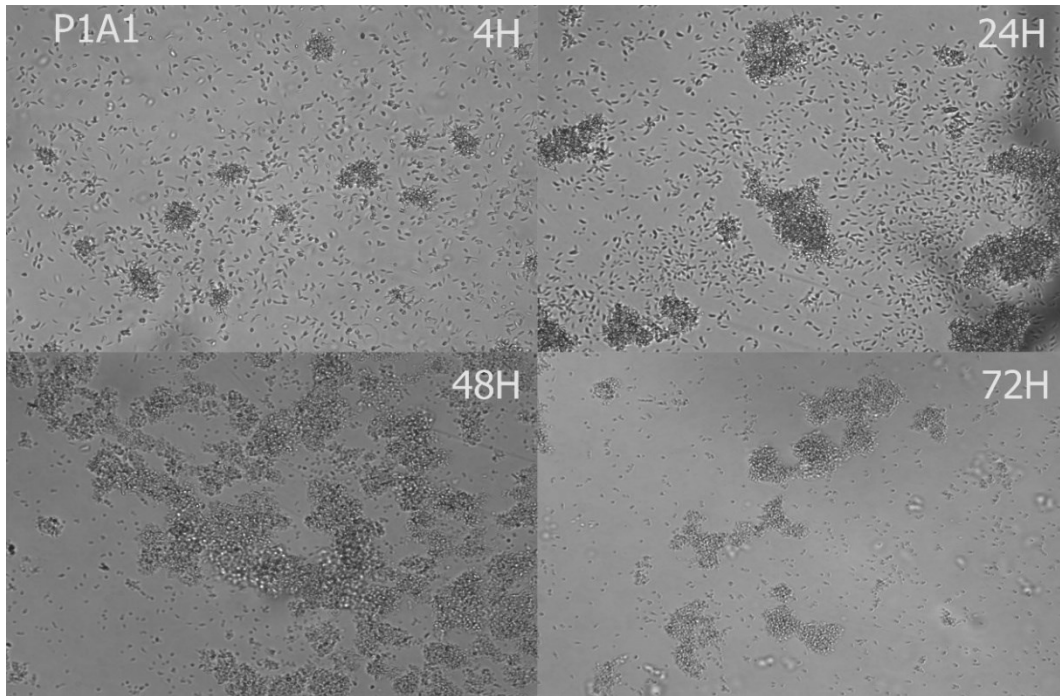
**Figure S3.** Representative *L. amazonensis* control with 2% DMSO after 72 h.



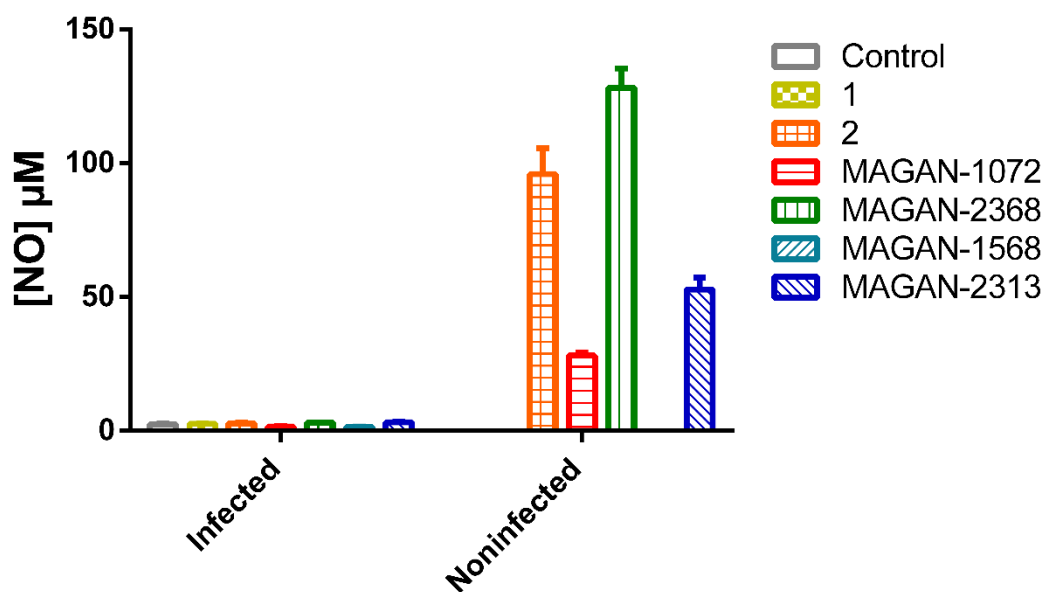
**Figure S4.** Example of an inactive pool against *L. amazonensis* at 166  $\mu\text{M}$  which would be attributed a value of 5 (similar to control, see excel files). Pool P1A4: plate 1, well A4.



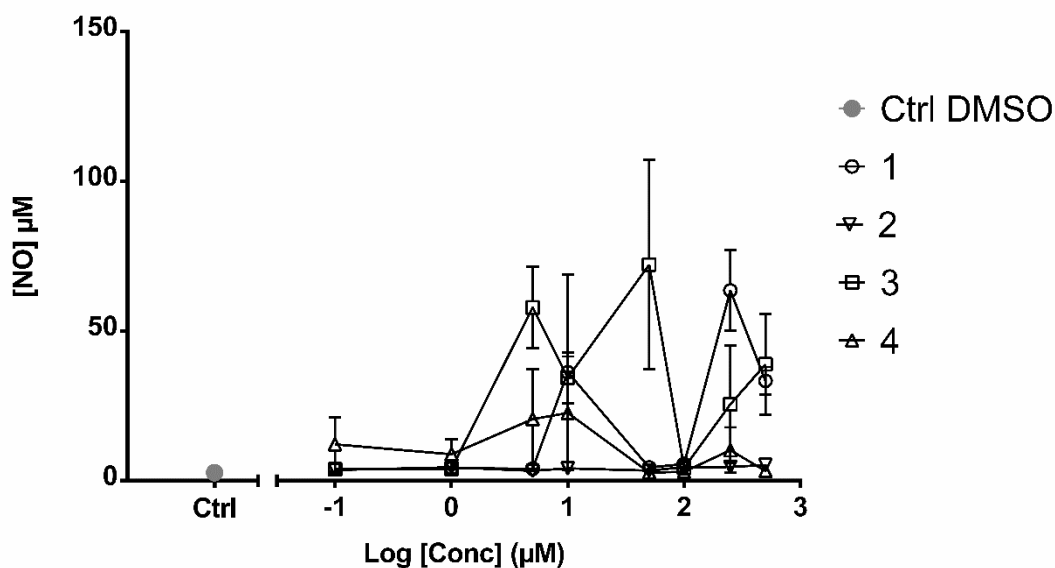
**Figure S5.** Example of a moderately active pool against *L. amazonensis* at 166  $\mu\text{M}$  which would be attributed a value of 3 (moderate health, see excel files). Pool P1C2: plate 1, well C2.



**Figure S6.** Example of highly effective pool against *L. amazonensis* at 166  $\mu\text{M}$  which would be attributed a value of 0 (dead, see excel files). Pool P1A1: plate 1, well A1.

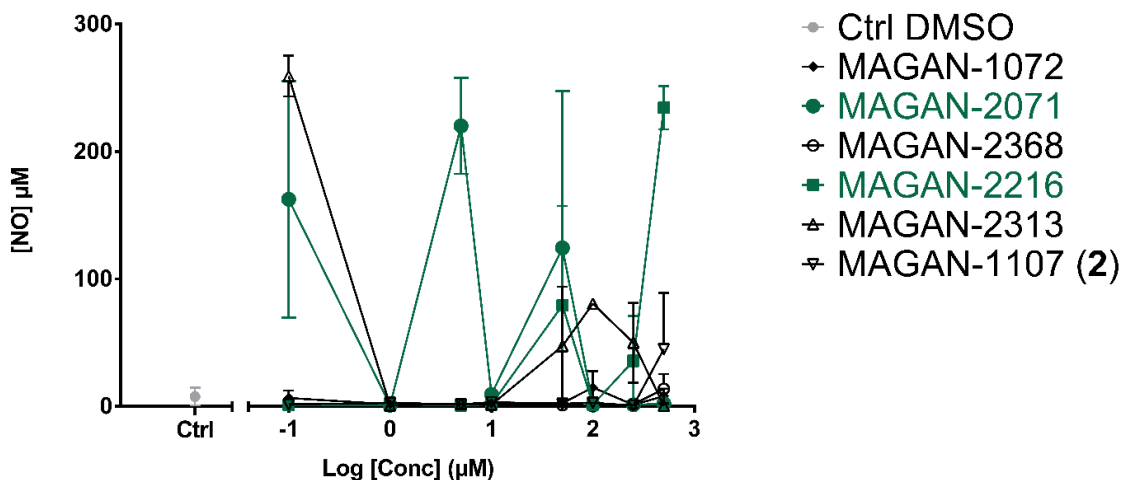


**Figure S7.** Nitric oxide secretion in infected and noninfected BMMs at compounds concentration of 166  $\mu\text{M}$ . Levels of NO are almost immeasurable in infected BMM media. Absence of bars for 1 and MAGAN-1568 are due to values of 0. Control condition represents 0.5% DMSO. Samples were prepared in triplicate.

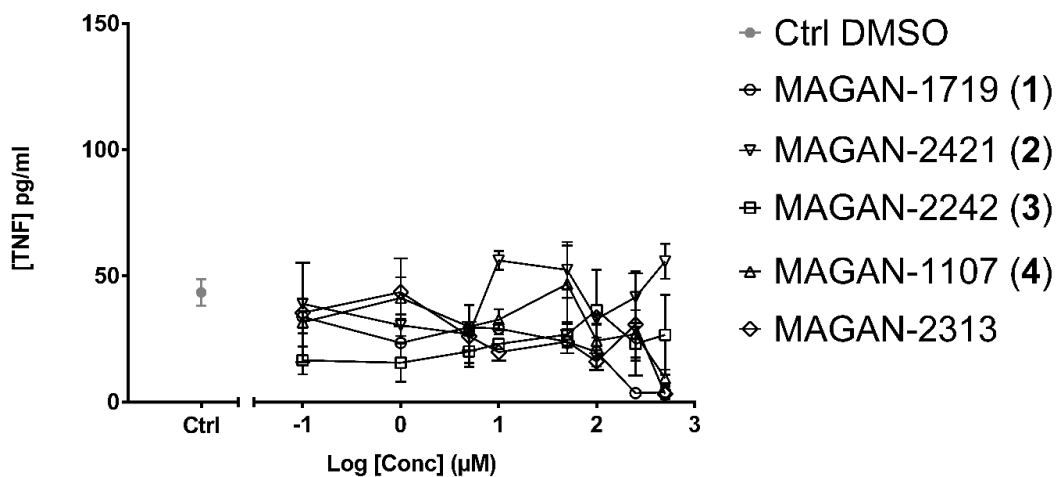


**Figure S8.** Nitric oxide secretion of the three hits in infected macrophages condition at multiple concentrations. Three compounds with similar molecular structures were added in as controls. No significant difference is observed between compounds or with vehicle control. NO secretion does not show any concentration-dependent effect. Control was the highest DMSO concentration used (0.5%). Samples were prepared in triplicate.



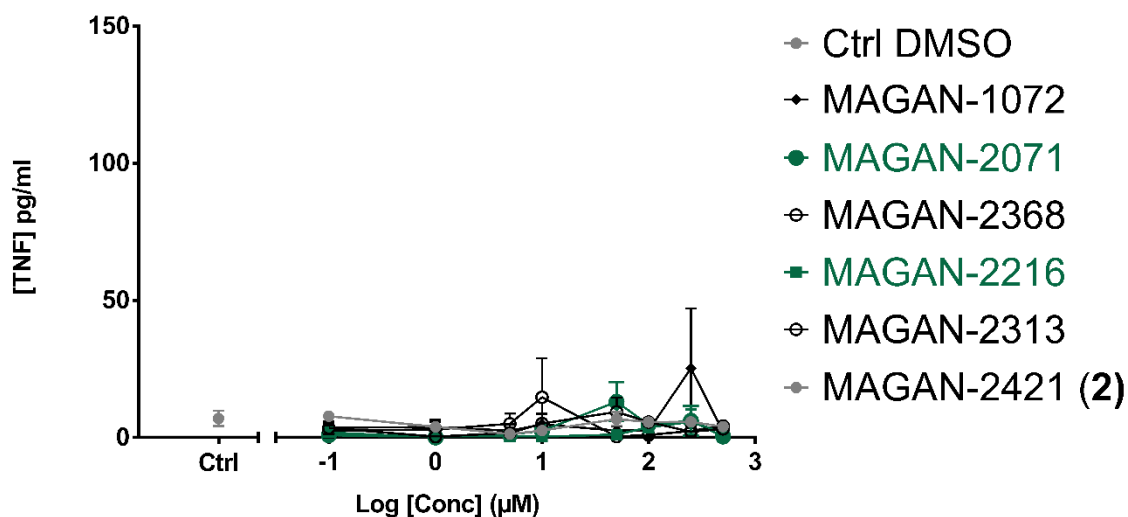


**Figure S9.** Nitric oxide secretion in infected macrophages of compounds that showed elevated inducement in primary screens. No significant difference is observed between most compounds and the control. Most effect variability is due to MAGAN-2071 and -2216 which were randomly selected singleton controls that did not show elevated NO secretion at 166 μM during primary screens. NO secretion does not show any concentration-dependent effect. Control was the highest DMSO concentration used (0.5%). Samples were prepared in triplicate.

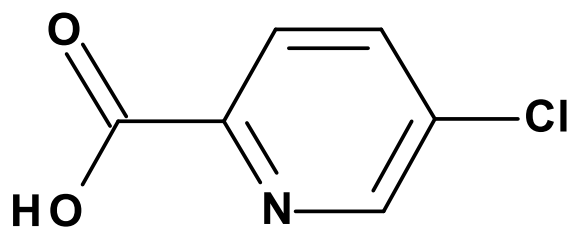


**Figure S10.** TNFα secretion of three initial hits and structurally similar inactive controls for *L. amazonensis* infected macrophages. No significant difference is observed between compounds

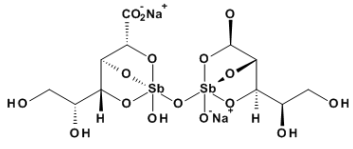
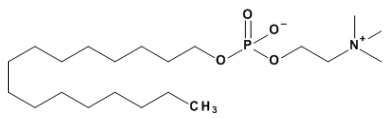
or with vehicle control. TNF secretion does not show any concentration-dependent effect. Vehicle control was the highest DMSO concentration used (0.5%). Samples were prepared in triplicate.



**Figure S11.** TNF $\alpha$  secretion of noninfected macrophages with compounds that showed elevated inducement in primary screens in noninfected macrophages. MAGAN-2071 and -2216 which were randomly selected singleton controls that did not show elevated NO secretion at 166  $\mu\text{M}$  during primary screens. No significant difference is observed between compounds or with vehicle control. Vehicle control was the highest DMSO concentration used (0.5%). Samples were prepared in triplicate.



**Figure S12.** Chemical structure of the known fragment aggregator (MAGAN-1376) in Fig. 6.

Compound	Structure	EC <sub>50</sub> (μM)
Sodium stibogluconate		13
Miltefosine		0.2

**Figure S13.** EC<sub>50</sub>s of reference drugs.

## 12 ARTICLE 6 – EXPANSION OF THE FRAGMENT-BASED PHENOTYPIC APPROACH TO OTHER DISEASE SYSTEMS

---

### Fragment-Based Phenotypic Lead Discovery To Identify New Drug Seeds That Target Infectious Diseases

Yann Ayotte<sup>1</sup>, Eve Bernet<sup>2</sup>, François Bilodeau<sup>3</sup>, Mena Cimino<sup>4</sup>, Dominic Gagnon<sup>5</sup>, Marthe Lebughe<sup>2</sup>, Maxime Mistretta<sup>4</sup>, Paul Ogadinma<sup>3</sup>, Sarah-Lisa Ouali<sup>1</sup>, Aïssatou Aïcha Sow<sup>1</sup>, Laurent Chatel-Chaix<sup>1</sup>, Albert Descoteaux<sup>1</sup>, Giulia Manina<sup>4</sup>, Dave Richard<sup>5</sup>, Frédéric Veyrier<sup>2</sup> & Steven R. LaPlante<sup>\*1</sup>

<sup>1</sup>Institut national de la recherche scientifique - Armand-Frappier Santé Biotechnologie Research Centre, 531 Boulevard des Prairies, Laval, Quebec, H7V 1B7, Canada

<sup>2</sup>Bacterial Symbionts Evolution, Institut national de la recherche scientifique, Armand-Frappier Santé Biotechnologie Research Centre, 531 Boulevard des Prairies, Laval, Quebec, H7V 1B7, Canada

<sup>3</sup>NMX Research and Solutions, Inc., 500 Boulevard Cartier Ouest, Laval, Quebec, H7V 5B7, Canada

<sup>4</sup>Microbial Individuality and Infection Group, Cell Biology and Infection Department, Institut Pasteur, 25-28 Rue du Docteur Roux 75015, Paris, France

<sup>5</sup>Centre de recherche du CHU de Québec-Université Laval, Département de Microbiologie-Infectiologie et d'Immunologie, Université Laval, Quebec, Quebec, G1V 0A6, Canada

Published in : ACS Chemical Biology 2021, 16(11), 2158-2163

DOI : 10.1021/acscchembio.1c00657

#### Authors contributions :

Yann Ayotte generated the figures, wrote the manuscript, and managed the various collaborative projects from the side of Steven Laplante's lab, under his supervision. Yann Ayotte performed the experiments for the *Leishmania* project under the supervision of François Bilodeau, Albert Descoteaux and Steven Laplante. Experiments for the *plasmodium* project were performed by Dominic Gagnon under supervision from Dave Richard. Experiments for the *Neisseria* project were performed by Eve Bernet and Marthe Lebughe under the supervision of Frédéric Veyrier. Chemical synthesis for the *Neisseria* project was performed by Sarah-Lisa Ouali under the supervision of Paul Ogadinma and François Bilodeau. Experiments for the *mycobacterium* project were performed by Maxime Mistretta and Mena Cimino under supervision from Giulia Manina.

Experiments for the flavivirus project were performed by Aïssatou Aïcha Sow under supervision from Laurent Chatel-Chaix.

## Abstract

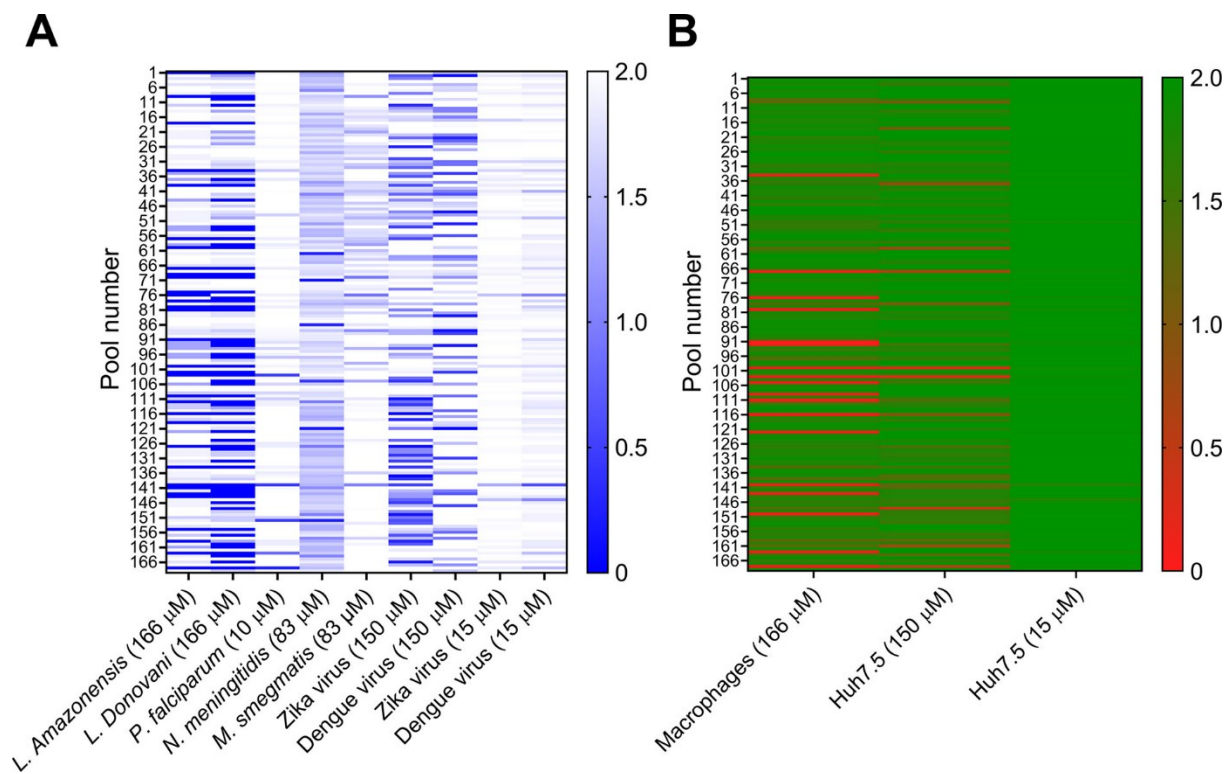
Fragment-based lead discovery has emerged over the last decades as one of the most powerful techniques for identifying starting chemical matter to target specific proteins or nucleic acids *in vitro*. However, the use of such low-molecular-weight fragment molecules in cell-based phenotypic assays has been historically avoided because of concerns that bioassays would be insufficiently sensitive to detect the limited potency expected for such small molecules and that the high concentrations required would likely implicate undesirable artifacts. Herein, we applied phenotype cell-based screens using a curated fragment library to identify inhibitors against a range of pathogens including *Leishmania*, *Plasmodium falciparum*, *Neisseria*, *Mycobacterium*, and flaviviruses. This proof-of-concept shows that fragment-based phenotypic lead discovery (FPLD) can serve as a promising complementary approach for tackling infectious diseases and other drug-discovery programs.

Fragment-based lead discovery has gained increased interest in the last decades as a tool to identify new chemical matter for disease targets. Because of their small size and low complexity, fragment compounds can be used to probe broad swaths of chemical space (fragment space) without the need for large chemical libraries—libraries of a hundred to a few thousand are typically sufficient to identify fragment binders.<sup>1,2</sup> However, it has long been assumed that the limited binding affinities of fragments would make them incompatible with cell-based screening.<sup>3</sup> From a more traditional screening viewpoint, the inherent weak affinities would be poorly detectable because of insufficient sensitivity of the assays, and the high concentrations required for measurements would certainly result in overwhelming artifacts.

On the other hand, compared with traditional high-throughput libraries, the small molecular size of fragments might also present advantages such as higher cell-membrane permeability.<sup>4</sup> Although small fragment-like compounds are certainly present in large library collections for high-throughput screens, examples of focused fragment screens involving cell-based assays are only lately emerging. Recent reports have used libraries of fragment-based chemical probes in combination with chemical proteomics to map interactions directly in human cells.<sup>5,6</sup> Although some hybrid approaches have been used to screen fragments in more biologically relevant conditions,<sup>7-9</sup> only a very few have screened fragments without any prior assumptions about the molecular mechanism of action. For example, FPLD was applied to screen *Mycobacterium tuberculosis* to identify compounds that exhibited favorable efficacy and

pharmacokinetic properties in mouse models.<sup>10</sup> In parallel, we successfully applied FPLD to inhibit *Leishmania* parasites<sup>11</sup> and other systems (*vide infra*).

Here, we report our findings from screens of a curated fragment library against a variety of infectious disease models that include parasites, bacteria, and viruses. Our approach was similar to the one that was successfully used to identify *Leishmania* parasite inhibitors.<sup>11</sup> To expedite the screening process of 1604 compounds, we first screened pools of compounds (169 pools containing 7–12 compounds per pool) and then deconvoluted those pools that were hits (referred to here as “hit pools”) via singleton assays. Figure 1A shows a comparison of the hit pools from various screening campaigns. One of our first observations was that, similarly to classical screens, optimal screening concentrations are system-dependent and might require optimization depending on the experimental setup. This system-dependent optimization may be explained by factors such as differences in membrane permeation, differential solubility of the compounds in the screening culture medium, or simply the intrinsic sensitivity of the system. For example, for replication screens against Zika and dengue viruses at 150  $\mu\text{M}$  compound concentration, it was obvious that the hit rates were relatively high, and some of the effects observed could be due to compound toxicity against the Huh7.5 host cells (Figure 1B). Reducing the screening concentration to 15  $\mu\text{M}$  resulted in a much more reasonable number of hits (Figure 1A) with minimal toxicity (Figure 1B). It is also noteworthy that although some pools showed activity/toxicity across several or even all conditions, most showed distinct activity profiles across the various systems.



**Figure 1.** Heatmap representations for the screens involving 169 pools screens (vertical axis). Results were normalized on a  $\log_{10}$  scale based on activity/toxicity where a value of 2 represents 100% viability or replication of either pathogens or mammalian cells and where a 0 represents complete cell death. Pool concentrations used for each system are indicated in parentheses. (A) Screens against multiple pathogens (horizontal axis). The most active pools are depicted in blue. (B) Toxicity counter-screen results against macrophages and Huh7.5 cells. The most cytotoxic pools are depicted in red.

### ***Leishmania***

We previously reported a proof-of-concept study where *Leishmania donovani* and *amazonensis* parasites were screened against the NMX <sup>1</sup>H Classic fragment library and yielded chemically similar hits.<sup>11</sup> Early medicinal chemistry explorations found tractable structure–activity relationships (SAR), where distinct single-atom changes resulted in a range of responses in activities. This, along with the observation that the compound series was nontoxic to bone-marrow-derived macrophage host cells, suggested that the hits and analogues were not promiscuous and likely hit a specific target.

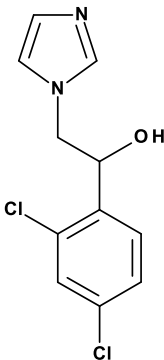
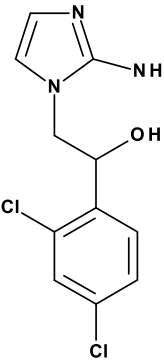
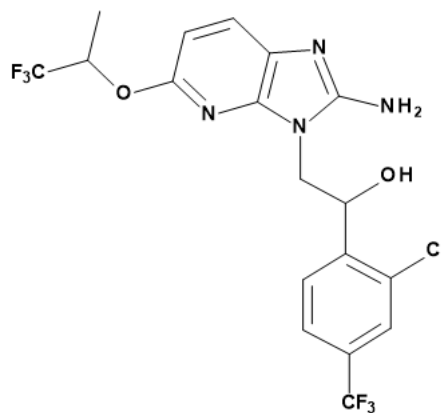
### ***Plasmodium falciparum***

The screen against *P. falciparum* resulted in five pools that showed >95% inhibition of parasite growth. Deconvolution of these hit pools to identify the active singletons led to five compounds



that exhibited more than 80% inhibition of growth and EC<sub>50</sub> activities in the low micromolar range. One of the best molecules (**1**) showed good selectivity indexes for 3D7 and Dd2 *P. falciparum* strains versus HEK293T and HeLa cell lines (Figure S1). During our study, we noted a publication from AstraZeneca where a high-throughput screen (HTS) identified a hit that was extremely close to our FPLD hit **1**. AstraZeneca discovered their hit (**2**) by screening their corporate library of ~500 000 compounds (Table 1). Interestingly, that study demonstrated that their hit could be optimized, giving compounds with increased potency in the low nanomolar range (compound **3**) that also exhibited good bioavailability and efficacy in the *P. falciparum* humanized mouse model.<sup>12</sup>

**Table 1.** Comparison of the best hit from our small fragment library and a previously reported hit from a large high-throughput screening library against *P. falciparum*.

Structure			
<b>Compound</b>	<b>1</b>	<b>2</b> (Ramachandran et al., 2014); Hit	<b>3</b> (Ramachandran et al., 2014); Optimized molecule
<b>Activity</b>	<i>Pf</i> IC <sub>50</sub> : 9.8 μM	<i>Pf</i> IC <sub>50</sub> : 0.38 μM	<i>Pf</i> IC <sub>50</sub> : 0.04 μM ED <sub>90</sub> : 28.6 mg/kg
<b>Library size</b>	1,604 (NMX 1H Classic library)	500,000 (AstraZeneca corporate library)	

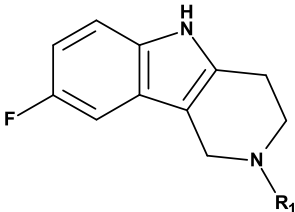
These findings are also consistent with the concept that screening small fragment libraries can be a fast and effective way to phenotype screen for starting chemical matter for a drug discovery program.<sup>13</sup> Our FPLD campaign screened a small library of only 1604 compounds, and we identified virtually the same hit as that discovered by the AstraZeneca HTS campaign that used ~500 000 compounds. Furthermore, the fact that they could optimize their hit to 40 nM shows that

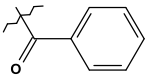
our hits were viable and exhibited specificity. Thus, this proof-of-concept suggests that phenotypic screens employing fragment libraries can identify real hits that can be followed up to achieve drug-like potency; thus, it can represent a convenient alternative approach for identifying new chemical matter for jump-starting drug discovery projects.

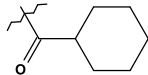
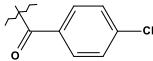
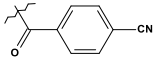
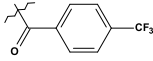
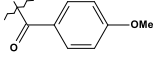
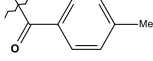
### ***Neisseria***

*Neisseria meningitidis* was also screened with the same pools of fragments (Figures S2 and S3), from which deconvolution efforts found that the most active hit was **4** (expected structure of **4a**) (Figure S4). Furthermore, counter-screening of all the hits against *E. coli* showed that compound **4** did not exhibit any significant effect on cell growth after 24 h (results not shown), suggesting a favorable degree of specificity of the effects observed. However, we were surprised to find that a freshly made stock solution of **4** showed a loss in activity beyond the detection limit of the assay ( $EC_{50} > 300 \mu\text{M}$ , Figure S4 and Table 2). Thus, it was clear that hit **4** no longer had the same structure as compound **4a**. One hypothesis proposed possible sample degradation such that **4a** slowly oxidized in solution to form compound **4b** (Figure S4). Oxidation of **4a** was performed experimentally, and the resulting product **4b** did indeed exhibit similar bactericidal activity as hit **4**. Surprisingly, despite the fact that core **4a** exhibited poor activity ( $EC_{50} > 300 \mu\text{M}$ ), addition of substituents at the R1 position (Table 2 and Figure S5) led to a number of active compounds **5–11**, which show distinct improvements in activity. The demonstration of structure–activity relationships (SAR) for this series of compounds is consistent with a specific interaction and binding mode. This series therefore warrants further optimization efforts.

**Table 2.**  $EC_{50}$  data for tetrahydro- $\gamma$ -carboline compounds against *N. meningitidis*.



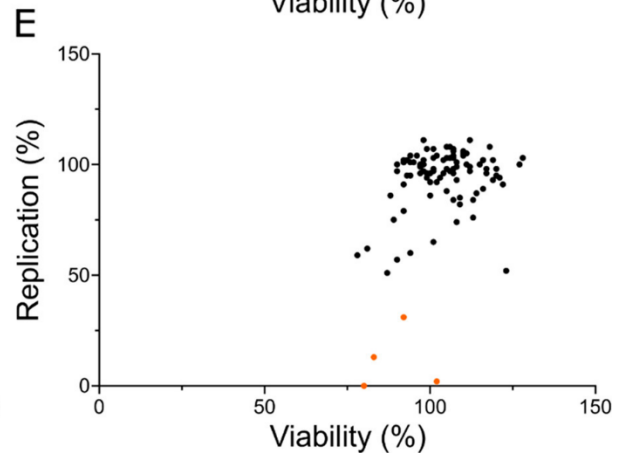
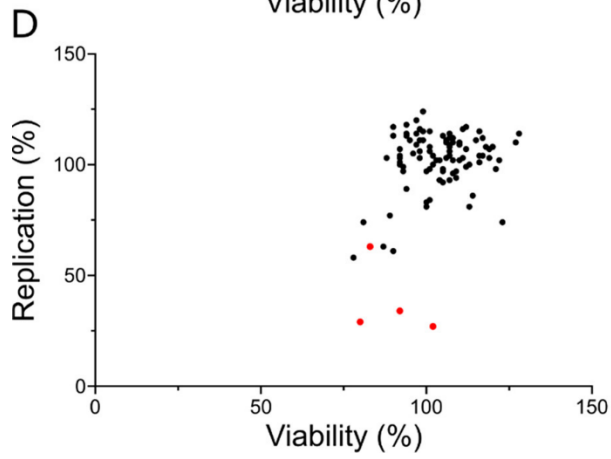
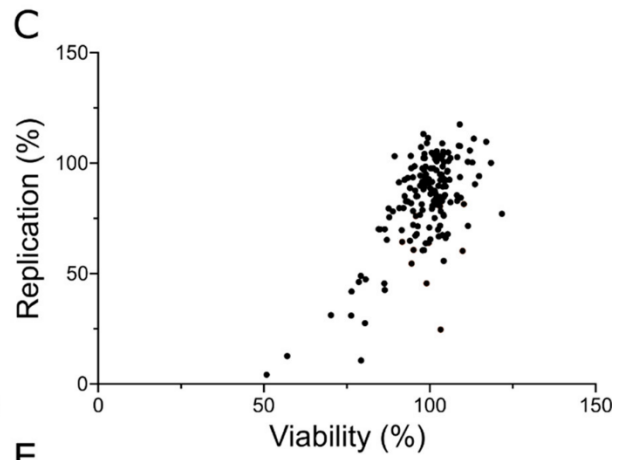
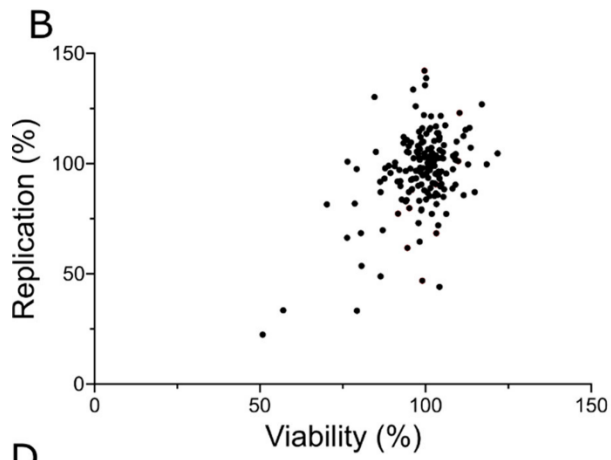
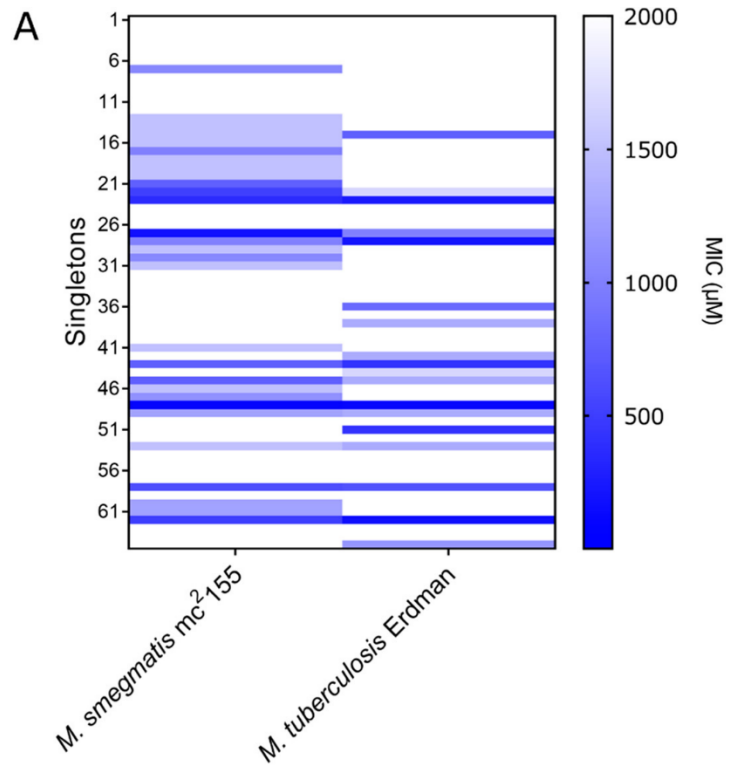
Compound	R <sub>1</sub>	EC <sub>50</sub> (μM)
<b>4a</b>	H	> 300
<b>5</b>		6.5

6		4.3
7		1.7
8		26.3
9		0.8
10		54.1
11		4.2

---

### ***Mycobacterium***

A phenotypic screen was also run against *M. smegmatis* using the same fragment pools. The six best pools also proved to be active against the highly related *M. tuberculosis* (>71  $\mu\text{M}$ ). Deconvolution of the best pools resulted in six individual fragments with micromolar minimum inhibitory concentrations (MICs) values  $125 < \text{MIC} < 417 \mu\text{M}$  against *M. tuberculosis* (Figure 2A). In these cases, no “stand out” hits were observed, which demonstrates, just as with conventional HTS screens, that not all fragment screens will identify outstanding hits. Interestingly, however, the four most effective compounds are able to inhibit both the fast-growing model organism and the slow-growing pathogenic strain. The weak hits still serve as starting points and are currently under investigation to identify more active analogues.



**Figure 2.** (A) Singleton deconvolution of the most interesting pools against *M. smegmatis* and *M. tuberculosis* results in several molecules with minimum inhibitory concentration (MIC) in the micromolar range. Some molecules are active against both *Mycobacterium* strains. (B–C) Results of the screens of fragment pools against Zika (B) and dengue (C) viruses at 15  $\mu\text{M}$  concentration (per compound). Singleton compounds of 10 pools were then tested at 50  $\mu\text{M}$  against Zika (D) and dengue (E) viruses. *X* axes represent the viability of Huh7.5 cells normalized to the DMSO vehicle control. *Y* axes represent the viral replication normalized to the DMSO vehicle.

## Flaviviruses

Viral replication screens were performed with Huh7.5 liver cells infected with Renilla luciferase-expressing reporter Zika virus (ZIKV) or dengue virus (DENV) using the same fragment library (Figure 2B–C).<sup>14,15</sup> Pools that resulted in a reduction of replication in the assay are shown in the heatmap of Figure 1A, which reveals some overlap of hits between both species—about 17 out of 25 active pools at 15  $\mu\text{M}$  are hits in both species. The panflaviviral activity of these pools is not surprising since DENV and ZIKV are closely genetically related and share many molecular mechanisms in their life cycle. However, eight pools were found to have effects only on DENV, suggesting that they might affect distinct targets reflecting their differences in their respective pathogenesis (e.g., neurovirulent vs hemorrhagic). Ten pools were selected for deconvolution secondary screening against the individual fragments. It was noteworthy that treatment with selected compounds conferred minimal toxicity in Huh7.5 cells and had a wide range of activity on the viruses. Four of the most interesting compounds (highlighted in red and orange in Figure 2D,E) are currently under investigation to identify more active analogues and investigate their mechanism of action.

In conclusion, we phenotypically screened a variety of microbial targets using 169 pools consisting of 1604 small-molecule fragment compounds. Rather than observe a flood of nonspecific inhibition and toxicity, as we anticipated could potentially occur, we found that these phenotypic assays resulted in limited (and hence manageable) numbers of hits. However, we noted that the hit rate needed to be “tuned” or lowered for some assays by reducing the compound concentrations. Moreover, the observed hits were for the most part distinct from one screen to the other, which was consistent with inhibition via a specific target. Also, counter screens showed no significant toxicity. Most importantly, follow up medicinal chemistry on some confirmed hits resulted in the observation of SAR and potency improvements. Thus, the proof-of-concept demonstrated here suggests that phenotypic screens employing fragment libraries can be a powerful way for identifying new chemical matter for jump-starting drug discovery projects.

One can envision some distinct opportunities when employing FPLD. The smaller size of fragments may inherently enable them to more easily cross cell membranes and inhibit new targets, rather than being largely limited to cell surface targets as with large drug-like compounds. Given that fragments are amenable for phenotypic screens, they could be practical for identifying new and essential targets without having to undergo exhaustive target characterization at the atomic level. Furthermore, this FPLD approach is not limited to microbes but can also prove to be relevant for a wide range of diseases such as cancer, neurological diseases, and cardiopathy drug discovery, which often lack well-defined clinical targets.<sup>16-24</sup> Finally, these and other pragmatic advantages promise to empower fragment-based screening efforts outside of pharma, such as in academia and smaller biotech companies.

### **Associated content**

The Supporting Information is available free of charge via the Internet.

Supplementary figures and experimental details

### **Methods**

Details of experimental materials and methods are provided in the Supporting Information.

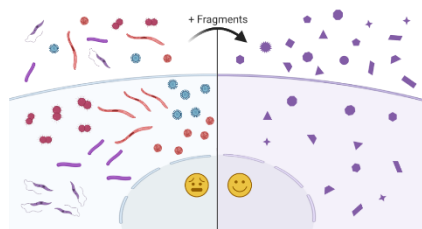
### **Acknowledgments**

We thank P. Shi and the World Reference Center for Emerging Viruses and Arboviruses (WRCEVA) for providing the ZIKV reporter system, and R. Bartenschlager (University of Heidelberg) for the DENV reporter system. We also thank PROTEO, NSERC, CQDM, MEI, MITACS, Institut Pasteur, INRS and NMX for financial support for this research. Table of content figure was created with BioRender.com.

### **Abbreviations**

FPLD, Fragment-based phenotypic lead discovery; FDA, Food and Drug Administration; DMSO, Dimethyl sulfoxide; SAR, Structure-activity relationship; MIC, Minimum inhibitory concentration; IC50, half maximal inhibitory concentration; CC50, half maximal cytotoxic concentration; *Pf*, *Plasmodium falciparum*; ED90, 90% effective dose; ZIKV, Zika virus; DENV, dengue virus

## TOC Figure



Depicted here are various microbes that get inhibited in presence of fragment molecules.

## References

- [1] Hajduk, P. J., and Greer, J. (2007) A decade of fragment-based drug design: strategic advances and lessons learned, *Nature Reviews Drug Discovery* 6, 211.
- [2] Rees, D. C., Congreve, M., Murray, C. W., and Carr, R. (2004) Fragment-based lead discovery, *Nature Reviews Drug Discovery* 3, 660-672.
- [3] Scott, D. E., Coyne, A. G., Hudson, S. A., and Abell, C. (2012) Fragment-Based Approaches in Drug Discovery and Chemical Biology, *Biochemistry* 51, 4990-5003.
- [4] Moreira, W., Lim, J. J., Yeo, S. Y., Ramanujulu, P. M., Dymock, B. W., and Dick, T. (2016) Fragment-Based Whole Cell Screen Delivers Hits against *M. tuberculosis* and Non-tuberculous Mycobacteria, *Front Microbiol* 7, 1392.
- [5] Cisar, J. S., and Cravatt, B. F. (2012) Fully Functionalized Small-Molecule Probes for Integrated Phenotypic Screening and Target Identification, *J. Am. Chem. Soc.* 134, 10385-10388.
- [6] Parker, C. G., Galmozzi, A., Wang, Y., Correia, B. E., Sasaki, K., Joslyn, C. M., Kim, A. S., Cavallaro, C. L., Lawrence, R. M., Johnson, et al. (2017) Ligand and Target Discovery by Fragment-Based Screening in Human Cells, *Cell* 168, 527-541.e529.
- [7] Schulze, J., Baukman, H., Wawrzinek, R., Fuchsberger, F. F., Specker, E., Aretz, J., Nazaré, M., and Rademacher, C. (2018) CellFy: A Cell-Based Fragment Screen against C-Type Lectins, *ACS Chem. Biol.* 13, 3229-3235.

- [8] Ruepp, M.-D., Brozik, J. A., de Esch, I. J. P., Farndale, R. W., Murrell-Lagnado, R. D., and Thompson, A. J. (2015) A fluorescent approach for identifying P2X1 ligands, *Neuropharmacology* 98, 13-21.
- [9] Stoddart, Leigh A., Vernall, Andrea J., Denman, Jessica L., Briddon, Stephen J., Kellam, B., and Hill, Stephen J. (2012) Fragment Screening at Adenosine-A3 Receptors in Living Cells Using a Fluorescence-Based Binding Assay, *Chem. Biol.* 19, 1105-1115.
- [10] Negatu, D. A., Liu, J. J. J., Zimmerman, M., Kaya, F., Dartois, V., Aldrich, C. C., Gengenbacher, M., and Dick, T. (2018) Whole-Cell Screen of Fragment Library Identifies Gut Microbiota Metabolite Indole Propionic Acid as Antitubercular, *Antimicrob. Agents Chemother.* 62.
- [11] Ayotte, Y., Bilodeau, F., Descoteaux, A., and LaPlante, S. R. (2018) Fragment-Based Phenotypic Lead Discovery: Cell-Based Assay to Target Leishmaniasis, *Chemmedchem* 13, 1377-1386.
- [12] Hameed P, S., Chinnapattu, M., Shanbag, G., Manjrekar, P., Koushik, K., Raichurkar, A., Patil, V., Jatheendranath, S., Rudrapatna, S. S., Barde, et al. (2014) Aminoazabenzimidazoles, a Novel Class of Orally Active Antimalarial Agents, *J. Med. Chem.* 57, 5702-5713.
- [13] Eggert, U. S. (2013) The why and how of phenotypic small-molecule screens, *Nat. Chem. Biol.* 9, 206.
- [14] Shan, C., Xie, X., Muruato, A. E., Rossi, S. L., Roundy, C. M., Azar, S. R., Yang, Y., Tesh, R. B., Bourne, N., Barrett, A. D., et al. (2016) An Infectious cDNA Clone of Zika Virus to Study Viral Virulence, Mosquito Transmission, and Antiviral Inhibitors, *Cell Host Microbe* 19, 891-900.
- [15] Chatel-Chaix, L., Fischl, W., Scaturro, P., Cortese, M., Kallis, S., Bartenschlager, M., Fischer, B., and Bartenschlager, R. (2015) A Combined Genetic-Proteomic Approach Identifies Residues within Dengue Virus NS4B Critical for Interaction with NS3 and Viral Replication, *J Virol* 89, 7170-7186.
- [16] Riggs, J. R., Nagy, M., Elsner, J., Erdman, P., Cashion, D., Robinson, D., Harris, R., Huang, D., Tehrani, L., Deyanat-Yazdi, G., et al. (2017) The Discovery of a Dual TTK Protein Kinase/CDC2-Like Kinase (CLK2) Inhibitor for the Treatment of Triple Negative Breast Cancer Initiated from a Phenotypic Screen, *J. Med. Chem.* 60, 8989-9002.
- [17] Drawnel, F. M., Boccardo, S., Prummer, M., Delobel, F., Graff, A., Weber, M., Gerard, R., Badi, L., Kam-Thong, T., Bu, L., et al. (2014) Disease modeling and phenotypic drug screening for diabetic cardiomyopathy using human induced pluripotent stem cells, *Cell Rep* 9, 810-821.



- [18] Moffat, J. G., Rudolph, J., and Bailey, D. (2014) Phenotypic screening in cancer drug discovery — past, present and future, *Nature Reviews Drug Discovery* 13, 588.
- [19] Xin, X., Wu, Y., Zang, R., and Yang, S.-T. (2019) A fluorescent 3D cell culture assay for high throughput screening of cancer drugs down-regulating survivin, *J. Biotechnol.* 289, 80-87.
- [20] Zheng, W., Thorne, N., and McKew, J. C. (2013) Phenotypic screens as a renewed approach for drug discovery, *Drug Discov Today* 18, 1067-1073.
- [21] Aggarwal, B. B., Sethi, G., Baladandayuthapani, V., Krishnan, S., and Shishodia, S. (2007) Targeting cell signaling pathways for drug discovery: An old lock needs a new key, *J. Cell. Biochem.* 102, 580-592.
- [22] Tardiff, D. F., and Lindquist, S. (2013) Phenotypic screens for compounds that target the cellular pathologies underlying Parkinson's disease, *Drug Discov Today Technol* 10, e121-128.
- [23] Zhang, M., Luo, G., Zhou, Y., Wang, S., and Zhong, Z. (2014) Phenotypic screens targeting neurodegenerative diseases, *J. Biomol. Screen.* 19, 1-16.
- [24] Schneider, M., Schüler, J., Höfflin, R., Korzeniewski, N., Grüllich, C., Roth, W., Teber, D., Hadaschik, B., Pahernik, S., Hohenfellner, M., et al. (2014) Phenotypic drug screening and target validation for improved personalized therapy reveal the complexity of phenotype-genotype correlations in clear cell renal cell carcinoma, *Urol Oncol.* 32, 877-884.

## 13 SUPPORTING INFORMATION - ARTICLE 6

---

### Fragment-Based Phenotypic Lead Discovery To Identify New Drug Seeds That Target Infectious Diseases

Yann Ayotte<sup>1</sup>, Eve Bernet<sup>2</sup>, François Bilodeau<sup>3</sup>, Mena Cimino<sup>4</sup>, Dominic Gagnon<sup>5</sup>, Marthe Lebughe<sup>2</sup>, Maxime Mistretta<sup>4</sup>, Paul Ogadinma<sup>3</sup>, Sarah-Lisa Ouali<sup>1</sup>, Aïssatou Aïcha Sow<sup>1</sup>, Laurent Chatel-Chaix<sup>1</sup>, Albert Descoteaux<sup>1</sup>, Giulia Manina<sup>4</sup>, Dave Richard<sup>5</sup>, Frédéric Veyrier<sup>2</sup> & Steven R. LaPlante<sup>\*1</sup>

<sup>1</sup>Institut national de la recherche scientifique - Armand-Frappier Santé Biotechnologie Research Centre, 531 Boulevard des Prairies, Laval, Quebec, H7V 1B7, Canada

<sup>2</sup>Bacterial Symbionts Evolution, Institut national de la recherche scientifique, Armand-Frappier Santé Biotechnologie Research Centre, 531 Boulevard des Prairies, Laval, Quebec, H7V 1B7, Canada

<sup>3</sup>NMX Research and Solutions, Inc., 500 Boulevard Cartier Ouest, Laval, Quebec, H7V 5B7, Canada

<sup>4</sup>Microbial Individuality and Infection Group, Cell Biology and Infection Department, Institut Pasteur, 25-28 Rue du Docteur Roux 75015, Paris, France

<sup>5</sup>Centre de recherche du CHU de Québec-Université Laval, Département de Microbiologie-Infectiologie et d'Immunologie, Université Laval, Quebec, Quebec, G1V 0A6, Canada

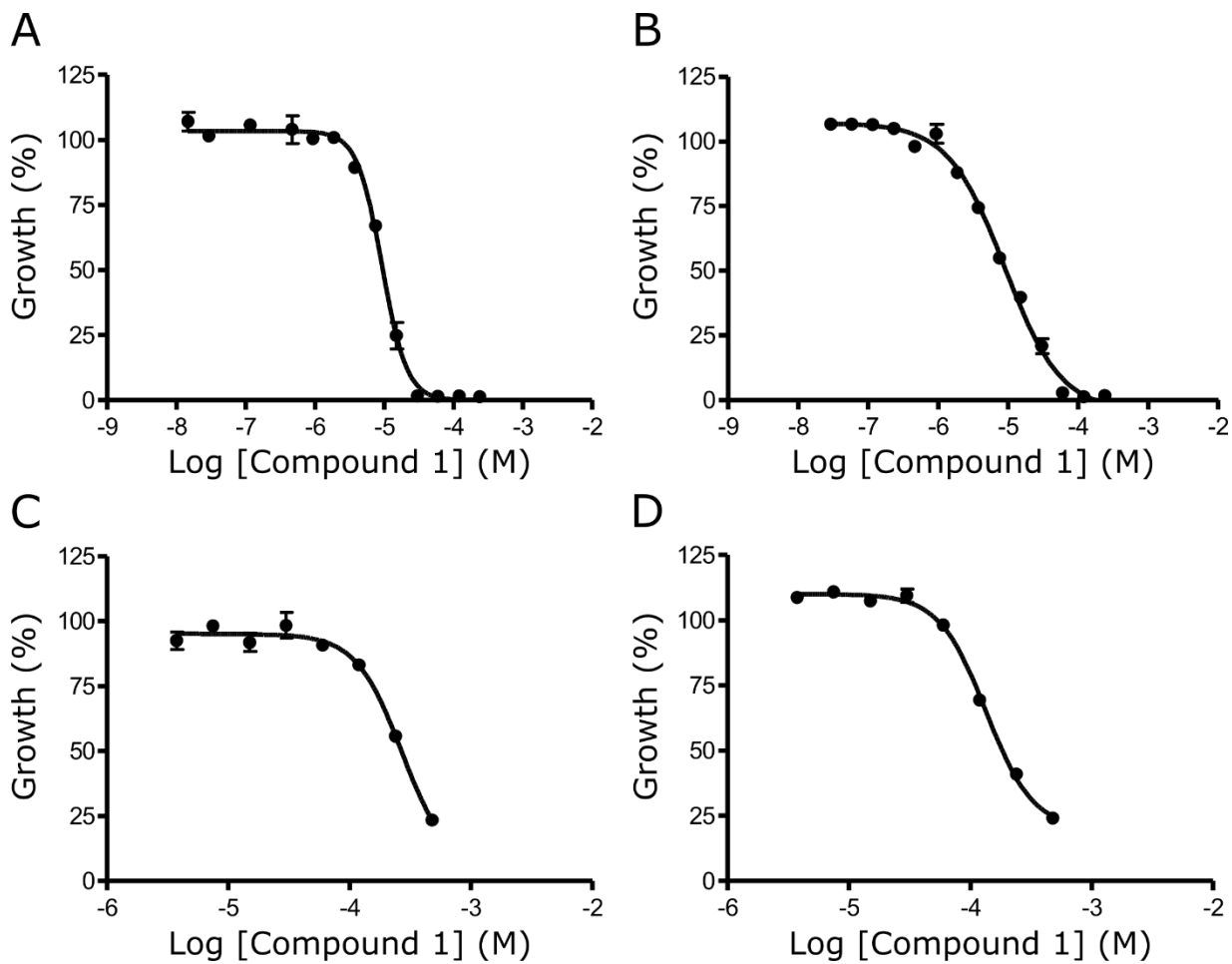
\*Corresponding author email: [steven.laplante@inrs.ca](mailto:steven.laplante@inrs.ca)

Published in : ACS Chemical Biology 2021, 16(11), 2158-2163

DOI : 10.1021/acscchembio.1c00657

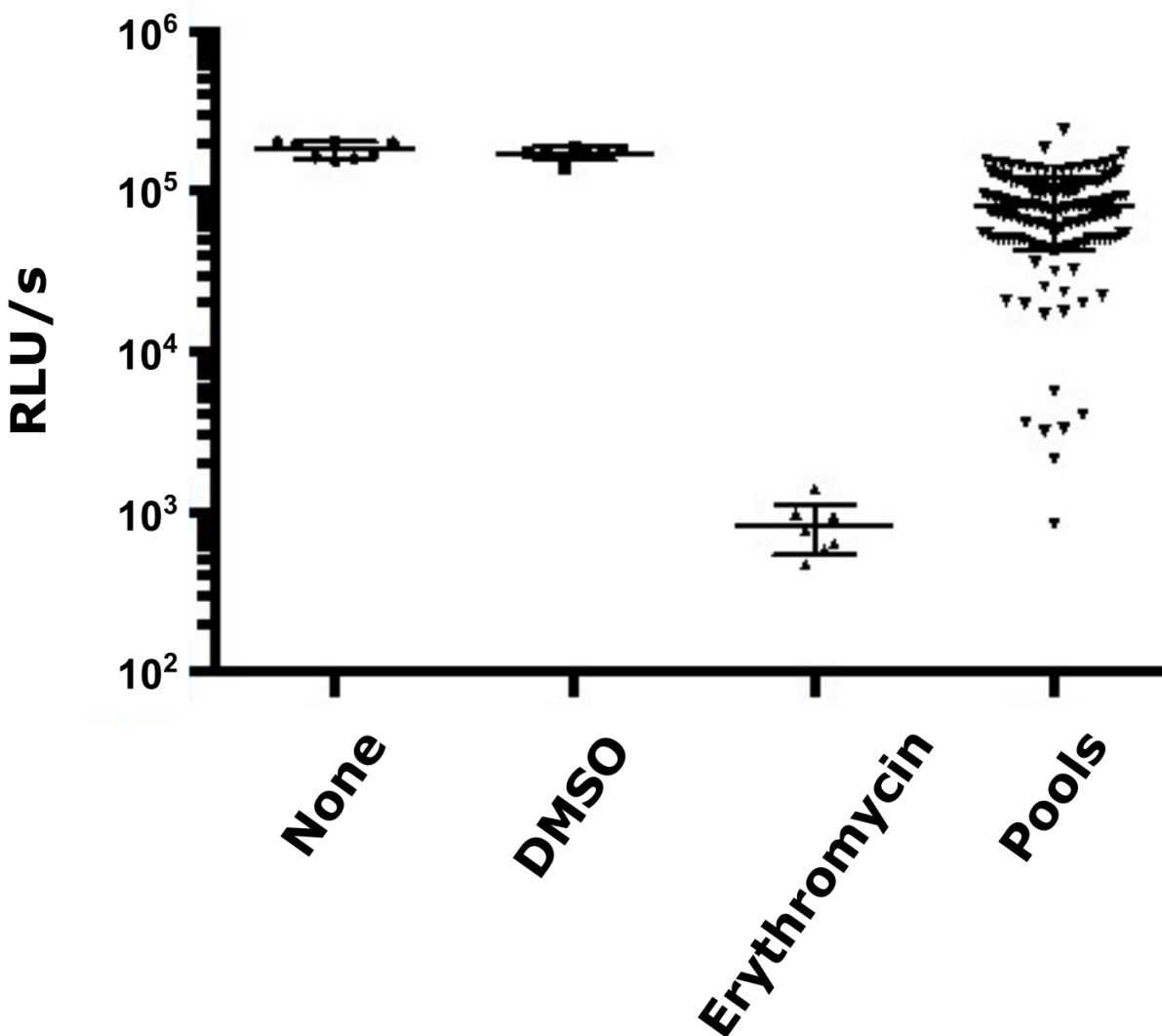
## Additional figures

### *Plasmodium falciparum*

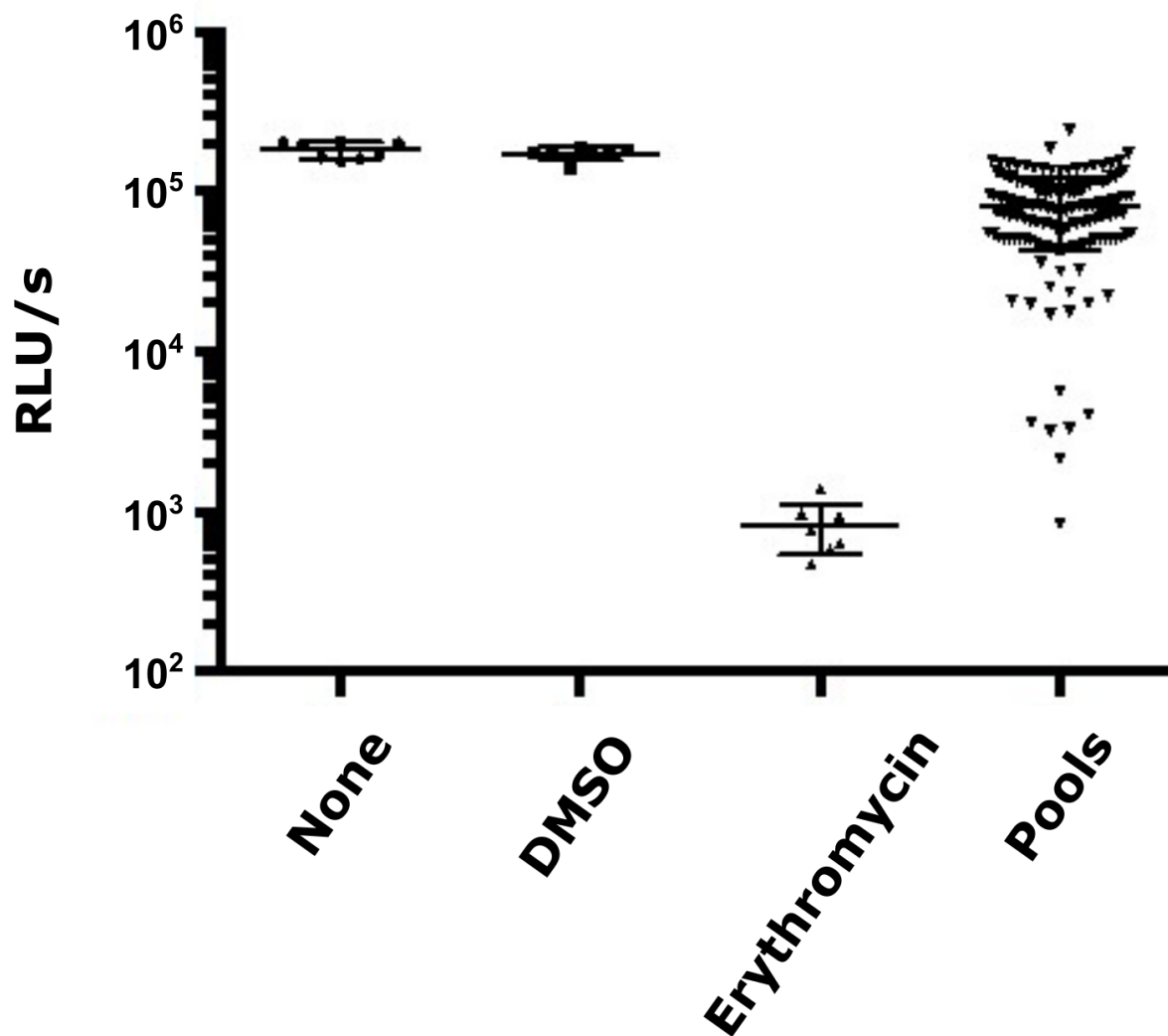


**Figure S1.** Dose-response curves for compound 1 against Pf3D7 (A), PfDd2 (B), HEK293T (C), and HeLa (D) cell lines. The x-axis is on a log<sub>10</sub> scale.

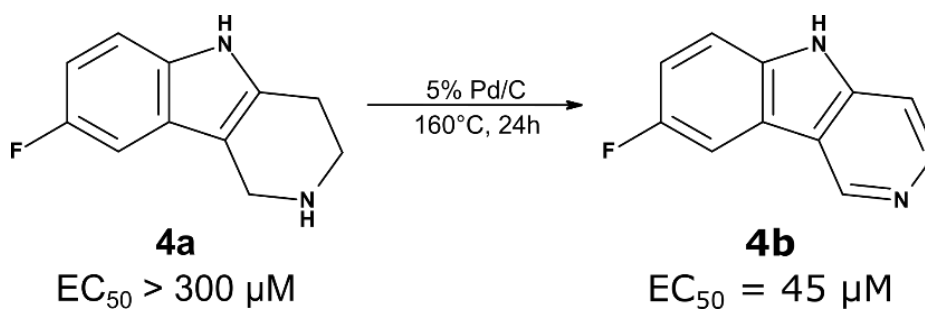
*Neisseria meningitidis*



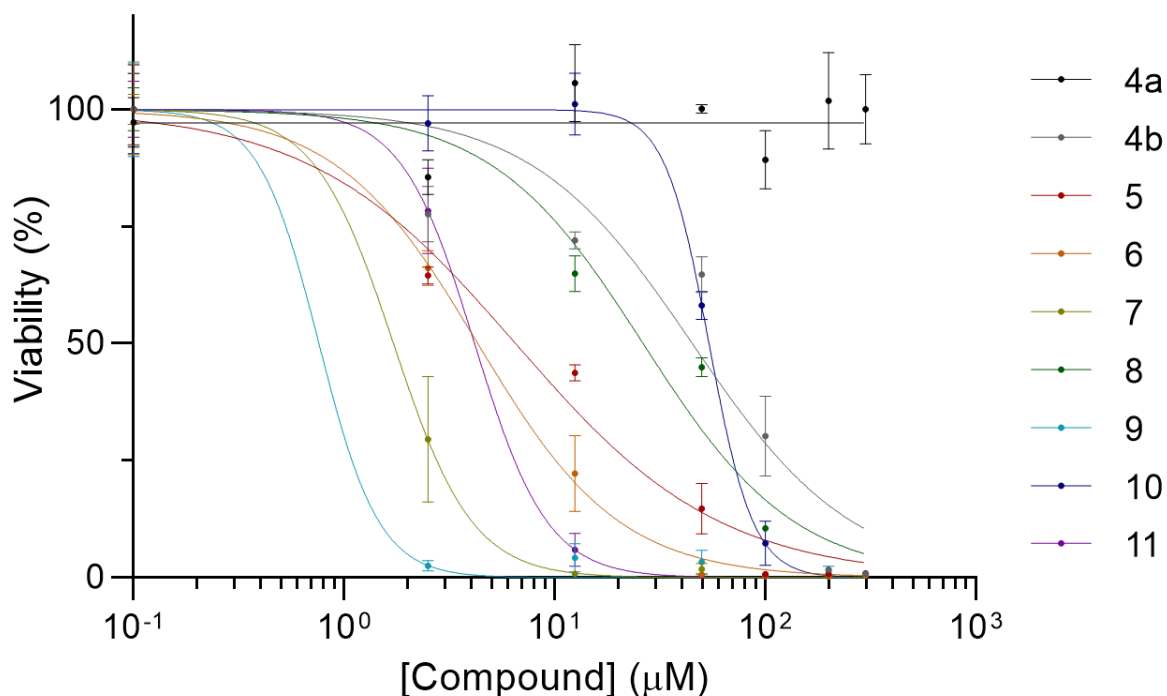
**Figure S2.** Library screening for the identification of compounds with the ability to inhibit the 24 h growth of *N. meningitidis*. Pools were tested at 83  $\mu$ M using a luciferase-based assay. *None*: no DMSO/compounds added; *DMSO*: 1% DMSO vehicle; *Erythromycin*: 4  $\mu$ M.



**Figure S3.** Singleton deconvolution for the identification of compounds with the ability to inhibit the 24 h growth of *N. meningitidis*. Singletons were tested at 83  $\mu\text{M}$  using a luciferase-based assay. None: no DMSO/compounds added; DMSO: 1% DMSO vehicle; Erythromycin: 4  $\mu\text{M}$ .



**Figure S4.** The most active inhibitor hit (4) of *N. meningitidis* growth was comprised of a mixture of inactive **4a** and active **4b**. Oxidation was experimentally reproduced and yielded the active molecule.



**Figure S5.** *N. meningitidis* growth inhibition curves after 16h. Data was normalized to DMSO vehicle control.

### Experimental information

#### General screening conditions

All fragment screens described in this report were performed as described in reference (1) using the NMX Research and Solutions Inc. Classic <sup>1</sup>H Library (<https://www.nmxresearch.com/>). A subset of the NMX library is available as Key Organics' BIONET library (<https://www.keyorganics.net/services/bionet-products/fragment-libraries/>). The library consists of 1604 fragments combined in 169 mixtures (pools) of 7-12 compounds per pool. Each fragment in these pools was each at 8.3 mM concentration in DMSO-d<sub>6</sub>. Singleton deconvolution was done

by testing all the individual fragments from the active pools. Singletons stock solutions were at 100 mM concentration in DMSO-d<sub>6</sub>.

### **Leishmania screens**

Experiments performed on *Leishmania* were previously described.<sup>1</sup>

### **Plasmodium screens**

#### **General**

Screens were performed using SYBR Gold assays on the asexual stage of *P. falciparum* 3D7 in infected red blood cells. Primary screen was done at 10 μM. Active pools were then deconvoluted and singleton compounds were re-tested at 10 μM. The most interesting fragments were then tested in dose-response curve to determine their EC<sub>50</sub>. Counter-screens were performed on HEK293T and HeLa cell lines to assess compound cytotoxicity. Best compounds were also tested against multidrug-resistant strain Dd2. Each condition was tested in triplicates and DMSO levels were kept below 0.5% (v/v) in the assays.

#### **Parasite culture**

*Plasmodium falciparum* 3D7 and Dd2 strains were obtained through BEI Resources, NIAID, NIH, contributed by Daniel J. Carucci and David Walliker (# MRA-102 and # MRA-150 respectively). Asexual stage *P. falciparum* parasites were cultured *in vitro* under standard conditions in RPMI-1640 (Gibco)-HEPES-buffered medium supplemented with 0.5% (w/v) Albumax (Invitrogen) at 4% hematocrit (human red blood cells (RBC) of O+ group). Parasites were kept at 37°C in a gas mixture composed of 5% oxygen, 5% carbon dioxide and 90% nitrogen. Parasites synchronization was done using D-sorbitol treatment as previously described.<sup>2</sup>

#### ***In vitro* 72 hours susceptibility assays**

Parasite susceptibility was measured as previously described with minor modifications.<sup>3</sup> Briefly, sorbitol synchronized 4-to-12 hours post-invasion ring stage parasites were exposed to a 16-point serial dilution of each molecule in a 96-well plate at 1% hematocrit and 0.5% starting parasitaemia in 100 μl final volume. Cells were incubated for 72 hours before being disrupted and the released

DNA stained by adding 25  $\mu$ l of 5X lysis buffer containing 0.16% saponin, 20mM Tris-HCl pH 7.5, 5mM EDTA, 1.6% Triton X-100, 5X SYBR™ Gold (Invitrogen). Each assay was done in triplicate. Plates were sealed and allowed to rest at room temperature for 24 hours. Relative fluorescence units (RFU) were measured with a VICTOR plate reader (PerkinElmer) at an excitation of 494 nm and emission of 530 nm. RFU were compared to untreated parasites and data plotted using GraphPad PRISM® software (GraphPad Software, San Diego, California USA, [www.graphpad.com](http://www.graphpad.com)) and  $IC_{50}$  determined using with the curve-fitting algorithm *log(inhibitor) versus response–Variable slope*.

### **Neisseria screens**

Screen and deconvolution were performed as previously described.<sup>4</sup>

### **Mycobacterium screens**

For the initial screen of the library, the amount of light produced after 24h and 48h of growth for *M. smegmatis* luminescent strain (harboring pLX-HSP60) was measured as previously described using decanal as a substrate.<sup>5</sup> The emitted light was measured using a 96-well plate luminometer (PerkinElmer Wallac 1420 Victor<sup>3</sup>) and results are expressed in relative light unit per second (RLU/s). All these assays were minimally performed in triplicates. Pool deconvolution was done against both wild-type *M. smegmatis* mc2155 and *M. tuberculosis* Erdman by a resazurin assay using serial dilution from 2000  $\mu$ M to 0.031  $\mu$ M. Ciprofloxacin and moxifloxacin were used as control drugs. In detail, exponentially growing cultures were diluted to  $OD_{600nm}$  0.005 using Middlebrook 7H9 broth, enriched with 0.5% (w/v) glycerol, 0.05% bovine serum albumin, 0.02% dextrose, 0.081% NaCl, in the absence or presence of 0.1% Tyloxapol. Diluted cultures were dispensed into a 96-well plate (100  $\mu$ l per well). The highest concentration of a given compound was dispensed into the first well of a column, containing 200  $\mu$ l of cell suspension. Drug was mixed and serially diluted through the whole column, except the last well, which was used as a control for cell viability. *M. smegmatis* plates were incubated at 37°C for 24 h, and *M. tuberculosis* plates were incubated at 37°C for 1 week. At the end of the incubation period, 5  $\mu$ l of 0.01% resazurin were added per well, mixed and plates were incubated at 37°C for 24 h. Colorimetric assay was interpreted as drug cidalty (blue) or cell viability (pink). MIC was scored as the lowest drug concentration causing cidalty. Experiments were repeated twice in *M. smegmatis* and four times in *M. tuberculosis*.



## **Flavivirus screens**

### **Cell culture**

Hepatocarcinoma Huh 7.5 cells were cultured in DMEM media containing 10% fetal-bovine serum, 1% non-essential amino acids and 1% penicillin-streptomycin. Cells were kept at 37°C with 5% CO<sub>2</sub>.

### **Cell viability assay**

Huh7.5 cells were cultured overnight in 10cm petri dishes. The following day, they were trypsinized, plated in 96 wells plate (15,000 cells/well) and treated with compounds at 150 µM and 15 µM. DMSO was used as negative control. Cell viability was assessed two days later by Cell-Titer Glo. Briefly, 50 uL Cell-Titer Glo reagent (1:1 PBS) were added to each well. Luminescence was evaluated using Spark® microplate reader (Tecan).

### **Anti-flaviviral activity**

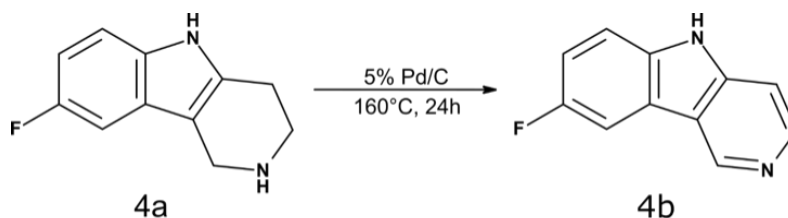
Viral replication was assessed using reporter viruses ZIKV-R2A and DENV-R2A. Viral stocks were produced as previously described.<sup>6, 7</sup> Huh7.5 were cultured overnight and infected with ZIKV-R2A or DENV-R2A at a multiplicity of infection (MOI) of 0.005. Cells were trypsinized 4 hours post-infection and plated in 96 wells plate (15,000 cells/well). Compounds were then added at indicated concentrations. At 48 hours post-infection, luciferase assay was performed, as previously described.<sup>8</sup> Briefly, cells were resuspended in 100 uL lysis buffer (0.1% Triton X-100, 25 mM glycylglycine [pH 7.8], 15 mM MgSO<sub>4</sub>, 4 mM EGTA (pH 8), and 1 mM DTT). 150 uL of assay buffer (25 mM glycylglycine [pH=7.8], 15 mM MgSO<sub>4</sub>, 4 mM EGTA [pH=8], 15 mM K<sub>2</sub>PO<sub>4</sub> [pH=7.8]) and coelenterazine (1.43 µM) was added to 30 uL lysate. Luminescence was measured using Spark® microplate reader (Tecan).

## **Synthetic procedures**

### **General**

All chemicals were obtained from commercial suppliers and used as received. NMR spectra were recorded on a 400 MHz Varian spectrometer and referenced to residual solvent (DMSO- $d_6$ :  $\delta$  2.50). TLCs were performed on 20  $\mu$ m thick Silica Plate. HPLC analysis was performed on a Hewlett-Packard 1100 system using a Phenetex C18 column equipped with a variable wavelength detector. Acetonitrile water mixture was used as eluent in a 70/30 ratio. LC-MS analysis was performed using an Agilent 1100 HPLC with Waters ZQ. The HPLC separation was carried out on an Agilent Zorbrax XDB-C18 column with dimension 4.6 x 30 mm, i.d., 3  $\mu$ m particle size. Initial solvent mixture for the elution was 95% water and 5% MeOH with 0.1% TFA in each, over 4 min at a flow rate of 2 mL/min.

### Synthesis of 8-Fluoro-5H-pyridyl[4,3-b]indole (**4b**)

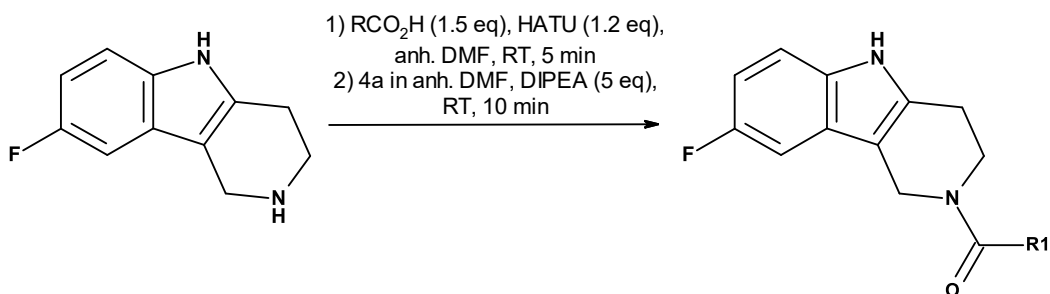


The amine  $R_2NH$  **4a** (50 mg, 0.26 mmol) and Pd/C (50 mg) were mixed and stirred overnight at 160°C. The reaction mixture was cooled to room temperature and MeOH (10 mL) was added.

The mixture was sonicated for 5 min. and then centrifuged. The MeOH supernatant was decanted and evaporated to obtain a pale-yellow solid. The crude product was purified on silica ( $R_f$  0.12 in EtOAc/MeOH 9:1) to afford the desired product **4b** (8 mg, 16% yield) as a pale-yellow solid.

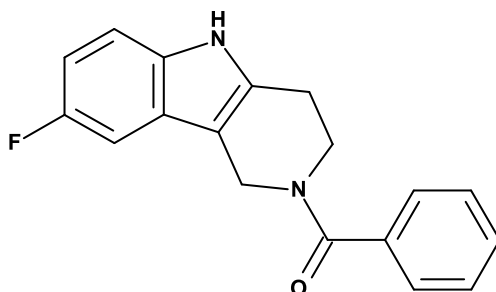
$^1H$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  11.77 (br s, 1H), 9.32-9.38 (m, 1H), 8.42 (d,  $J$ =5.86 Hz, 1H), 8.09 (dd,  $J$  = 2.34, 9.38 Hz, 1H), 7.56 (dd,  $J$  = 4.49, 8.79 Hz, 1H), 7.47 (dd,  $J$  = 0.98, 5.67 Hz, 1H), 7.32 (dt,  $J$  = 2.34, 9.18 Hz, 1H). LCMS (ES $^+$ )  $m/z$  found 187.2 [M+H]; retention time 1.26 s, 100% pure. HPLC, retention time 2.75 s.

### General procedure for the amide formation



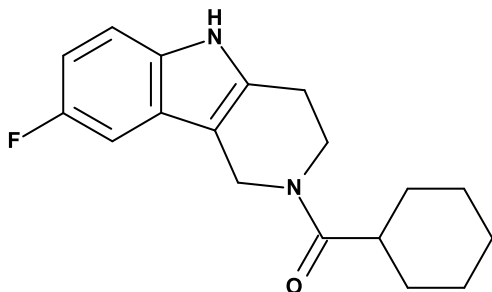
The appropriate carboxylic acid (1.5 eq.) and HATU (1.2eq.) were dissolved in DMF (1 mL). The colorless solution was allowed to stir for 5 minutes at RT. The amine (0.26 mmol, 1 eq.) in DMF (1 mL) was added followed by the addition of DIPEA (0.2 mL). After 10 min., the resultant yellow solution was transferred to saturated brine (15mL). EtOAc (15mL) was added and the organic phase washed thrice with brine. The EtOAc phase was further washed twice with saturated  $\text{Na}_2\text{CO}_3$  and  $\text{NH}_4\text{Cl}$ . It was then passed over  $\text{Na}_2\text{SO}_4$  and then evaporated to dryness. After column chromatography on silica, the desired products were obtained as two isomeric mixture (A and B) in varying proportions.

**(8-Fluoro-1,3,4,5-tetrahydropyrido[4,3-b]indol-2-yl)-phenyl-methanone (5)**



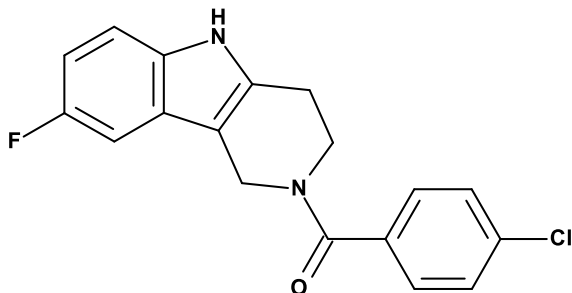
Yield: 40 mg, 52% as a yellow solid;  $R_f=0.23$  (1:1 hexane/EtOAc);  $^1\text{H NMR}$  (400 MHz,  $\text{DMSO-d}_6$ ): Isomer A (major 62%)  $\delta$  11.07 (br s, 1H), 7.54-6.8 (m, 8H), 4.75 (br s, 2H), 3.64 (bs, 2H), 2.86 (m, 2H). Isomer B (minor 38%)  $\delta$  11.07 (bs, 1H), 7.54-6.8 (m, 8H), 4.55 (br s, 2H), 4.01 (bs, 2H), 2.86 (m, 2H). LCMS ( $\text{ES}^+$ )  $m/z$  found 295.2  $[\text{M}+\text{H}]$ ; retention time 1.79 s, 100% pure. HPLC, retention time 5.03 s

**Cyclohexyl-(8-fluoro-1,3,4,5-tetrahydropyrido[4,3-b]indol-2-yl)methanone (6)**



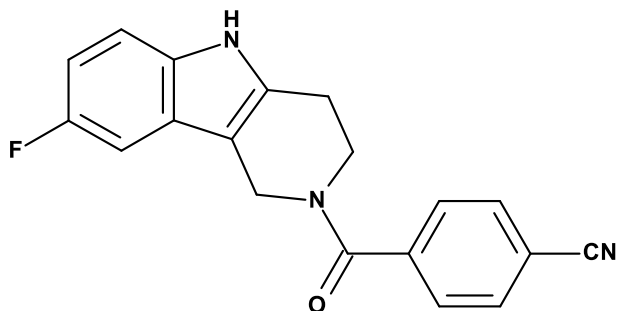
Yield: 26 mg, 33% as a pale yellow oil;  $R_f=0.35$  (9:1 EtOAc/MeOH);  $^1\text{H NMR}$  (400 MHz, DMSO- $d_6$ ) Isomer A (minor 47%)  $\delta$  11.03 (br s, 1H), 6.83-7.30 (m, 3H), 4.67 (s, 2H), 3.82 (t,  $J = 5.86$  Hz, 2H), 2.85 (br s, 1H), 2.71-2.73 (m, 2H), 1.33-1.36 (m, 10H). Isomer B (major 53%)  $\delta$  11.03 (br s, 1H), 6.83-7.30 (m, 3H), 4.57 (s, 2H), 3.82 (t,  $J = 5.86$  Hz, 2H), 2.85 (br s, 1H), 2.71-2.73 (m, 2H), 1.63-1.72 (m, 10H). LCMS ( $\text{ES}^+$ )  $m/z$  found 301.3 [ $\text{M}+\text{H}$ ]; retention time 1.98 s, 96% pure. HPLC, retention time 5.31 s.

**(4-Chlorophenyl)-(8-fluoro-1,3,4,5-tetrahydropyrido[4,3-b]indol-2-yl)methanone (7)**



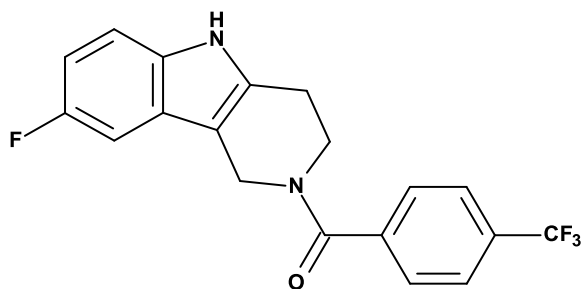
Yield: 42 mg, 48% as a yellow solid;  $R_f=0.32$  (1:1 hexane/EtOAc);  $^1\text{H NMR}$  (400 MHz, DMSO- $d_6$ ) Isomer A (major 63%)  $\delta$  11.07 (br s, 1H), 6.86-7.53 (m, 7H), 4.75 (m, 2H), (m, 1H), 3.56 (br s, 2H), 2.86 (br s, 2H). Isomer B (minor 37%)  $\delta$  11.07 (br s, 1H), 6.86-7.53 (m, 7H), 4.46 (m, 2H), 3.99 (br s, 2H), 2.86 (br s, 2H). LCMS ( $\text{ES}^+$ )  $m/z$  found 329.2 [ $\text{M}+\text{H}$ ]; retention time 1.97 s. HPLC, retention time 4.51 s.

**(4-Cyanophenyl)-(8-fluoro-1,3,4,5-tetrahydropyrido[4,3-b]indol-2-yl)methanone (8)**



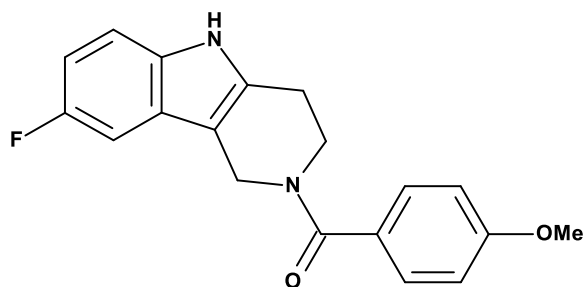
Yield: 51 mg, 61% as a yellow oil;  $R_f=0.32$  (2:8 hexane/EtOAc);  $^1\text{H NMR}$  (400 MHz,  $\text{DMSO-d}_6$ ): Isomer A (major 62%)  $\delta$  11.09 (br s, 1H), 6.88-7.97 (m, 7H), 4.78 (br s, 2H), 3.58 (br t,  $J=5.28$  Hz, 2H), 2.84-2.89 (m, 2H). Isomer B (minor 38%)  $\delta$  11.09 (br s, 1H), 6.88-7.97 (m, 7H), 4.48 (br s, 2H), 4.02 (br t,  $J=5.28$  Hz, 2H), 2.84-2.89 (m, 2H). LCMS ( $\text{ES}^+$ )  $m/z$  found 320.2  $[\text{M}+\text{H}]$ ; retention time 1.80 s, 100% pure. HPLC, retention time 3.88 s.

**(4-Trifluoromethylphenyl)-(8-fluoro-1,3,4,5-tetrahydropyrido[4,3-b]indol-2-yl)methanone (9)**



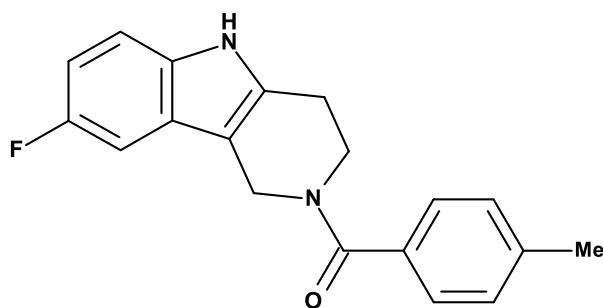
Yield: 58 mg, 61% as a pale yellow solid;  $R_f=0.40$  (1:1 hexane/EtOAc);  $^1\text{H NMR}$  (400 MHz,  $\text{DMSO-d}_6$ ) Isomer A (major 63%)  $\delta$  11.08 (br s, 1H), 6.77-7.86 (m, 7H), 4.79 (br s, 2H), 3.60 (br s, 2H), 2.85 (br s, 2H). Isomer B (major 37%)  $\delta$  11.08 (br s, 1H), 6.77-7.86 (m, 7H), 4.52 (br s, 2H), 4.03 (br s, 2H), 2.91 (br s, 2H). LCMS ( $\text{ES}^+$ )  $m/z$  found 363.2  $[\text{M}+\text{H}]$ ; retention time 2.00 s. HPLC, retention time 2.24 s.

**(4-Methoxyphenyl)-(8-fluoro-1,3,4,5-tetrahydropyrido[4,3-b]indol-2-yl)methanone (10)**



Yield: 45 mg, 53% as a yellow oil;  $R_f=0.22$  (1:1 hexane/EtOAc);  $^1\text{H NMR}$  (400 MHz,  $\text{DMSO-d}_6$ )  $\delta$  11.05 (br s, 1H), 6.78-7.93 (m, 7H), 4.65 (br s, 2H), 3.77-3.80 (br s, 2H), 3.79 (s, 3H), 2.84 (m, 2H). LCMS ( $\text{ES}^+$ )  $m/z$  found 325.3  $[\text{M}+\text{H}]$ ; retention time 1.83 s, 100% pure. HPLC, retention time 2.66 s.

**(4-Methylphenyl)-(8-fluoro-1,3,4,5-tetrahydropyrido[4,3-b]indol-2-yl)methanone (11)**



Yield: 32 mg, 40% as a yellow solid;  $R_f=0.26$  (1:1 hexane/EtOAc);  $^1\text{H NMR}$  (400 MHz,  $\text{DMSO-d}_6$ ) Isomer A (major 61%)  $\delta$  11.07 (br s, 1H), 6.81-7.42 (m, 7H), 4.72 (br s, 2H), 3.65 (br s, 2H), 2.86 (br s, 2H), 2.26 (s, 3H). Isomer B (minor 39%)  $\delta$  11.07 (br s, 1H), 6.81-7.42 (m, 7H), 4.57 (br s, 2H), 3.98 (br s, 2H), 2.86 (br s, 2H), 2.36 (s, 3H). LCMS ( $\text{ES}^+$ )  $m/z$  found 309.2  $[\text{M}+\text{H}]$ ; retention time 1.91 s, 98% pure. HPLC, retention time 3.76s.

## **References**

- [1] Ayotte, Y., Bilodeau, F., Descoteaux, A., and LaPlante, S. R. (2018) Fragment-Based Phenotypic Lead Discovery: Cell-Based Assay to Target Leishmaniasis, *Chemmedchem* 13, 1377-1386.
- [2] Lambros, C., and Vanderberg, J. P. (1979) Synchronization of *Plasmodium falciparum* erythrocytic stages in culture, *J Parasitol* 65, 418-420.
- [3] Johnson, J. D., Denuff, R. A., Gerena, L., Lopez-Sanchez, M., Roncal, N. E., and Waters, N. C. (2007) Assessment and continued validation of the malaria SYBR green I-based fluorescence assay for use in malaria drug screening, *Antimicrob Agents Chemother* 51, 1926-1933.
- [4] Bernet, E., Lebughe, M., Vincent, A. T., Haghdoost, M. M., Golbaghi, G., Laplante, S., Castonguay, A., and Veyrier, F. J. (2021) Sodium Tetraphenylborate Displays Selective Bactericidal Activity against *Neisseria meningitidis* and *N. gonorrhoeae* and Is Effective at Reducing Bacterial Infection Load, *Antimicrobial Agents and Chemotherapy* 65, e00254-00220.
- [5] Veyrier, F., Saïd-Salim, B., and Behr, M. A. (2008) Evolution of the Mycobacterial SigK Regulon, *Journal of Bacteriology* 190, 1891-1899.
- [6] Shan, C., Xie, X., Muruato, A. E., Rossi, S. L., Roundy, C. M., Azar, S. R., Yang, Y., Tesh, R. B., Bourne, N., Barrett, A. D., Vasilakis, et al. (2016) An Infectious cDNA Clone of Zika Virus to Study Viral Virulence, Mosquito Transmission, and Antiviral Inhibitors, *Cell Host Microbe* 19, 891-900.
- [7] Fischl, W., and Bartenschlager, R. (2013) High-Throughput Screening Using Dengue Virus Reporter Genomes, In *Antiviral Methods and Protocols* (Gong, E. Y., Ed.), pp 205-219, Humana Press, Totowa, NJ.
- [8] Chatel-Chaix, L., Fischl, W., Scaturro, P., Cortese, M., Kallis, S., Bartenschlager, M., Fischer, B., and Bartenschlager, R. (2015) A Combined Genetic-Proteomic Approach Identifies Residues within Dengue Virus NS4B Critical for Interaction with NS3 and Viral Replication, *J Virol* 89, 7170-7186.

## 14 GENERAL DISCUSSION AND CONCLUSION

---

### 14.1 Compound Aggregation

We have reported the implementation of an T2-CPMG NMR assay to detect compound self-aggregation. The assay has various advantages such as high sensitivity to intermolecular interactions, while being quick to perform. This makes it an advantageous method to be used in a screening context or for library curation. Despite all this, there are some limitations to the experiment. Firstly, intrinsic exchange phenomena (e.g. rotamers, tautomers, etc.) can result in faster relaxation and may result in falsely flagging a compound as an aggregator. However, if such molecules possess multiple protons, it is often possible to determine if intrinsic exchange is occurring, as not all protons would experience comparable decay rates (i.e. protons involved in the exchange may exhibit very fast decay rates, while other protons may decay at the same rate as what can be expected for a “well-behaved” compound). Similarly, resonances close to the water suppression area can also have their relaxation profile impacted, resulting in potential false-positive flags for aggregation. Therefore, looking at the overall profile of the molecule is important and additional experiments should be performed to confirm aggregation

The published T2-CPMG assay is based on the use of aromatic protons to monitor aggregation since their analysis is more straightforward in large part due to the lack of overlap with common buffer components. However, as some compounds do not possess any aromatic  $^1\text{H}$ , similar guidelines would be required to expand the method to aliphatic  $^1\text{H}$ . Several challenges would be expected with the use of aliphatic  $^1\text{H}$  to monitor aggregation, such as the overlap with signals from the buffer or solvent. Also, it is conceivable that different guidelines could be needed based on the type of aliphatic protons evaluated (i.e. CH, CH<sub>2</sub>, CH<sub>3</sub>) do to differences in intrinsic relaxation rates. A similar assay could be also implemented using  $^{19}\text{F}$  NMR. However, given the large contribution of chemical shift anisotropy (CSA) on the relaxation of the  $^{19}\text{F}$  spin, developing guidelines would likely prove challenging and may require the use of computational prediction to correct for these CSA effects in order to obtain  $^{19}\text{F}$  T2 values that are more representative of tumbling rates and exchange phenomena.

In addition to the T2-CPMG assay, we also reported the implementation of an NMR-focused protocol to monitor aggregation of small molecules. This protocol is designed to be a practical workflow which provides guidelines as to which experiments can be performed, in which order, and what are the possible conclusions based on the data that emanates from these various assays.



Despite being fairly comprehensive, this protocol could be further expanded in the future by integrating additional assays. For example, by adding an enzyme-inhibition assay such as a AmpC  $\beta$ -lactamase assay to determine if an aggregate exhibits evidence of non-specific inhibition of enzymatic activity <sup>75</sup>. Similarly, non-functionalized gold sensors such as the ones used in surface plasmon resonance could be used to detect non-specific interaction of these aggregates with the sensor chip <sup>67</sup>. The described detergent assay could also be expanded to include multiple detergent concentrations as some aggregate molecules can be more resistant to detergent, while others could even be stabilized by the presence of detergent via surface passivation <sup>75,173</sup>.

Data generated by the aforementioned aggregation assays could be also fed to artificial intelligence (AI) models to improve predictive models. In the past decades, various tools have been published that try to predict misbehaved compounds based on computational models and machine learning <sup>174-176</sup>. However, one major challenge pertains to the fact that aggregators do not appear to be limited to specific scaffold and are likely highly diverse in structure. Indeed, a report has analyzed the chemical scaffolds of 319 aggregators and found that almost 75% of scaffolds were unique (i.e. found in only one aggregator) <sup>177</sup>. The fact that aggregation is condition-dependent also increases the challenge associated with predicting the phenomenon as it is not simply a binary event likely explains and likely partly explains why the successes have so far been limited. Moreover, a lot of the experimental data fed into AI originated from DLS, which suggests that the current published models may be overlooking some types of aggregates that are not detected by this technique, such as small aggregates.

Despite all the problems associated with this phenomenon, aggregation is also observed for several approved drugs which means that at a later stage in drug development, evidence of aggregation does not necessarily result in a “death sentence” for a molecule, particularly if the concentration at which the molecule aggregates is significantly higher than the concentration at which it exerts its therapeutic effects <sup>71,178-180</sup>. There have also been some published examples where the aggregation properties of some molecules appeared to be advantageous. For example, some aggregates of non-nucleoside reverse transcriptase inhibitors, a class of anti-AIDS drug, have been found to exhibit improved solubility under certain conditions. Indeed, aggregates with sizes ranging from 30-100 nm were associated with favorable absorption, while larger aggregates of >250 nm were associated with poor absorption <sup>59,181</sup>.

These kinds of observations have resulted in some efforts in trying to leverage the properties of aggregating molecules to improve drug delivery <sup>182</sup>. In a similar fashion, it was reported that some aggregates could be optimized to be used as stabilizing agents to increase protein stability <sup>183,184</sup>.

Another example comes from a team at Janssen that reported a cocrystal structure obtained between TNF $\alpha$  and a compound aggregate where the aggregate induced a quaternary change in TNF $\alpha$ , leading to inhibition of the interaction of TNF $\alpha$  with its TNFR1 and TNFR2 receptors <sup>185</sup>. Similarly, some aggregators were also reported to inhibit amyloid fiber formation, likely via protein sequestration mechanism <sup>186</sup>, which opens up the questions of whether this mechanism could be exploited for the treatment of diseases such as Alzheimer, although recent controversies are casting some doubts over the link between amyloid fibers and the disease <sup>187</sup>.

Albeit interesting, translating aggregation phenomena into a positive clinical outlook represents a daunting and risky task as it is very difficult to predict the potential adverse effects that such molecules could induce. Optimizing aggregates to improve activity or pharmacological properties would also be very challenging and may only be possible through a lot of trial-and-error. Nevertheless, as stated in the published reports, at the early stage of drug discovery, it is probably safer to exclude aggregating molecules and prioritize well-behaved compounds as aggregation at this stage is likely to occur in a similar concentration range as the compounds' binding affinities/activities and could result in artifactual data.

## 14.2 “NMR for SAR” platform

We have implemented an NMR-centric platform coined “NMR for SAR” which takes advantages of ligand-observed NMR techniques, effectively circumventing several of the limitations of the traditional “SAR by NMR” approach that focuses on protein-observed 2D screening. This platform presents several advantages: There are no theoretical limit on the size of the protein being screened, the protein does not need to be isotopically labeled (even though availability of labeled protein can be useful for additional downstream investigation), the quantities of protein required for a screen are significantly reduced, compounds of weaker binding affinities can be detected. Most importantly, binding scores are extracted from the data which provide information as to the binding affinities of the molecules. The scoring is calculated based on single-concentration data, effectively increasing the throughput of the platform. Moreover, since this scoring system relies on the aforementioned concept of protein-bound ligand, this reduces the compound solubility requirements which would limit binding assessment or affinity determination of relatively weaker compounds using other biophysical techniques or protein-detected NMR due to the necessity in sufficiently saturating the protein with ligand.

Since NMR provides direct atomic information of the sample, the platform also allows robust monitoring of ligand concentration, purity and integrity. Similarly, the protein integrity (folding, signal intensity, purity) is also evaluated in parallel on the same samples. Moreover, most biophysical techniques used for drug screening require labeling or immobilization of the protein target. In contrast, NMR is a label- and immobilization-free method (although both are also possible by NMR) which allows the study of a more native form of the protein and would be expected to result in better translatability of the results.

We applied this platform on HRas<sup>G12V</sup> and identified a ~7-10 mM fragment hit which was then optimized using the aforementioned binding scores to guide synthetic chemistry efforts. By testing less than 200 analogs we obtained  $\mu$ M-nM affinity compounds which also exhibit activity in a nucleotide release functional assay, as well as preferential inhibition of cancerous cell lines. The binding affinity of key compounds were also orthogonally validated using other biophysical techniques and correlated very well with the NMR binding scores. X-ray efforts have also been deployed, however, despite being successful at elucidating a crystallographic structure of the apo HRas<sup>G12V-GDP</sup> protein, attempts in obtaining cocrystals with ligands have not been fruitful so far. This highlights the challenges of obtaining cocrystal structures to guide structure-based drug design. Nevertheless, the platform could successfully be applied, even in the absence of X-ray data and further evaluation of these compounds will be needed to better understand their activity.

Nevertheless, just like any technique, this platform presents some limitations. Firstly, the scoring system assumes that the compounds being investigated are in a fast-exchange binding regime with the target protein. As the binding affinities improve (usually around the lower  $\mu$ M affinity range), the validity of ligand-observed scoring is expected to diminish and other methods or conditions should be used to assess the binding affinities of these tighter binders. Another limitation pertains to the fact that in comparison with protein-observed experiments, ligand-detected methods can be more skewed by contribution from non-specific or multi-site binding which can therefore impact the binding scores. As such, the use of such a platform on more complex systems requires more caution and the binding stoichiometry should be regularly assessed using other methods. Moreover, whereas there is no theoretical limit on the size of the protein that can be targeted using ligand-detected NMR approaches, 1D and particularly 2D protein-observed experiment will become more difficult to perform with increasing protein size. Similarly, such protein-observed experiment may not be possible on protein systems that are very challenging to produce,

### 14.3 Fluorine NMR referencing

In the published report, we tested ten fluorine reference candidates to assess their suitability for NMR screening uses. Four of these molecules were deprioritized due to limited aqueous solubility and another one due to incompatibility with magnesium chloride and calcium chloride salts. Among the remaining five candidates, none could be considered as an ideal reference since they all possessed their own disadvantages. We highlighted these limitations, as well as the advantages of each of these candidates so that users can better decide which reference molecule may suit their specific needs.

We also described a workflow that could be used to screen other reference candidates in the hopes of finding even better performing molecules. Also, given that none of the resulting five candidates in our study possess a monofluoride moiety such as a CF, it could be pertinent to expand the search for monofluorinated references, especially if screening using narrower bandwidths which may not allow for concomitant detection of CF and CF<sub>3</sub> for example. However, with the ever-expanding adoption of broadband excitation pulses, this should become less and less of an issue.

### 14.4 Fragment-based phenotypic screening

We have described several proof-of-concept studies where a curated fragment library was used to perform phenotypic screens against various pathogens: *Leishmania* and *Plasmodium* parasites, *Neisseria* and *Mycobacterium* bacteria, as well as Zika and dengue viruses. Overall, all these screening efforts resulted in the identification of interesting hits which could serve as starting points for further chemical optimization. Early SAR evaluation was performed for some chemical series and some SAR trends seemed to emerge, suggesting that the compounds were exerting their activity via productive interactions and that the effects were not simply due to unspecific binding. Therefore, combining the use of fragment-based libraries to phenotypic approaches appears to be a viable option to probe unexplored molecular targets, while mitigating the usually lower throughput of phenotypic approaches.

Interestingly, one of the best hits in the *Plasmodium falciparum* screen was very closely related to a hit previously identified through the high-content screening of an AstraZeneca library of half a million molecules that was not fragment-oriented. Despite being fragment-sized, this “AstraZeneca” hit was then optimized into a molecule exhibiting good bioavailability and efficacy

in a mouse model <sup>188</sup>. This further supported our claim that fragment molecules can yield productive hits, even in phenotypic assays.

Another noteworthy observation is that several of the molecules identified in our screening and analoging efforts exhibited relatively potent cell activity, which would not necessarily have been expected for such small-molecular weight molecules. However, the IC<sub>50</sub> values observed here are in comparable ranges with previously reported fragment hits identified in target screens and directly tested in various antiparasitic assays against *Trypanosoma brucei* and *cruzi*, *Leishmania infantum*, as well as *Plasmodium falciparum* <sup>189</sup>.

One limitation of phenotypic approaches pertains to the challenge of optimizing a molecule in the absence of detailed knowledge about its mechanism of action <sup>129</sup>. Therefore, in absence of known target, X-ray crystallography or modeling efforts cannot be initiated to help medicinal chemists grow the molecules based on structural information. Despite such challenges, it is still possible to perform SAR in absence of structural data. Indeed, many drugs were discovered decades before the advent of X-ray crystallography of proteins and similarly, there has been significant efforts invested in drugging some GPCRs before any 3D structure of these proteins became available <sup>190-192</sup>.

The frequent preference toward target-based approaches is understandable from a scientific and intellectual standpoint as they tend to be more hypothesis-based in comparison with phenotypic approaches which are often more of a “black box” and can involve a significant amount of serendipity. This likely is one of the main reasons why, from the perspective of financial investments, a lot of decision-makers in pharmaceutical companies will consider target identification a prerequisite before being ready to invest significant amounts of resources into a chemical series <sup>193</sup>.

Therefore, there will usually be efforts to identify the molecular target(s) of compounds arising from phenotypic screens. Indeed, target identification represents a very active field and there have been significant advances in techniques for target identification and these technologies will continue improving, which should further help advance compounds identified via phenotypic methods in the future <sup>193-197</sup>. However, this can be a quite challenging tasks, especially in the case of relatively lower potency compounds/hits <sup>198</sup>. Ever improving artificial intelligence applications may provide opportunities to shed some light on the potential mechanism of action (MoA).

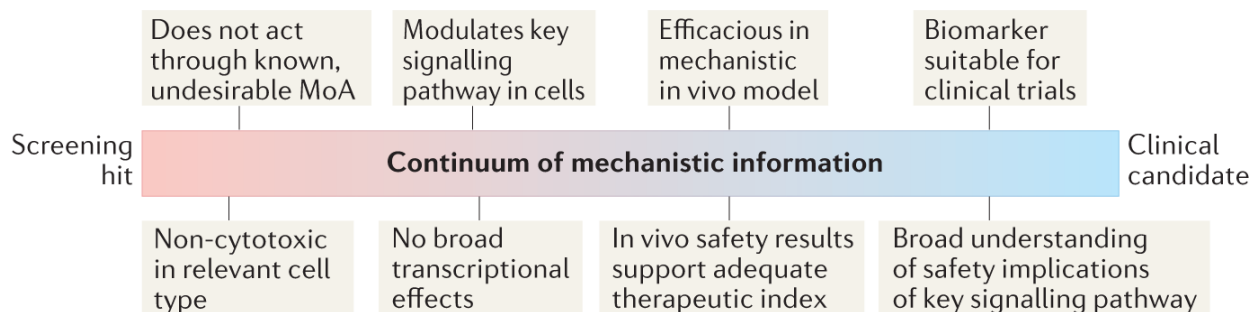
Nonetheless, there are a lot of debates and discussions around the concept of “target identification” considering that knowledge of MoA is not a purely binary concept. As such, Figure

8.1 depicts an alternative view which was proposed where target identification efforts actually represent a continuum where knowledge is accumulated up to a point where enough information is known to provide confidence that a compound can be pushed forward to the clinic <sup>199</sup>. This alternative thinking is supported by the fact that identification of the drug target does not necessarily translate directly to the understanding of the mechanism of action as there can be a disconnect between the observed effects and the identified target(s). One example of this was noted for daclatasvir, an drug against the hepatitis C virus, where the identification of NS5A as a target did not fully explain the high potency and efficacy observed for the compound <sup>200</sup>.

**a Legacy thinking**



**b Emerging thinking**



**Figure 14.1 Target identification – Legacy versus emerging thinking.**

Target identification used to be viewed particularly as a binary concept (a), but a continuum of information would be a better representation (b). Reproduced from <https://doi.org/10.1038/s41573-022-00472-w> with permission.

These kinds of observations are the reason why the idea of a drug being specific to only a specific molecular target is increasingly being questioned. Indeed, it was reported by the mining of public databases that FDA-approved drugs were known to interact with an average of six molecular targets <sup>201</sup>. These polypharmalogical effects can obviously be associated with adverse effects, but they can also explain the mechanism of action of certain drugs. Increasing examples of the latter scenario are being reported, such as the example of statins which were mostly discovered

due to their inhibition of HMG-CoA reductase, but which cannot completely explain their biological efficacy <sup>202</sup>.

However, it must be highlighted that knowledge of the mechanism of action is not a prerequisite for drug approval and an analysis of NMEs drugs approved by the FDA between 2000 and 2012 estimated that about 8% of molecules did not have known MoA <sup>203</sup>. Some recent examples include lacosamide <sup>204,205</sup> and lenalidomide <sup>206</sup> which have been approved in 2008 and 2005 respectively, but for which the MoA is either yet to be identified, or has been identified several years after approval.

This partly explains why PDD tends to have better translatability for more complex diseases such as for central nervous system (CNS) diseases where multiple targets modulation may be required for proper disease modulation, as evidenced by the fact that the majority of marketed CNS drugs act on multiple targets <sup>134,207</sup>. Moreover, due to the challenges associated with crossing the blood-brain barrier, smaller-sized molecules such as fragments would be expected to improve activity in more biologically complex models. Smaller molecules such as fragments therefore hold the potential to probe multiple targets in more biologically relevant models, which would have been considered a problem in the past due to concerns for off-targets effects, but in recent years there has been a gradual recognition of the potential for such polypharmalogical modulation <sup>208</sup>. Indeed, there has even been a recent report from Lipinski (who coined the aforementioned “Lipinski’s Rule of 5”) highlighting the untapped potential of low molecular weight compounds such as fragments in phenotypic assays <sup>209</sup>.

In conclusion, target- and phenotypic-based approaches both have advantages and have both been successful in bringing chemical matters into the clinic. These approaches should therefore not be considered as being in opposition to one another, but rather as complementary. For example, PDD has the potential to yield tool compounds or probes that can result in the identification of new protein targets involved in diseases <sup>210,211</sup>, which can then be tackled using target-based approaches. In addition, there are also emerging hybrid approaches and an ever-increasing portion of drug discovery efforts fall into an intermediate category between pure phenotypic and target-based approaches which has been coined by Moffat et al. as “mechanism-informed phenotypic drug discovery” <sup>212</sup>.

## 15 REFERENCES

---

- 1 Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nature Reviews Drug Discovery* **6**, 211, doi:10.1038/nrd2220 (2007).
- 2 McGovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *Journal of Medicinal Chemistry* **45**, 1712-1722, doi:10.1021/jm010533y (2002).
- 3 Huth, J. R. *et al.* ALARM NMR: A Rapid and Robust Experimental Method To Detect Reactive False Positives in Biochemical Screens. *Journal of the American Chemical Society* **127**, 217-224, doi:10.1021/ja0455547 (2005).
- 4 Rishton, G. M. Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discovery Today* **8**, 86-96, doi:<https://doi.org/10.1016/S1359644602025722> (2003).
- 5 Rishton, G. M. Reactive compounds and in vitro false positives in HTS. *Drug Discovery Today* **2**, 382-384, doi:[https://doi.org/10.1016/S1359-6446\(97\)01083-0](https://doi.org/10.1016/S1359-6446(97)01083-0) (1997).
- 6 Hajduk, P. J. & Burns, D. J. Integration of NMR and high-throughput screening. *Comb. Chem. High Throughput Screen.* **5**, 613-621, doi:10.2174/1386207023329996 (2002).
- 7 Jencks, W. P. On the attribution and additivity of binding energies. *Proceedings of the National Academy of Sciences* **78**, 4046-4050, doi:10.1073/pnas.78.7.4046 (1981).
- 8 Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-1534 (1996).
- 9 Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings PII of original article: S0169-409X(96)00423-1. The article was originally published in *Advanced Drug Delivery Reviews* 23 (1997) 3–25.1. *Adv. Drug Del. Rev.* **46**, 3-26, doi:[https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0) (2001).
- 10 Owens, J. Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discovery Today* **8**, 12-16, doi:[https://doi.org/10.1016/S1359-6446\(02\)02556-4](https://doi.org/10.1016/S1359-6446(02)02556-4) (2003).
- 11 Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today* **8**, 876-877, doi:[https://doi.org/10.1016/S1359-6446\(03\)02831-9](https://doi.org/10.1016/S1359-6446(03)02831-9) (2003).
- 12 Köster, H. *et al.* A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *J. Med. Chem.* **54**, 7784-7796, doi:10.1021/jm200642w (2011).
- 13 Martin, Y. C. Challenges and prospects for computational aids to molecular diversity. *Perspect. Drug Discovery Des.* **7**, 159-172, doi:10.1007/BF03380186 (1996).
- 14 Reymond, J.-L., van Deursen, R., Blum, L. C. & Ruddigkeit, L. Chemical space as a source for new drugs. *MedChemComm* **1**, 30-38, doi:10.1039/C0MD00020E (2010).
- 15 Fink, T., Bruggesser, H. & Reymond, J.-L. Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons. *Angew. Chem. Int. Ed.* **44**, 1504-1508, doi:<https://doi.org/10.1002/anie.200462457> (2005).



- 16 Murray, C. W., Verdonk, M. L. & Rees, D. C. Experiences in fragment-based drug discovery. *Trends Pharmacol. Sci.* **33**, 224-232, doi:<https://doi.org/10.1016/j.tips.2012.02.006> (2012).
- 17 Ma, R., Wang, P., Wu, J. & Ruan, K. Process of Fragment-Based Lead Discovery-A Perspective from NMR. *Molecules (Basel, Switzerland)* **21**, 854, doi:10.3390/molecules21070854 (2016).
- 18 Boyd, S. M. & de Kloe, G. E. Fragment library design: efficiently hunting drugs in chemical space. *Drug Discovery Today: Technologies* **7**, e173-e180, doi:<https://doi.org/10.1016/j.ddtec.2010.11.010> (2010).
- 19 Hann, M. M., Leach, A. R. & Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **41**, 856-864, doi:10.1021/ci000403i (2001).
- 20 Schuffenhauer, A. *et al.* Library design for fragment based screening. *Curr. Top. Med. Chem.* **5**, 751-762 (2005).
- 21 Leach, A. R. & Hann, M. M. Molecular complexity and fragment-based drug discovery: ten years on. *Curr. Opin. Chem. Biol.* **15**, 489-496, doi:<https://doi.org/10.1016/j.cbpa.2011.05.008> (2011).
- 22 Kuntz, I. D., Chen, K., Sharp, K. A. & Kollman, P. A. The maximal affinity of ligands. *Proceedings of the National Academy of Sciences* **96**, 9997-10002, doi:10.1073/pnas.96.18.9997 (1999).
- 23 Hopkins, A. L., Groom, C. R. & Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **9**, 430-431, doi:[https://doi.org/10.1016/S1359-6446\(04\)03069-7](https://doi.org/10.1016/S1359-6446(04)03069-7) (2004).
- 24 Keserü, G. M. & Makara, G. M. The influence of lead discovery strategies on the properties of drug candidates. *Nature Reviews Drug Discovery* **8**, 203-212, doi:10.1038/nrd2796 (2009).
- 25 Leeson, P. D. & Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews Drug Discovery* **6**, 881-890, doi:10.1038/nrd2445 (2007).
- 26 Brandt, P., Geitmann, M. & Danielson, U. H. Deconstruction of Non-Nucleoside Reverse Transcriptase Inhibitors of Human Immunodeficiency Virus Type 1 for Exploration of the Optimization Landscape of Fragments. *J. Med. Chem.* **54**, 709-718, doi:10.1021/jm101052g (2011).
- 27 Mortenson, P. N. & Murray, C. W. Assessing the lipophilicity of fragments and early hits. *J. Comput. Aided Mol. Des.* **25**, 663-667, doi:10.1007/s10822-011-9435-z (2011).
- 28 Abad-Zapatero, C. & Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* **10**, 464-469, doi:[https://doi.org/10.1016/S1359-6446\(05\)03386-6](https://doi.org/10.1016/S1359-6446(05)03386-6) (2005).
- 29 Verdonk, M. L. & Rees, D. C. Group Efficiency: A Guideline for Hits-to-Leads Chemistry. *Chemmedchem* **3**, 1179-1180, doi:<https://doi.org/10.1002/cmdc.200800132> (2008).
- 30 Wenlock, M. C., Austin, R. P., Barton, P., Davis, A. M. & Leeson, P. D. A Comparison of Physicochemical Property Profiles of Development and Marketed Oral Drugs. *J. Med. Chem.* **46**, 1250-1256, doi:10.1021/jm021053p (2003).

- 31 Petros, A. M. *et al.* Discovery of a Potent Inhibitor of the Antiapoptotic Protein Bcl-xL from NMR and Parallel Synthesis. *J. Med. Chem.* **49**, 656-663, doi:10.1021/jm0507532 (2006).
- 32 Ichihara, O., Barker, J., Law, R. J. & Whittaker, M. Compound Design by Fragment-Linking. *Molecular Informatics* **30**, 298-306, doi:<https://doi.org/10.1002/minf.201000174> (2011).
- 33 Frank, A. O. *et al.* Discovery of a potent inhibitor of replication protein a protein-protein interactions using a fragment-linking approach. *J. Med. Chem.* **56**, 9242-9250, doi:10.1021/jm401333u (2013).
- 34 Oltersdorf, T. *et al.* An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **435**, 677-681, doi:10.1038/nature03579 (2005).
- 35 Scott, D. E., Coyne, A. G., Hudson, S. A. & Abell, C. Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry* **51**, 4990-5003, doi:10.1021/bi3005126 (2012).
- 36 Hung, A. W. *et al.* Application of fragment growing and fragment linking to the discovery of inhibitors of Mycobacterium tuberculosis pantothenate synthetase. *Angew. Chem. Int. Ed. Engl.* **48**, 8452-8456, doi:10.1002/anie.200903821 (2009).
- 37 Sutherland, A. G. *et al.* Structure-based design of carboxybiphenylindole inhibitors of the ZipA-FtsZ interaction. *Organic & Biomolecular Chemistry* **1**, 4138-4140, doi:10.1039/B312016C (2003).
- 38 Nikiforov, P. O. *et al.* A fragment merging approach towards the development of small molecule inhibitors of Mycobacterium tuberculosis EthR for use as ethionamide boosters. *Organic & Biomolecular Chemistry* **14**, 2318-2326, doi:10.1039/C5OB02630J (2016).
- 39 Miyake, Y. *et al.* Identification of novel lysine demethylase 5-selective inhibitors by inhibitor-based fragment merging strategy. *Biorg. Med. Chem.* **27**, 1119-1129, doi:<https://doi.org/10.1016/j.bmc.2019.02.006> (2019).
- 40 Brough, P. A. *et al.* Combining Hit Identification Strategies: Fragment-Based and in Silico Approaches to Orally Active 2-Aminothieno[2,3-d]pyrimidine Inhibitors of the Hsp90 Molecular Chaperone. *J. Med. Chem.* **52**, 4794-4809, doi:10.1021/jm900357y (2009).
- 41 Rees, D. C., Congreve, M., Murray, C. W. & Carr, R. Fragment-based lead discovery. *Nature Reviews Drug Discovery* **3**, 660, doi:10.1038/nrd1467 (2004).
- 42 Baker, M. Fragment-based lead discovery grows up. *Nature Reviews Drug Discovery* **12**, 5-7, doi:10.1038/nrd3926 (2013).
- 43 Orita, M., Warizaya, M., Amano, Y., Ohno, K. & Niimi, T. Advances in fragment-based drug discovery platforms. *Expert Opinion on Drug Discovery* **4**, 1125-1144, doi:10.1517/17460440903317580 (2009).
- 44 Tsai, J. *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proceedings of the National Academy of Sciences* **105**, 3041-3046, doi:10.1073/pnas.0711741105 (2008).
- 45 Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467**, 596-599, doi:10.1038/nature09454 (2010).
- 46 Coyne, A. G., Scott, D. E. & Abell, C. Drugging challenging targets using fragment-based approaches. *Curr. Opin. Chem. Biol.* **14**, 299-307, doi:<https://doi.org/10.1016/j.cbpa.2010.02.010> (2010).

- 47 Ciulli, A. & Abell, C. Fragment-based approaches to enzyme inhibition. *Curr. Opin. Biotechnol.* **18**, 489-496, doi:10.1016/j.copbio.2007.09.003 (2007).
- 48 Hajduk, P. J. *et al.* Discovery of Potent Nonpeptide Inhibitors of Stromelysin Using SAR by NMR. *J. Am. Chem. Soc.* **119**, 5818-5827, doi:10.1021/ja9702778 (1997).
- 49 Magee, T. V. Progress in discovery of small-molecule modulators of protein–protein interactions via fragment screening. *Bioorg. Med. Chem. Lett.* **25**, 2461-2468, doi:<https://doi.org/10.1016/j.bmcl.2015.04.089> (2015).
- 50 Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* **21**, 1102-1114, doi:10.1016/j.chembiol.2014.09.001 (2014).
- 51 Zhu, Z. *et al.* Discovery of Cyclic Acylguanidines as Highly Potent and Selective  $\beta$ -Site Amyloid Cleaving Enzyme (BACE) Inhibitors: Part I—Inhibitor Design and Validation. *J. Med. Chem.* **53**, 951-965, doi:10.1021/jm901408p (2010).
- 52 Wang, Y. S. *et al.* Application of fragment-based NMR screening, X-ray crystallography, structure-based design, and focused chemical library design to identify novel microM leads for the development of nM BACE-1 (beta-site APP cleaving enzyme 1) inhibitors. *J. Med. Chem.* **53**, 942-950, doi:10.1021/jm901472u (2010).
- 53 Roberts, A. W. *et al.* Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **374**, 311-322, doi:10.1056/NEJMoa1513257 (2016).
- 54 Lin, M., Shapiro, M. J. & Wareing, J. R. Diffusion-Edited NMR–Affinity NMR for Direct Observation of Molecular Interactions. *J. Am. Chem. Soc.* **119**, 5249-5250, doi:10.1021/ja963654+ (1997).
- 55 Fejzo, J. *et al.* The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* **6**, 755-769, doi:[https://doi.org/10.1016/S1074-5521\(00\)80022-8](https://doi.org/10.1016/S1074-5521(00)80022-8) (1999).
- 56 Dalvit, C. *et al.* NMR-Based Screening with Competition Water–Ligand Observed via Gradient Spectroscopy Experiments: Detection of High-Affinity Ligands. *J. Med. Chem.* **45**, 2610-2614, doi:10.1021/jm011122k (2002).
- 57 Coan, K. E. D. & Shoichet, B. K. Stoichiometry and Physical Chemistry of Promiscuous Aggregate-Based Inhibitors. *Journal of the American Chemical Society* **130**, 9606-9612, doi:10.1021/ja802977h (2008).
- 58 Mosquera-Giraldo, L. I. & Taylor, L. S. Glass–Liquid Phase Separation in Highly Supersaturated Aqueous Solutions of Telaprevir. *Mol. Pharm.* **12**, 496-503, doi:10.1021/mp500573z (2015).
- 59 Frenkel, Y. V. *et al.* Concentration and pH Dependent Aggregation of Hydrophobic Drug Molecules and Relevance to Oral Bioavailability. *J. Med. Chem.* **48**, 1974-1983, doi:10.1021/jm049439i (2005).
- 60 Trasi, N. S. & Taylor, L. S. Thermodynamics of Highly Supersaturated Aqueous Solutions of Poorly Water-Soluble Drugs-Impact of a Second Drug on the Solution Phase Behavior and Implications for Combination Products. *J. Pharm. Sci.* **104**, 2583-2593, doi:10.1002/jps.24528 (2015).
- 61 Indulkar, A. S., Box, K. J., Taylor, R., Ruiz, R. & Taylor, L. S. pH-Dependent Liquid–Liquid Phase Separation of Highly Supersaturated Solutions of Weakly Basic Drugs. *Mol. Pharm.* **12**, 2365-2377, doi:10.1021/acs.molpharmaceut.5b00056 (2015).

- 62 Shoichet, B. K. Screening in a spirit haunted world. *Drug Discov. Today* **11**, 607-615, doi:10.1016/j.drudis.2006.05.014 (2006).
- 63 Aldrich, C. *et al.* The Ecstasy and Agony of Assay Interference Compounds. *ACS central science* **3**, 143-147, doi:10.1021/acscentsci.7b00069 (2017).
- 64 Feng, B. Y. *et al.* A High-Throughput Screen for Aggregation-Based Inhibition in a Large Compound Library. *J. Med. Chem.* **50**, 2385-2390, doi:10.1021/jm061317y (2007).
- 65 Shoichet, B. K. Interpreting Steep Dose-Response Curves in Early Inhibitor Discovery. *J. Med. Chem.* **49**, 7274-7277, doi:10.1021/jm061103g (2006).
- 66 McGovern, S. L., Helfand, B. T., Feng, B. & Shoichet, B. K. A Specific Mechanism of Nonspecific Inhibition. *J. Med. Chem.* **46**, 4265-4272, doi:10.1021/jm030266r (2003).
- 67 Giannetti, A. M., Koch, B. D. & Browner, M. F. Surface Plasmon Resonance Based Assay for the Detection and Characterization of Promiscuous Inhibitors. *J. Med. Chem.* **51**, 574-580, doi:10.1021/jm700952v (2008).
- 68 Coan, K. E. D., Maltby, D. A., Burlingame, A. L. & Shoichet, B. K. Promiscuous Aggregate-Based Inhibitors Promote Enzyme Unfolding. *J. Med. Chem.* **52**, 2067-2075, doi:10.1021/jm801605r (2009).
- 69 Sassano, M. F., Doak, A. K., Roth, B. L. & Shoichet, B. K. Colloidal Aggregation Causes Inhibition of G Protein-Coupled Receptors. *J. Med. Chem.* **56**, 2406-2414, doi:10.1021/jm301749y (2013).
- 70 Duan, D. *et al.* Internal Structure and Preferential Protein Binding of Colloidal Aggregates. *ACS Chem Biol* **12**, 282-290, doi:10.1021/acscchembio.6b00791 (2017).
- 71 Owen, S. C., Doak, A. K., Wassam, P., Shoichet, M. S. & Shoichet, B. K. Colloidal Aggregation Affects the Efficacy of Anticancer Drugs in Cell Culture. *ACS Chem Biol* **7**, 1429-1435, doi:10.1021/cb300189b (2012).
- 72 Owen, S. C. *et al.* Colloidal drug formulations can explain "bell-shaped" concentration-response curves. *ACS Chem Biol* **9**, 777-784, doi:10.1021/cb4007584 (2014).
- 73 Tres, F., Posada, M. M., Hall, S. D., Mohutsky, M. A. & Taylor, L. S. The Effect of Promiscuous Aggregation on in Vitro Drug Metabolism Assays. *Pharm. Res.* **36**, 170, doi:10.1007/s11095-019-2713-5 (2019).
- 74 Ryan, A. J., Gray, N. M., Lowe, P. N. & Chung, C.-w. Effect of Detergent on "Promiscuous" Inhibitors. *J. Med. Chem.* **46**, 3448-3451, doi:10.1021/jm0340896 (2003).
- 75 Feng, B. Y. & Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors. *Nature Protocols* **1**, 550-553, doi:10.1038/nprot.2006.77 (2006).
- 76 Wang, J. & Matayoshi, E. Solubility at the molecular level: development of a critical aggregation concentration (CAC) assay for estimating compound monomer solubility. *Pharm. Res.* **29**, 1745-1754, doi:10.1007/s11095-012-0730-8 (2012).
- 77 Bhattacharjee, S. DLS and zeta potential – What they are and what they are not? *J. Controlled Release* **235**, 337-351, doi:<https://doi.org/10.1016/j.jconrel.2016.06.017> (2016).
- 78 Hassan, P. A., Rana, S. & Verma, G. Making Sense of Brownian Motion: Colloid Characterization by Dynamic Light Scattering. *Langmuir* **31**, 3-12, doi:10.1021/la501789z (2015).

- 79 Panchal, J., Kotarek, J., Marszal, E. & Topp, E. M. Analyzing Subvisible Particles in Protein Drug Products: a Comparison of Dynamic Light Scattering (DLS) and Resonant Mass Measurement (RMM). *The AAPS Journal* **16**, 440-451, doi:10.1208/s12248-014-9579-6 (2014).
- 80 Geißler, D. *et al.* Effect of fluorescent staining on size measurements of polymeric nanoparticles using DLS and SAXS. *Analytical Methods* **7**, 9785-9790, doi:10.1039/C5AY02005K (2015).
- 81 Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Reviews* **8**, 409-427, doi:10.1007/s12551-016-0218-6 (2016).
- 82 Murugesan, J. R. *et al.* Revealing dye and dye-drug aggregation into nano-entities using NMR. *Dyes and Pigments* **153**, 300-306, doi:10.1016/j.dyepig.2018.02.026 (2018).
- 83 Ganesh, A. N., McLaughlin, C. K., Duan, D., Shoichet, B. K. & Shoichet, M. S. A New Spin on Antibody-Drug Conjugates: Trastuzumab-Fulvestrant Colloidal Drug Aggregates Target HER2-Positive Cells. *ACS Appl Mater Interfaces* **9**, 12195-12202, doi:10.1021/acsami.6b15987 (2017).
- 84 LaPlante, S. R. *et al.* Monitoring drug self-aggregation and potential for promiscuity in off-target in vitro pharmacology screens by a practical NMR strategy. *J. Med. Chem.* **56**, 7073-7083, doi:10.1021/jm4008714 (2013).
- 85 LaPlante, S. R. *et al.* Compound aggregation in drug discovery: implementing a practical NMR assay for medicinal chemists. *J. Med. Chem.* **56**, 5142-5150, doi:10.1021/jm400535b (2013).
- 86 Fielding, L. NMR Methods for the Determination of Protein- Ligand Dissociation Constants. *Curr. Top. Med. Chem.* **3**, 39-53, doi:<http://dx.doi.org/10.2174/1568026033392705> (2003).
- 87 Pellecchia, M., Sem, D. S. & Wüthrich, K. Nmr in drug discovery. *Nature Reviews Drug Discovery* **1**, 211, doi:10.1038/nrd748 (2002).
- 88 Boehr, D. D., Dyson, H. J. & Wright, P. E. An NMR Perspective on Enzyme Dynamics. *Chem. Rev.* **106**, 3055-3079, doi:10.1021/cr050312q (2006).
- 89 Campos-Olivas, R. NMR screening and hit validation in fragment based drug discovery. *Curr. Top. Med. Chem.* **11**, 43-67, doi:10.2174/156802611793611887 (2011).
- 90 Harner, M. J., Frank, A. O. & Fesik, S. W. Fragment-based drug discovery using NMR spectroscopy. *J. Biomol. NMR* **56**, 65-75, doi:10.1007/s10858-013-9740-z (2013).
- 91 Navratilova, I. & Hopkins, A. L. Fragment Screening by Surface Plasmon Resonance. *ACS Medicinal Chemistry Letters* **1**, 44-48, doi:10.1021/ml900002k (2010).
- 92 Shepherd, C. A., Hopkins, A. L. & Navratilova, I. Fragment screening by SPR and advanced application to GPCRs. *Prog. Biophys. Mol. Biol.* **116**, 113-123, doi:<https://doi.org/10.1016/j.pbiomolbio.2014.09.008> (2014).
- 93 Chilingaryan, Z., Yin, Z. & Oakley, A. J. Fragment-based screening by protein crystallography: successes and pitfalls. *Int J Mol Sci* **13**, 12857-12879, doi:10.3390/ijms131012857 (2012).
- 94 Caliandro, R., Belviso, D. B., Aresta, B. M., de Candia, M. & Altomare, C. D. Protein crystallography and fragment-based drug design. *Future Med Chem* **5**, 1121-1140, doi:10.4155/fmc.13.84 (2013).

- 95 Recht, M. I., Nienaber, V. & Torres, F. E. in *Methods Enzymol.* Vol. 567 (ed Andrew L. Feig) 47-69 (Academic Press, 2016).
- 96 Silvestre, H. L., Blundell, T. L., Abell, C. & Ciulli, A. Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. *Proc Natl Acad Sci U S A* **110**, 12984-12989, doi:10.1073/pnas.1304045110 (2013).
- 97 Linke, P. *et al.* An Automated Microscale Thermophoresis Screening Approach for Fragment-Based Lead Discovery. *J. Biomol. Screen.* **21**, 414-421, doi:10.1177/10870571115618347 (2016).
- 98 Mercier, K. A., Shortridge, M. D. & Powers, R. A multi-step NMR screen for the identification and evaluation of chemical leads for drug discovery. *Comb Chem High Throughput Screen* **12**, 285-295, doi:10.2174/138620709787581738 (2009).
- 99 Hajduk, P. J. *et al.* High-throughput nuclear magnetic resonance-based screening. *J. Med. Chem.* **42**, 2315-2317, doi:10.1021/jm9901475 (1999).
- 100 Wider, G. NMR techniques used with very large biological macromolecules in solution. *Methods Enzymol.* **394**, 382-398, doi:10.1016/s0076-6879(05)94015-9 (2005).
- 101 Hajduk, P. J. *et al.* NMR-Based Screening of Proteins Containing <sup>13</sup>C-Labeled Methyl Groups. *J. Am. Chem. Soc.* **122**, 7898-7904, doi:10.1021/ja000350l (2000).
- 102 Philip, J. H. & David, J. B. Integration of NMR and High-Throughput Screening. *Combinatorial Chem. High Throughput Screening* **5**, 613-621, doi:<http://dx.doi.org/10.2174/1386207023329996> (2002).
- 103 Jahnke, W. & Widmer, H. Protein NMR in biomedical research. *Cellular and Molecular Life Sciences CMLS* **61**, 580-599, doi:10.1007/s00018-003-3382-3 (2004).
- 104 Dalvit, C., Caronni, D., Mongelli, N., Veronesi, M. & Vulpetti, A. NMR-Based Quality Control Approach for the Identification of False Positives and False Negatives in High Throughput Screening. *Curr. Drug Disc. Technol.* **3**, 115-124, doi:<http://dx.doi.org/10.2174/157016306778108875> (2006).
- 105 Shortridge, M. D., Hage, D. S., Harbison, G. S. & Powers, R. Estimating Protein-Ligand Binding Affinity Using High-Throughput Screening by NMR. *J. Comb. Chem.* **10**, 948-958, doi:10.1021/cc800122m (2008).
- 106 Hann, M. *et al.* Strategic Pooling of Compounds for High-Throughput Screening. *Journal of Chemical Information and Computer Sciences* **39**, 897-902, doi:10.1021/ci990423o (1999).
- 107 Lepre, C. A. Library design for NMR-based screening. *Drug Discovery Today* **6**, 133-140, doi:[https://doi.org/10.1016/S1359-6446\(00\)01616-0](https://doi.org/10.1016/S1359-6446(00)01616-0) (2001).
- 108 Sugiki, T., Furuita, K., Fujiwara, T. & Kojima, C. Current NMR Techniques for Structure-Based Drug Discovery. *Molecules* **23**, 148 (2018).
- 109 Viegas, A., Macedo, A. L. & Cabrita, E. J. in *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols* (ed Ana Cecília A. Roque) 81-100 (Humana Press, 2010).
- 110 Claudio, D. & Anna, V. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **62**, 2218-2244, doi:10.1021/acs.jmedchem.8b01210 (2019).

- 111 Troelsen, N. S. *et al.* The 3F Library: Fluorinated Fsp3-Rich Fragments for Expeditious <sup>19</sup>F NMR Based Screening. *Angew. Chem. Int. Ed.* **59**, 2204-2210, doi:<https://doi.org/10.1002/anie.201913125> (2020).
- 112 Hofmann, M. H., Gerlach, D., Misale, S., Petronczki, M. & Kraut, N. Expanding the Reach of Precision Oncology by Drugging All KRAS Mutants. *Cancer Discovery* **12**, 924-937, doi:10.1158/2159-8290.Cd-21-1331 (2022).
- 113 Moore, A. R., Rosenberg, S. C., McCormick, F. & Malek, S. RAS-targeted therapies: is the undruggable drugged? *Nature Reviews Drug Discovery* **19**, 533-552, doi:10.1038/s41573-020-0068-6 (2020).
- 114 Herdeis, L., Gerlach, D., McConnell, D. B. & Kessler, D. Stopping the beating heart of cancer: KRAS reviewed. *Curr. Opin. Struct. Biol.* **71**, 136-147, doi:<https://doi.org/10.1016/j.sbi.2021.06.013> (2021).
- 115 Lanman, B. A. *et al.* Discovery of a Covalent Inhibitor of KRASG12C (AMG 510) for the Treatment of Solid Tumors. *J. Med. Chem.* **63**, 52-65, doi:10.1021/acs.jmedchem.9b01180 (2020).
- 116 Maurer, T. *et al.* Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proceedings of the National Academy of Sciences* **109**, 5299-5304, doi:doi:10.1073/pnas.1116510109 (2012).
- 117 Sun, Q. *et al.* Discovery of Small Molecules that Bind to K-Ras and Inhibit Sos-Mediated Activation. *Angew. Chem. Int. Ed.* **51**, 6140-6143, doi:<https://doi.org/10.1002/anie.201201358> (2012).
- 118 Kessler, D. *et al.* Drugging an undruggable pocket on KRAS. *Proceedings of the National Academy of Sciences* **116**, 15823-15829, doi:doi:10.1073/pnas.1904529116 (2019).
- 119 Harvey, J. J. An Unidentified Virus which causes the Rapid Production of Tumours in Mice. *Nature* **204**, 1104-1105, doi:10.1038/2041104b0 (1964).
- 120 Johnson, D. E. *et al.* Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers* **6**, 92, doi:10.1038/s41572-020-00224-3 (2020).
- 121 Saginala, K. *et al.* Epidemiology of Bladder Cancer. *Med Sci (Basel)* **8**, doi:10.3390/medsci8010015 (2020).
- 122 Hobbs, G. A., Der, C. J. & Rossman, K. L. RAS isoforms and mutations in cancer at a glance. *J. Cell Sci.* **129**, 1287-1292, doi:10.1242/jcs.182873 (2016).
- 123 Tengel, T. *et al.* Use of <sup>19</sup>F NMR spectroscopy to screen chemical libraries for ligands that bind to proteins. *Organic & Biomolecular Chemistry* **2**, 725-731, doi:10.1039/B313166A (2004).
- 124 Harris, R. K. *et al.* Further conventions for NMR shielding and chemical shifts IUPAC recommendations 2008. *Solid State Nucl. Magn. Reson.* **33**, 41-56, doi:<https://doi.org/10.1016/j.ssnmr.2008.02.004> (2008).
- 125 Markley, J. L. *et al.* Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. *J. Biomol. NMR* **12**, 1-23, doi:10.1023/a:1008290618449 (1998).
- 126 Rigger, R. *et al.* Certified Reference Material for Use in <sup>1</sup>H, <sup>31</sup>P, and <sup>19</sup>F Quantitative NMR, Ensuring Traceability to the International System of Units. *J. AOAC Int.* **100**, 1365-1375, doi:10.5740/jaoacint.17-0093 (2019).

- 127 Pina, A. S., Hussain, A. & Roque, A. C. A. in *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols* (ed Ana Cecília A. Roque) 3-12 (Humana Press, 2010).
- 128 Swinney, D. C. in *Phenotypic Drug Discovery* 1-19 (The Royal Society of Chemistry, 2021).
- 129 Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nature Reviews Drug Discovery* **10**, 507-519, doi:10.1038/nrd3480 (2011).
- 130 Schnur, D. M. Recent trends in library design: 'rational design' revisited. *Curr Opin Drug Discov Devel* **11**, 375-380 (2008).
- 131 Eggert, U. S. The why and how of phenotypic small-molecule screens. *Nature Chemical Biology* **9**, 206, doi:10.1038/nchembio.1206 (2013).
- 132 Parker, C. G. *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527-541.e529, doi:<https://doi.org/10.1016/j.cell.2016.12.029> (2017).
- 133 Schulze, J. *et al.* CellFy: A Cell-Based Fragment Screen against C-Type Lectins. *ACS Chem Biol* **13**, 3229-3235, doi:10.1021/acscchembio.8b00875 (2018).
- 134 Shao, L. *et al.* In vivo phenotypic drug discovery: applying a behavioral assay to the discovery and optimization of novel antipsychotic agents. *MedChemComm* **7**, 1093-1101, doi:10.1039/C6MD00128A (2016).
- 135 Houghten, R. A. *et al.* Direct Phenotypic Screening in Mice: Identification of Individual, Novel Antinociceptive Compounds from a Library of 734 821 Pyrrolidine Bis-piperazines. *ACS Combinatorial Science* **18**, 51-64, doi:10.1021/acscombsci.5b00126 (2016).
- 136 Pieper, A. A. *et al.* Discovery of a Proneurogenic, Neuroprotective Chemical. *Cell* **142**, 39-51, doi:10.1016/j.cell.2010.06.018 (2010).
- 137 Zimmermann, G. R., Lehár, J. & Keith, C. T. Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discovery Today* **12**, 34-42, doi:<https://doi.org/10.1016/j.drudis.2006.11.008> (2007).
- 138 Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **4**, 682-690, doi:10.1038/nchembio.118 (2008).
- 139 Gilbert, I. H. Drug discovery for neglected diseases: molecular target-based and phenotypic approaches. *J. Med. Chem.* **56**, 7719-7726, doi:10.1021/jm400362b (2013).
- 140 Balaña-Fouce, R., Pérez Pertejo, M. Y., Domínguez-Asenjo, B., Gutiérrez-Corbo, C. & Reguera, R. M. Walking a tightrope: drug discovery in visceral leishmaniasis. *Drug Discovery Today* **24**, 1209-1216, doi:<https://doi.org/10.1016/j.drudis.2019.03.007> (2019).
- 141 Alves, F. *et al.* Recent Development of Visceral Leishmaniasis Treatments: Successes, Pitfalls, and Perspectives. *Clin. Microbiol. Rev.* **31**, e00048-00018, doi:doi:10.1128/CMR.00048-18 (2018).
- 142 Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **44**, 235-249, doi:[https://doi.org/10.1016/S1056-8719\(00\)00107-6](https://doi.org/10.1016/S1056-8719(00)00107-6) (2000).
- 143 Barry, C. E. *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature Reviews Microbiology* **7**, 845-855, doi:10.1038/nrmicro2236 (2009).



- 144 Campbell, J. High-throughput assessment of bacterial growth inhibition by optical density measurements. *Curr Protoc Chem Biol* **3**, 100115, doi:10.1002/9780470559277.ch100115 (2011).
- 145 Pink, R., Hudson, A., Mouriès, M.-A. & Bendig, M. Opportunities and Challenges in Antiparasitic Drug Discovery. *Nature Reviews Drug Discovery* **4**, 727-740, doi:10.1038/nrd1824 (2005).
- 146 Global Forum for Health, R. & World Health, O. (World Health Organization, Geneva, 2004).
- 147 Mendes, V. & Blundell, T. L. Targeting tuberculosis using structure-guided fragment-based drug design. *Drug Discovery Today* **22**, 546-554, doi:<https://doi.org/10.1016/j.drudis.2016.10.003> (2017).
- 148 Brennan, P. J. & Nikaido, H. THE ENVELOPE OF MYCOBACTERIA. *Annual Review of Biochemistry* **64**, 29-63, doi:10.1146/annurev.bi.64.070195.000333 (1995).
- 149 Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nature Reviews Drug Discovery* **6**, 29-40, doi:10.1038/nrd2201 (2007).
- 150 Gwynn, M. N., Portnoy, A., Rittenhouse, S. F. & Payne, D. J. Challenges of antibacterial discovery revisited. *Annals of the New York Academy of Sciences* **1213**, 5-19, doi:<https://doi.org/10.1111/j.1749-6632.2010.05828.x> (2010).
- 151 Gopal, P. & Dick, T. Reactive dirty fragments: implications for tuberculosis drug discovery. *Current Opinion in Microbiology* **21**, 7-12, doi:<https://doi.org/10.1016/j.mib.2014.06.015> (2014).
- 152 Jackson, M., McNeil, M. R. & Brennan, P. J. Progress in targeting cell envelope biogenesis in Mycobacterium tuberculosis. *Future Microbiology* **8**, 855-875, doi:10.2217/fmb.13.52 (2013).
- 153 E. Barry, C. Lessons from Seven Decades of Antituberculosis Drug Discovery. *Curr. Top. Med. Chem.* **11**, 1216-1225, doi:<http://dx.doi.org/10.2174/156802611795429158> (2011).
- 154 Burza, S., Croft, S. L. & Boelaert, M. Leishmaniasis. *The Lancet* **392**, 951-970, doi:10.1016/S0140-6736(18)31204-2 (2018).
- 155 Roatt, B. M. *et al.* Recent advances and new strategies on leishmaniasis treatment. *Appl. Microbiol. Biotechnol.* **104**, 8965-8977, doi:10.1007/s00253-020-10856-w (2020).
- 156 Rich, S. M. *et al.* The origin of malignant malaria. *Proceedings of the National Academy of Sciences* **106**, 14902-14907, doi:doi:10.1073/pnas.0907740106 (2009).
- 157 Zavala, F. RTS,S: the first malaria vaccine. *The Journal of Clinical Investigation* **132**, doi:10.1172/JCI156588 (2022).
- 158 Varo, R., Chaccour, C. & Bassat, Q. Update on malaria. *Medicina Clínica*, doi:10.1016/j.medcli.2020.05.010.
- 159 Plewes, K., Leopold, S. J., Kingston, H. W. F. & Dondorp, A. M. Malaria: What's New in the Management of Malaria? *Infectious Disease Clinics of North America* **33**, 39-60, doi:<https://doi.org/10.1016/j.idc.2018.10.002> (2019).
- 160 Houben, R. M. & Dodd, P. J. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med* **13**, e1002152, doi:10.1371/journal.pmed.1002152 (2016).

- 161 Dheda, K., Barry, C. E., 3rd & Maartens, G. Tuberculosis. *The Lancet* **387**, 1211-1226, doi:10.1016/S0140-6736(15)00151-8 (2016).
- 162 Gagneux, S. Ecology and evolution of Mycobacterium tuberculosis. *Nature Reviews Microbiology* **16**, 202-213, doi:10.1038/nrmicro.2018.8 (2018).
- 163 Hollingshead, S. & Tang, C. M. in *Neisseria meningitidis: Methods and Protocols* (eds Kate L. Seib & Ian R. Peak) 1-16 (Springer New York, 2019).
- 164 Dretler, A. W., Roupael, N. G. & Stephens, D. S. Progress toward the global control of Neisseria meningitidis: 21st century vaccines, current guidelines, and challenges for future vaccine development. *Hum Vaccin Immunother* **14**, 1146-1160, doi:10.1080/21645515.2018.1451810 (2018).
- 165 Bekkat-Berkani, R. *et al.* Public health perspective of a pentavalent meningococcal vaccine combining antigens of MenACWY-CRM and 4CMenB. *J. Infect.* **85**, 481-491, doi:10.1016/j.jinf.2022.09.001 (2022).
- 166 Martínón-Torres, F. *et al.* Evolving strategies for meningococcal vaccination in Europe: Overview and key determinants for current and future considerations. *Pathogens and Global Health* **116**, 85-98, doi:10.1080/20477724.2021.1972663 (2022).
- 167 Stanaway, J. D. *et al.* The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *The Lancet Infectious Diseases* **16**, 712-723, doi:[https://doi.org/10.1016/S1473-3099\(16\)00026-8](https://doi.org/10.1016/S1473-3099(16)00026-8) (2016).
- 168 Ahmed, A. M. *et al.* Prevalence and burden of dengue infection in Europe: A systematic review and meta-analysis. *Rev. Med. Virol.* **30**, e2093, doi:<https://doi.org/10.1002/rmv.2093> (2020).
- 169 Norshidah, H., Vignesh, R. & Lai, N. S. Updates on Dengue Vaccine and Antiviral: Where Are We Heading? *Molecules* **26**, 6768 (2021).
- 170 Zhou, K. *et al.* Current Progress in the Development of Zika Virus Vaccines. *Vaccines* **9**, 1004 (2021).
- 171 Bernatchez, J. A. *et al.* Drugs for the Treatment of Zika Virus Infection. *J. Med. Chem.* **63**, 470-489, doi:10.1021/acs.jmedchem.9b00775 (2020).
- 172 Duffy, M. R. *et al.* Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med* **360**, 2536-2543, doi:10.1056/NEJMoa0805715 (2009).
- 173 Ganesh, A. N. *et al.* Colloidal Drug Aggregate Stability in High Serum Conditions and Pharmacokinetic Consequence. *ACS Chem Biol* **14**, 751-757, doi:10.1021/acschembio.9b00032 (2019).
- 174 Roche, O. *et al.* Development of a Virtual Screening Method for Identification of "Frequent Hitters" in Compound Libraries. *J. Med. Chem.* **45**, 137-142, doi:10.1021/jm010934d (2002).
- 175 Irwin, J. J. *et al.* An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **58**, 7076-7087, doi:10.1021/acs.jmedchem.5b01105 (2015).
- 176 Yang, Z.-Y. *et al.* Structural Analysis and Identification of Colloidal Aggregators in Drug Discovery. *Journal of Chemical Information and Modeling* **59**, 3714-3726, doi:10.1021/acs.jcim.9b00541 (2019).

- 177 Rao, H. *et al.* Identification of small molecule aggregators from large compound libraries by support vector machines. *J. Comput. Chem.* **31**, 752-763, doi:<https://doi.org/10.1002/jcc.21347> (2010).
- 178 McGovern, S. L. & Shoichet, B. K. Kinase Inhibitors: Not Just for Kinases Anymore. *J. Med. Chem.* **46**, 1478-1483, doi:10.1021/jm020427b (2003).
- 179 Seidler, J., McGovern, S. L., Doman, T. N. & Shoichet, B. K. Identification and Prediction of Promiscuous Aggregating Inhibitors among Known Drugs. *J. Med. Chem.* **46**, 4477-4486, doi:10.1021/jm030191r (2003).
- 180 Doak, A. K., Wille, H., Prusiner, S. B. & Shoichet, B. K. Colloid formation by drugs in simulated intestinal fluid. *J. Med. Chem.* **53**, 4259-4265, doi:10.1021/jm100254w (2010).
- 181 Volovik Frenkel, Y., Gallicchio, E., Das, K., Levy, R. M. & Arnold, E. Molecular Dynamics Study of Non-nucleoside Reverse Transcriptase Inhibitor 4-[[4-[[4-[(E)-2-Cyanoethenyl]-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino]benzotrile (TMC278/Rilpivirine) Aggregates: Correlation between Amphiphilic Properties of the Drug and Oral Bioavailability. *J. Med. Chem.* **52**, 5896-5905, doi:10.1021/jm900282z (2009).
- 182 Ganesh, A. N. *et al.* Leveraging Colloidal Aggregation for Drug-Rich Nanoparticle Formulations. *Mol. Pharm.* **14**, 1852-1860, doi:10.1021/acs.molpharmaceut.6b01015 (2017).
- 183 McLaughlin, C. K. *et al.* Stable Colloidal Drug Aggregates Catch and Release Active Enzymes. *ACS Chem Biol* **11**, 992-1000, doi:10.1021/acscchembio.5b00806 (2016).
- 184 Torosyan, H. & Shoichet, B. K. Protein Stability Effects in Aggregate-Based Enzyme Inhibition. *J. Med. Chem.* **62**, 9593-9599, doi:10.1021/acs.jmedchem.9b01019 (2019).
- 185 Blevitt, J. M. *et al.* Structural Basis of Small-Molecule Aggregate Induced Inhibition of a Protein-Protein Interaction. *J. Med. Chem.* **60**, 3511-3517, doi:10.1021/acs.jmedchem.6b01836 (2017).
- 186 Feng, B. Y. *et al.* Small-molecule aggregates inhibit amyloid polymerization. *Nat. Chem. Biol.* **4**, 197-199, doi:10.1038/nchembio.65 (2008).
- 187 Piller, C. Blots on a field? *Science* **377**, 358-363, doi:10.1126/science.add9993 (2022).
- 188 Hameed P, S. *et al.* Aminoazabenzimidazoles, a Novel Class of Orally Active Antimalarial Agents. *J. Med. Chem.* **57**, 5702-5713, doi:10.1021/jm500535j (2014).
- 189 Blaazer, A. R. *et al.* Fragment-Based Screening in Tandem with Phenotypic Screening Provides Novel Antiparasitic Hits. *J. Biomol. Screen.* **20**, 131-140, doi:10.1177/1087057114549735 (2015).
- 190 Okada, T. *et al.* X-Ray Diffraction Analysis of Three-Dimensional Crystals of Bovine Rhodopsin Obtained from Mixed Micelles. *Journal of Structural Biology* **130**, 73-80, doi:<https://doi.org/10.1006/jsbi.1999.4209> (2000).
- 191 Cherezov, V. *et al.* High-Resolution Crystal Structure of an Engineered Human  $\beta$ 2-Adrenergic G Protein-Coupled Receptor. *Science* **318**, 1258-1265, doi:doi:10.1126/science.1150577 (2007).
- 192 Rosenbaum, D. M. *et al.* GPCR Engineering Yields High-Resolution Structural Insights into  $\beta$ 2-Adrenergic Receptor Function. *Science* **318**, 1266-1273, doi:doi:10.1126/science.1150609 (2007).

- 193 Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* **16**, 531-543, doi:10.1038/nrd.2017.111 (2017).
- 194 Fetz, V., Prochnow, H., Brönstrup, M. & Sasse, F. Target identification by image analysis. *Natural Product Reports* **33**, 655-667, doi:10.1039/C5NP00113G (2016).
- 195 Schirle, M. & Jenkins, J. L. Identifying compound efficacy targets in phenotypic drug discovery. *Drug Discovery Today* **21**, 82-89, doi:<https://doi.org/10.1016/j.drudis.2015.08.001> (2016).
- 196 Moore, J. D. The impact of CRISPR–Cas9 on target identification and validation. *Drug Discovery Today* **20**, 450-457, doi:<https://doi.org/10.1016/j.drudis.2014.12.016> (2015).
- 197 Lee, H. & Lee, J. W. Target identification for biologically active small molecules using chemical biology approaches. *Arch. Pharmacol Res.* **39**, 1193-1201, doi:10.1007/s12272-016-0791-z (2016).
- 198 Lee, J. & Bogoy, M. Target deconvolution techniques in modern phenotypic profiling. *Curr. Opin. Chem. Biol.* **17**, 118-126, doi:<https://doi.org/10.1016/j.cbpa.2012.12.022> (2013).
- 199 Vincent, F. *et al.* Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nature Reviews Drug Discovery*, doi:10.1038/s41573-022-00472-w (2022).
- 200 Belema, M. & Meanwell, N. A. Discovery of Daclatasvir, a Pan-Genotypic Hepatitis C Virus NS5A Replication Complex Inhibitor with Potent Clinical Effect. *J. Med. Chem.* **57**, 5057-5071, doi:10.1021/jm500335h (2014).
- 201 Mestres, J., Gregori-Puigjané, E., Valverde, S. & Solé, R. V. The topology of drug–target interaction networks: implicit dependence on drug properties and target families. *Molecular BioSystems* **5**, 1051-1057, doi:10.1039/B905821B (2009).
- 202 Kell, D. B. Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening and knowledge of transporters: where drug discovery went wrong and how to fix it. *The FEBS Journal* **280**, 5957-5980, doi:<https://doi.org/10.1111/febs.12268> (2013).
- 203 Munos, B. A Forensic Analysis of Drug Targets From 2000 Through 2012. *Clinical Pharmacology & Therapeutics* **94**, 407-411, doi:<https://doi.org/10.1038/clpt.2013.126> (2013).
- 204 Rogawski, M. A., Tofighy, A., White, H. S., Matagne, A. & Wolff, C. Current understanding of the mechanism of action of the antiepileptic drug lacosamide. *Epilepsy Res* **110**, 189-205, doi:10.1016/j.eplepsyres.2014.11.021 (2015).
- 205 Labau, J. I. R. *et al.* Lacosamide Inhibition of Na(V)1.7 Channels Depends on its Interaction With the Voltage Sensor Domain and the Channel Pore. *Front Pharmacol* **12**, 791740, doi:10.3389/fphar.2021.791740 (2021).
- 206 Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305-309, doi:10.1126/science.1244917 (2014).
- 207 Prior, M. *et al.* Back to the future with phenotypic screening. *ACS Chem Neurosci* **5**, 503-513, doi:10.1021/cn500051h (2014).
- 208 Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nature Reviews Drug Discovery* **5**, 993-996, doi:10.1038/nrd2199 (2006).

- 209 Lipinski, C. A. & Reaume, A. G. Phenotypic screening of low molecular weight compounds is rich ground for repurposed, on-target drugs. *Frontiers in Pharmacology* **13**, doi:10.3389/fphar.2022.917968 (2022).
- 210 Castoreno, A. B. & Eggert, U. S. Small molecule probes of cellular pathways and networks. *ACS Chem Biol* **6**, 86-94, doi:10.1021/cb1002976 (2011).
- 211 Mayer, T. U. *et al.* Small Molecule Inhibitor of Mitotic Spindle Bipolarity Identified in a Phenotype-Based Screen. *Science* **286**, 971-974, doi:doi:10.1126/science.286.5441.971 (1999).
- 212 Moffat, J. G., Rudolph, J. & Bailey, D. Phenotypic screening in cancer drug discovery — past, present and future. *Nature Reviews Drug Discovery* **13**, 588-602, doi:10.1038/nrd4366 (2014).