Practical Aspects of Designing the 2nd Generation BIONET Premium Fragment Library

Steven R. LaPlante[1], Patrick McCarren[2], Francois Bilodeau[1], Andrew Lowerson[3], Michael Serrano-Wu[2]

[1]NMX Research and Solutions, Laval, Quebec, Canada; University of Quebec, INRS-IAF, Laval, Quebec, Canada.
[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[3]Key Organics, Camelford, Cornwall, United Kingdom

Introduction

Fragment libraries are commonly assembled by Rule of 3 filtering followed by manual curation, however robust experimental data that ensures the proper physicochemical attributes needed for high-concentration screening is often lacking and replaced instead by in silico calculations of uncertain predictive value. A fragment collection with experimentally-determined aqueous solubility will address a major source of false positives and attrition in fragment screening libraries: aggregation, stability, and solubility. ^{1}H NMR spectral data in aqueous buffer will further enable practitioners to rapidly build fragment pools and initiate screening. Diversity selection methods in shape, scaffold, fingerprint, and predicted property space combined with industry-standard substructure filtering were used to select 1981 Key Organics fragments for experimental profiling. ^{1}H NMR analysis allowed the careful selection of highly-soluble fragments with desirable physicochemical and stability characteristics. All fragments are soluble in DMSO (200mM). Importantly, the curated molecules are enriched in cyclic scaffolds commonly found in drug candidates and span chemical space that minimally overlaps with existing commercial collections.

Cheminformatic curation for fragment prioritization and library characterization.

We began with the virtual assessment of the Key Organics Fragment Collection (> 20K fragments) that would be physically available in large quantities then filtering this to a fragment-like space. The "Rule of Three" was used to bias towards fragments likely to be fragment hits[1]. Further, a design element of the Astex screening library was adopted that further reduces the size of fragments ≤ 16 heavy atoms, above which much lower hit rates were observed[2]. All properties were calculated using Pipeline Pilot[3].

The parameters used in the design of Fragment Library:

- Heavy atoms ≤16
- logP ≤ 3,
- Hydrogen bond donors ≤3
- Hydrogen bond acceptors ≤3
- Polar surface area ≤ 60
- Rotatable bonds ≤ 3

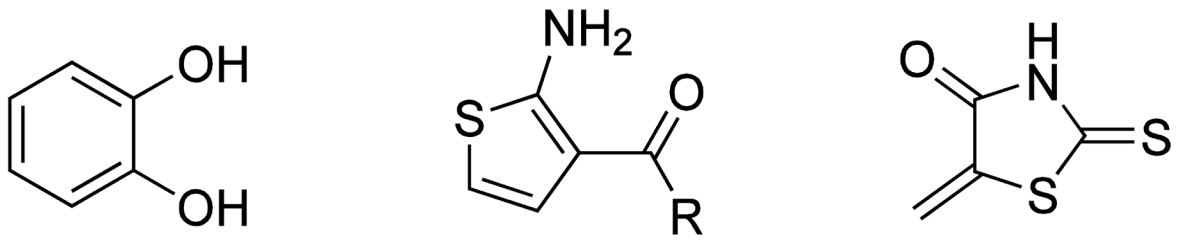
As part of our Fragment selection process, industry-standard substructure filtering - including PAINS filtering - was implemented and as a result the BIONET 2nd Generation Premium Fragment Library does not include substructures identified as promiscuous or reactive by the following empirically determined rejection rules:

- Lilly MedChem Rules[4]
- PAINS[5]
- BMS[6]

Focus on Pan Assay Interference Compounds (PAINS) substructure filtering – a deciding factor in the quality of a fragment library.

PAINS are compounds that frequently show up as screening hits, but that act through non-specific mechanisms such as covalent attachment to proteins or generation of hydrogen peroxide. The problem with PAINS is that they may show convincing biochemical and even cell based activity, but mechanistically be useless for further advancement to drugs or even chemical probes. PAINS remain common in many vendors Fragment Libraries. PAINS fragments have been identified and substructure filters constructed that recognise these fragments.

Figure 1: Example of PAINS



Finally, medicinal chemists with fragment screening experience visually inspected and evaluated the fragments for liabilities and general unattractiveness and suspect fragments were removed.

Diversity in structure and calculated 2D and 3D property space, as well as incorporation of BioCores[7], were criteria for selection to be tested experimentally (Figure 1). BioCores are defined as a heteroaromatic ring connected by a carbon linker to a saturated heterocyclic molecule. Incorporation of BioCores into the scaffolds of a molecular library enhances their natural/drug like properties. This pattern is frequently observed in natural products but not frequently in marketed drugs or patents, increasing their potential for patentability. BioCores are underrepresented in commercially available fragment libraries.

An iterative algorithm was used that scored points for fragments satisfying an unexplored scaffold, 2D fingerprints, property bin, combinations of property bins, or drug substructure. For 2D properties, the number of atoms, molecular weight, calculated logP, rotatable bonds; number of hydrogen bond donors and acceptors, and polar surface area were used. For 3D shape, the normalized PMI ratios were used based on the lowest energy conformation[8]. ECFP4 fingerprints and Bemis-Murcko scaffolds[9] were used to describe the 2D structure. For each iteration, the best scoring example of a scaffold was selected in which BioCores were chosen first. The components of the score were weighted to select fragments that were substructures of marketed drugs and satisfied new 3D space first, followed by the number of new 2D fingerprints, followed by the coverage of 2D property bins:

1. Group by scaffolds
2. Visit scaffolds in the following order:
   I. Prefer by BioCore
   II. Prefer by having a drug fragment
3. For each scaffold in sequence, look for fingerprints and properties not seen before or deficient
4. Visit scaffolds again until reaching 3 members
5. Verify final

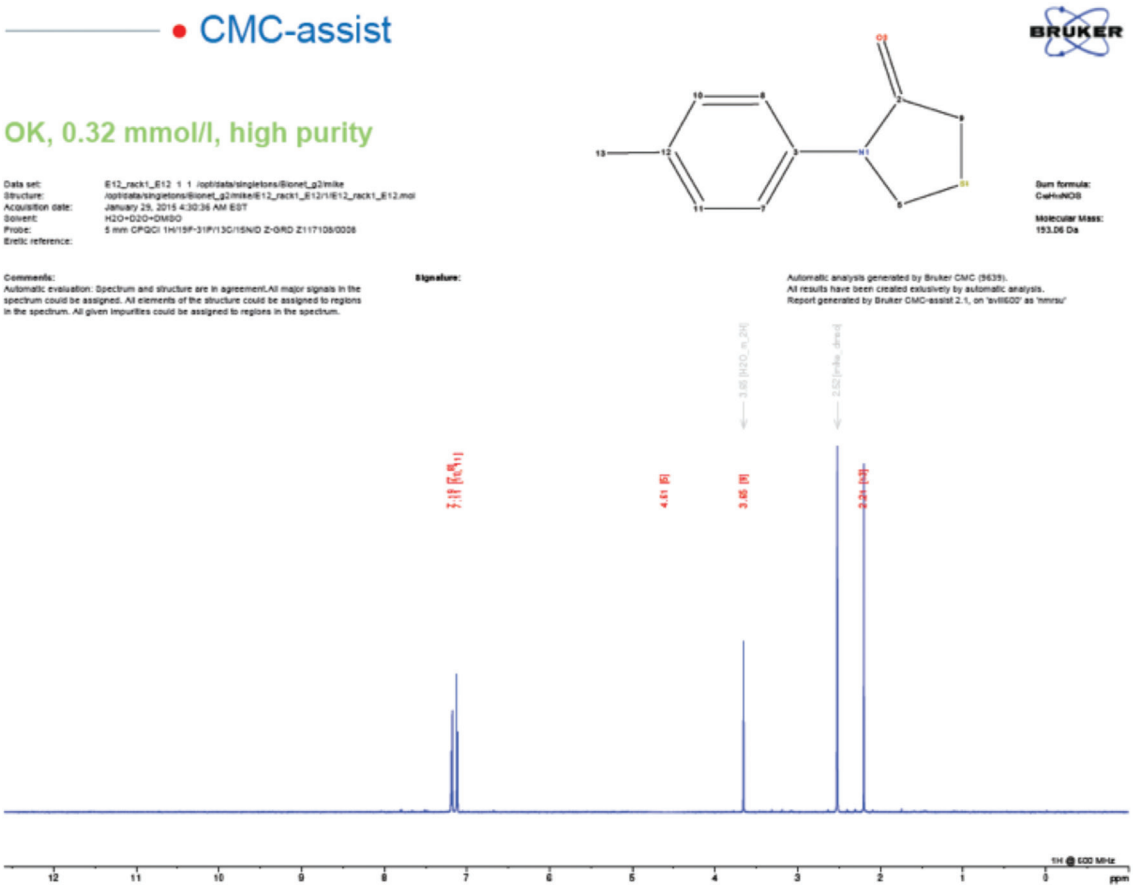
Approximately 1981 fragments were then physically acquired for experimental curation. After experimental curation, the final set was investigated again for diversity using closest similarities using ECFP_4 fingerprints and MACCS 166-bit keys. Highly similar fragments were removed with the exception of those that represented different geometric pharmacophore patterns.

^{1}H NMR curation for fragment prioritization and library characterization.

The 1981 fragments selected from the cheminformatics analyses were then subjected to experimental curation to prioritize fragments that would be amenable to rigorous biophysical analysis in physiologically-relevant aqueous solution. For example, fragments that were highly soluble in water and DMSO (200mM) were favored given that most fragment screens and assays require high concentrations. Other fragments were reprioritized due to self-aggregation, poor solubility or instability - given that these properties render then unamenable to screening and have notoriously been associated with a range of undesirable attributes.

The experimental curation method of choice was ^{1}H NMR spectroscopy given that simple ^{1}H NMR spectra in water allowed for the rapid evaluation of structural integrity, purity, solubility, stability, aggregation tendencies and chemical shift positions (see Figure 2). For this, NMR samples were made for each of the 1981 top contenders. Each singleton sample consisted of 300 µM fragment in buffer (50 mM sodium phosphate pH 7.4, 100 mM NaCl). ^{1}H NMR spectra were acquired at 600 MHz spectrometer equipped with a helium cryoprobe that significantly increased signal-to-noise. Simple 1D ^{1}H NMR spectra were acquired along with 1D ^{1}H CPMG for size-filtering aggregation analyses.

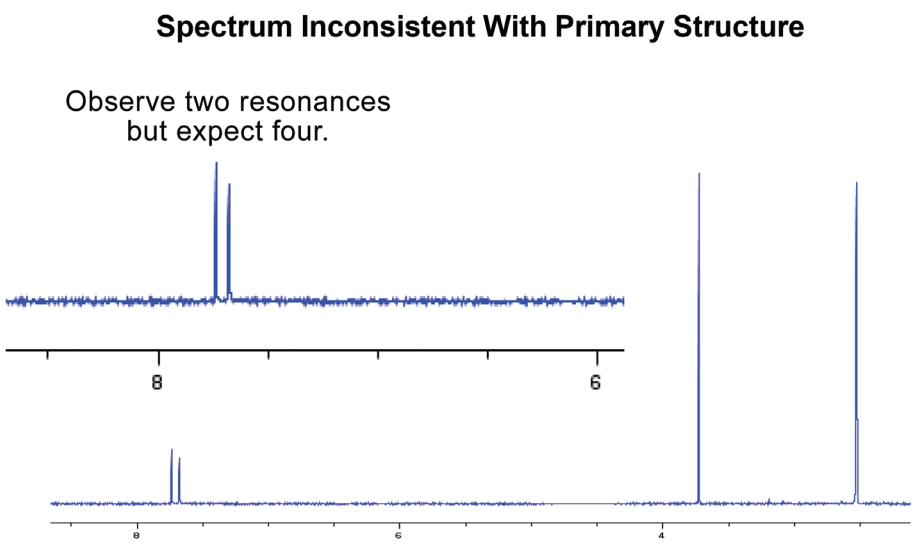
Figure 2: Example report provided by the CMC Assist automation module.



Data analyses involved combinations of manual and automation tools. The CMC Assist automation software purchased from Bruker Spectrospin Inc. had multiple practical uses. It allowed for an automatic readout of the fragment concentration that was experimentally derived from integrating the NMR resonances of each singleton sample, and referencing to standardized samples using the Erectic module (Bruker Spectrospin Inc.)[10]. The CMC Assist module also allowed for verification of each singleton spectrum to determine if the spectral attributes were consistent with the proposed primary structure of the corresponding fragment. This exercise was also complemented by an automated analyses using Spectral DB software (ACD Inc.). Finally, manual visual inspections were also performed by an NMR spectroscopist and a medicinal chemist. The latter was critical for removing fragments that were perceived to have known potential chemical liabilities.

Based on these interactive analyses, 300 fragments were subsequently removed from the library. Some examples are shown on the following pages.

Figure 3: ^{1}H NMR spectrum is inconsistent with the primary structure.



Figures 4 and 5 illustrate how the NMR spectra of two fragments reveal degradation and hence justified removal of these fragments from the library.

Figure 4: Fragment degradation revealed by the observation of an additional NMR resonance.

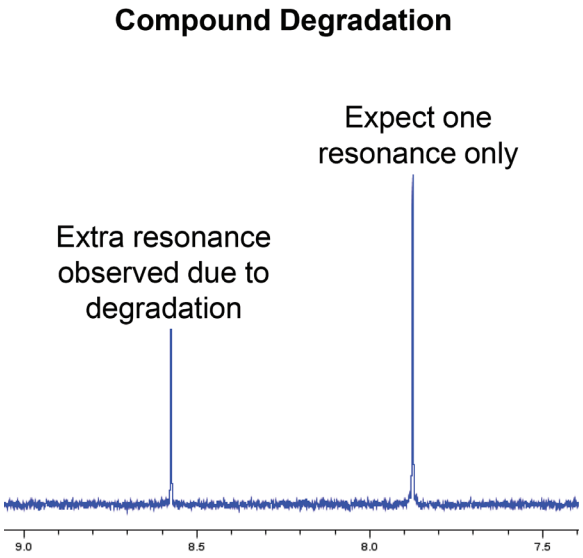
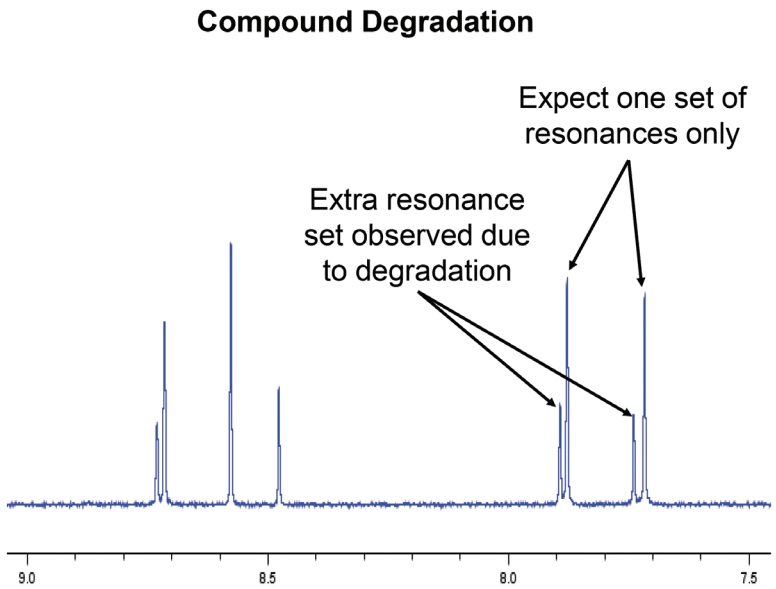
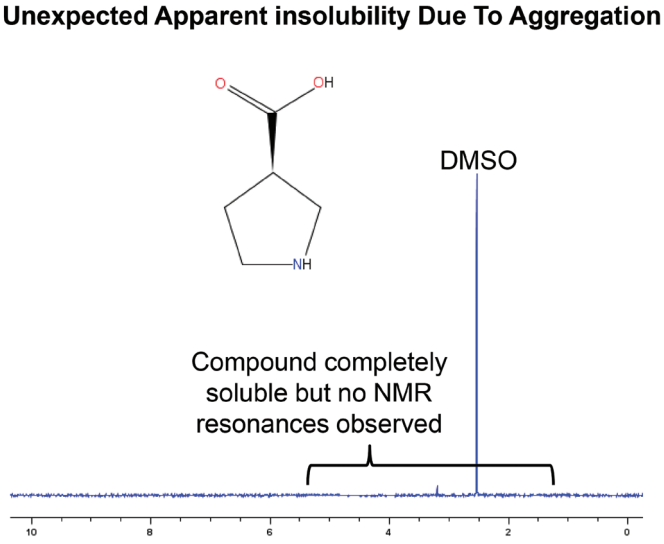


Figure 5: Fragment degradation revealed by the observation of an extra set of NMR resonances.



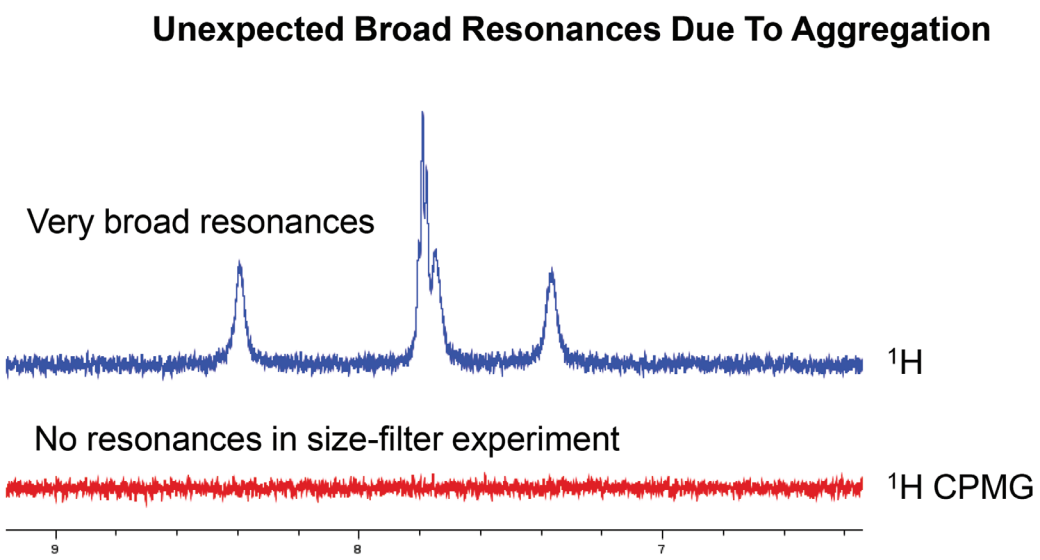
Many fragments were removed due to insolubility. Evidence was easily noted by the observation of precipitation and the absence of NMR resonances. However, there were also samples where no precipitate was notable and no resonances were observed. This is illustrated in Figure 6 for a fragment that has a carboxylic acid moiety which should have imparted solubility. This fragment most likely formed soluble large aggregates and was therefore removed from the library.

Figure 6: Example of a sample that lacked precipitation and NMR resonances.



Other samples were assessed to have aggregation properties and removed from the library. Figure 7 showed that the ^{1}H NMR spectrum (blue) had very broad resonances which were unexpected for a small molecule in solution, but rather indicative of a slow-tumbling aggregate. These resonances were completely lacking in the experimental CPMG spectrum (red), which serves as a size-filter experiment to eliminate resonance of slow-tumbling entities.

Figure 7:
Example of a sample that had unexpected broad resonances due to aggregation (blue spectrum). A CPMG experiment is shown in the red spectrum.



In all, NMR-based triaging resulted in the removal of fragments from the library due to, (1) insufficient solubility, (2) chemical degradation, (3) potential structural liabilities, (4) poor resolution of NMR resonances, (5) signs of aggregation.

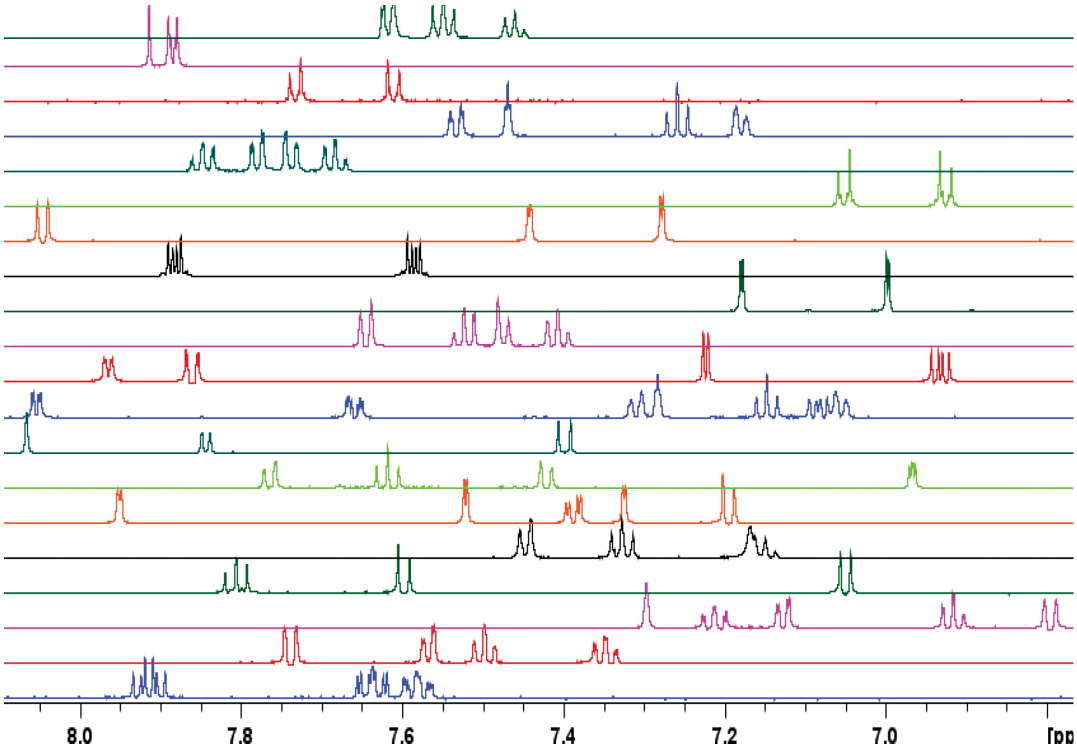
The remaining 1681 fragments were then evaluated with regards to the desired solubility range that would be appropriate for biophysical studies. Thus, it was decided that only fragments having an Erectic-derived concentration of ≥ 200 µM would be selected and promoted to the next rounds of evaluation. This resulted in the removal of another 399 fragments leaving a remaining 1282 fragments.

Given the stringent filtering steps described above, a final round of cheminformatics analyses was undertaken to properly evaluate the remaining 1282 fragments with regards to maximizing library diversity and distinction from other commercial libraries. This led to the removal of 112 fragments based on a cheminformatics similarity analyses (considered not sufficiently diverse from related fragments in the library) and 4 were removed due to suspected undesirable properties. In the end, this finally resulted in a library of 1166 well-behaved soluble fragments that are amenable for biophysical studies and screens that require high concentrations.

Figure 8 shows the superposition of the ^{1}H NMR spectra of 20 fragments from passed all of the curation filters. A visual inspection of the spectra reveals that the fragments have sharp resonances as expected for fast-tumbling lone molecules in solution. The spectra also show that the fragments are pure, soluble and stable in aqueous buffer.

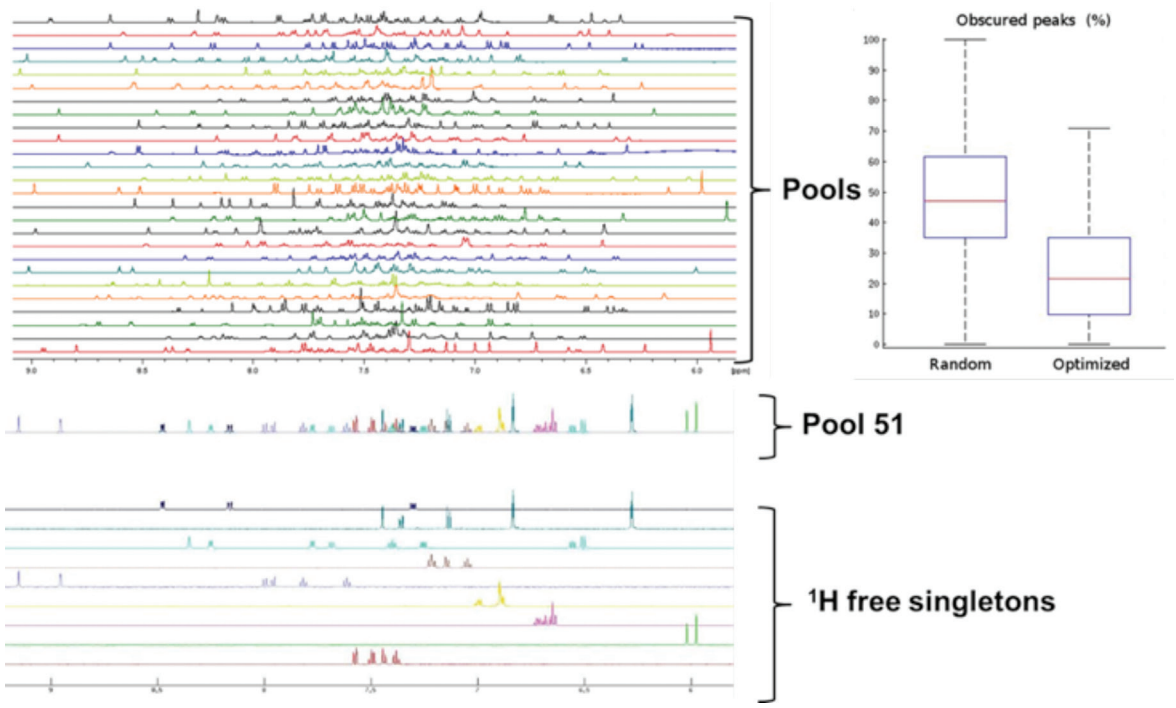
The ^{1}H NMR spectra of each singleton fragment in buffer is available upon request. Note that the NMR files also include analyses data derived from the CMC Assist module which automatically picked resonance peaks and tabulated the data within the spectral files. These can then be used for chemical shift encoding purposes.

Figure 8: Superposition of the 600 MHz ^{1}H NMR spectra of a subset of individual fragments.



Access to this distribution of resonance shifts then allows for the possibility of "smart pooling" of fragments that effectively can minimize the overlap of resonances. Fragment pooling is a common strategy employed for NMR-based fragment screening as a means to minimize the use of expensive NMR time and amount of target protein. Minimizing the overlap of resonance via smart pooling helps with the subsequent deconvolution steps (i.e. to identify the singleton binder) when a binding hit is identified from screening pools of fragments. An example of how the NMR data acquired here was used to pool fragments is shown in Figure 9.

Figure 9: Intelligent Pooling Developed.



Salient statistics of the final library.

In all, the Key Organics 'BIONET' Fragment collection was filtered to a much smaller subset of candidate fragments utilizing Ro3 and industry standard substructure filtering including removal of PAINS. For example, the selected subset strictly meets "Astex Rule of 3" including polar surface area ≤ 60 and number of atoms ≤ 16. A visualization of the final 2nd Generation BIONET Fragment Library properties and diversity is provided in Figure 10, 11 and 12 on the following pages.

After experimental NMR-based curation, the final library consisted of 1166 fragments that had ^{1}H-NMR-derived solubilities of at least 200 µM in standard aqueous buffer. The library included 445 fragments found in approved drugs and >1100 sharing cores found in drugs. It also included 37 "BioCore" containing patterns. It was interesting that 937 unique fragments were not found in commercially available fragment collections. There was low overlap between any one particular commercial libraries.

Figure 10: Overview of calculated properties of the final library

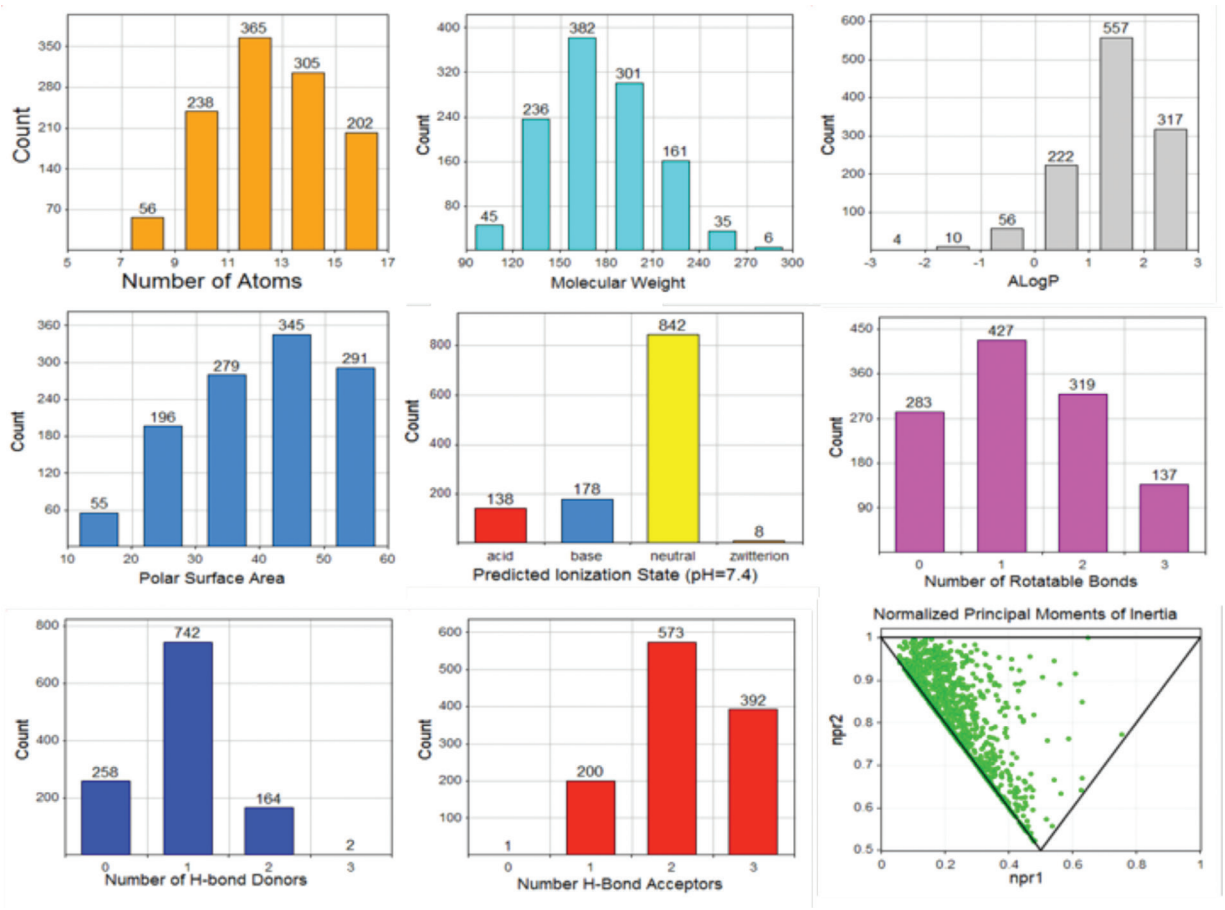


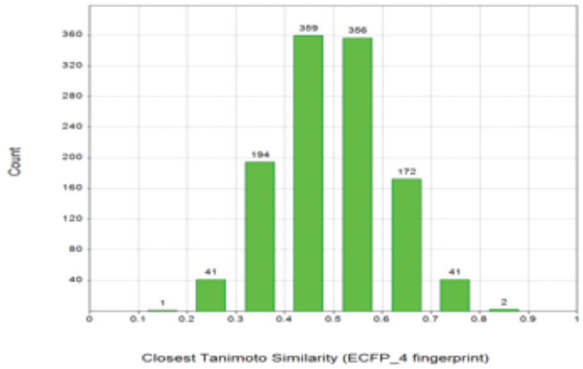
Figure 11: Overview of the range of properties of the final library

Property	Average	Min	Max
Molecular Weight	177	95	290
Heavy Atom Count	12.1	7	16
Polar Surface Area	40.4	12	60
cLogP	1.4	-2.8	3.0
# H-bond acceptors	2.1	0	3
# H-bond donors	0.9	0	3
# rotatable bonds	1.3	0	3

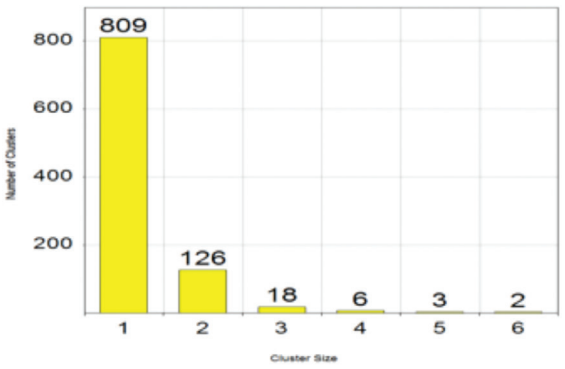
- Diversity coefficient (average distance using MACCS 166-bit keys) = 0.72
- # clusters at 0.85 Tanimoto similarity (MACCS 166-bit MOE) = 964 clusters / 1166 fragments 83%

Figure 12: Diversity statistics of the final library.

Closest similarities using ECFP_4



MACCS 166-bit cluster size, min Tanimoto 0.85



Summary

A 2nd generation BIONET Premium Fragment Library has been constructed employing Rule of Three and industry standard substructure filtering including PAINS analysis. Diversity selection utilized methods in shape, scaffold, fingerprint and predicted property space. All 1166 fragments in the 2nd Generation Premium Fragment Library have been analyzed by ^{1}H NMR for:

- Structure verification
- Purity
- Solubility
- Lack of aggregation.

Physiochemical Properties of the library:

- Heavy atoms ≤16
- logP ≤ 3,
- Hydrogen bond donors ≤3
- Hydrogen bond acceptors ≤3
- Polar surface area ≤ 60
- Rotatable bonds ≤ 3

Substructure Filtering:

- Lilly MedChem Rules[4]
- PAINS[5]
- BMS[6]

Diversity Statistics:

- Diversity coefficient (average distance using MACCS 166-bit keys) = 0.72
- # clusters at 0.85 Tanimoto similarity (MACCS 166-bit MOE) = 964 clusters / 1166 fragments 83%

DMSO Solubility:

All 1166 Fragments are soluble in DMSO at 200mM

Data Provision:

All customers will be supplied with the following data package for each aqueous soluble fragment purchased:

- Aqueous buffer ^{1}H NMR pdf
- Aqueous buffer ^{1}H NMR raw data file
- ^{1}H NMR chemical shifts supplied in an excel file

References

1. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. *Drug Discov. Today* **2003**, 8 (19), 876–877.
2. Hall, R. J.; Mortenson, P. N.; Murray, C. W. *Prog. Biophys. Mol. Biol.* **2014**, 116 (2-3), 82–91.
3. Pipeline Pilot v 8.5; note H-bond acceptors and donors used the default Accelrys Num_H_ Acceptors and Num_H_Donors which are different from the Lipinski definition.
4. Bruns, R. F.; Watson, I. A. *J. Med. Chem.* **2012**, 55 (22), 9763–9772.
5. Baell, J. B.; Holloway, G. A. *J. Med. Chem.* 2**010**, 53 (7), 2719–2740.
6. Pearce, B. C.; Sofia, M. J.; Good, A. C.; Drexler, D. M.; Stock, D. A. J. *Chem. Inf. Model.* **2006**, 46 (3), 1060–1068.
7. Kombarov, R.; Altieri, A.; Genis, D.; Kirpichenok, M.; Kochubey, V.; Rakitina, N.; Titarenko, Z. *Mol. Divers.* **2010**, 14 (1), 193–200.
8. Conformations generated in MOE 2014 using the MMFFs force field.
9. Bemis-Murcko scaffolds were calculated as in the defaults for the Pipeline Pilot Generate Fragments, which keeps atom type.
10. Akoka, S.; Barantin, L.; Trierweiler, M. Anal. Chem. **1999**, 71, 2554-2557.

Key Organics

Chemistry | Innovation | Quality

BIONET

Fragment Libraries

For more information please contact us at:

T: +44 (0)1840 212137 F: +44 (0)1840 213712 E: fragments@keyorganics.net www.keyorganics.net

Key Organics Ltd., Highfield Road Industrial Estate, Camelford, Cornwall PL32 9RA, United Kingdom