Analysis of the Sygnature Discovery Screening Set

Compound profile of 61 libraries delivered to the ELF

Christopher Pearce, Claire Cariou-Mumford, Megan Lightfoot, Kam Chohan, Daniel Hamza

Sygnature Discovery Ltd, BioCity, Pennyfoot Street, Nottingham, NG1 1GF, United Kingdom

The European Lead Factory

The European Lead Factory (ELF) is a project of the Innovative Medicines Initiative (IMI) and is aimed at boosting drug discovery in Europe through the generation of a state-of-the-art high-throughput screening (HTS) collection of 500,000 compounds. 300,000 of these compounds are being donated by pharmaceutical companies and 200,000 are being generated in collaboration between academic groups and small and medium enterprises (SMEs) with Sygnature Discovery committed to delivering 40,000 compounds by the end of 2017. Here, the properties of 61 libraries synthesised at Sygnature Discovery, below termed the Sygnature set, are discussed.

	Sygnature Set
Number of libraries	61
Number of compounds	34352
Average number of compounds per library	563
Average LC-MS purity (%)	98
Average amount (µmol)	22
Average mass (mg)	9.3

TABLE 1 QC and statistics of the final compounds in the Sygnature set.

Properties of the Sygnature Discovery set

The Sygnature set comprises 61 libraries with on average 563 compounds per library at 98% purity and in amounts of 22 µmol (Table 1). Figure 1 shows the distribution of some physicochemical properties known to influence the pharmacokinetics of small molecule drugs. The set conforms well to the Lipinski rule of five parameters for oral drugs with 72% of compounds having MW ≤ 450 Da, 96% cLogP ≤ 4, 99.9% TPSA ≤ 140 Å^2, 99% H-bond donors ≤ 3 and 83% H-bond acceptors ≤ 5. An analysis of charge shows 0.5% are acids, 45% bases, 54% neutrals and 0.5% zwitterions.

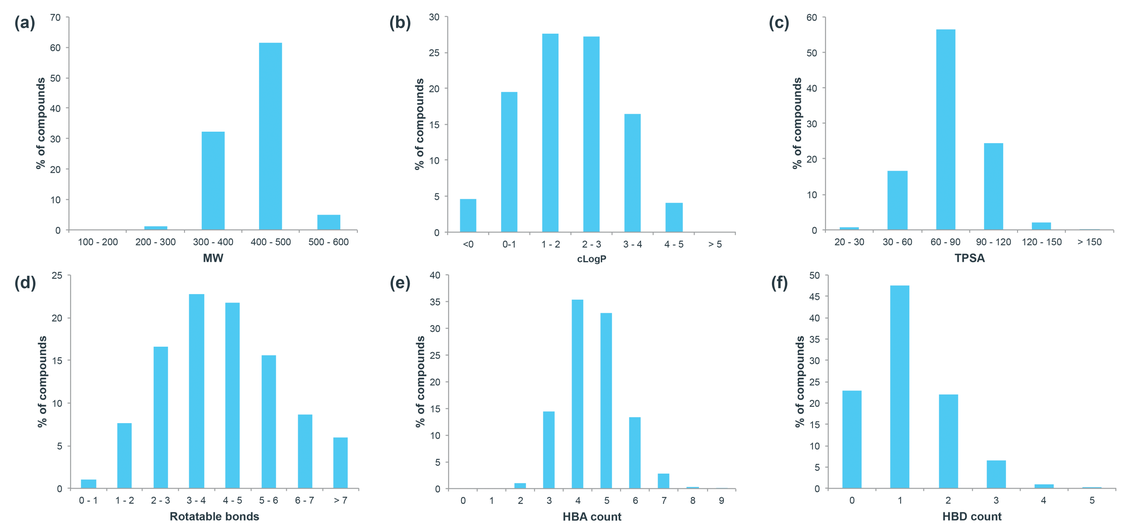

FIGURE 1 Physicochemical properties of the Sygnature set.[1] **(a)** Molecular weight (MW) **(b)** Calculated LogP (cLogP) **(c)** Topological polar surface area (TPSA) **(d)** Number of rotatable bonds **(e)** Number of hydrogen bond acceptors (HBA) **(f)** Number of hydrogen bond donors (HBD).

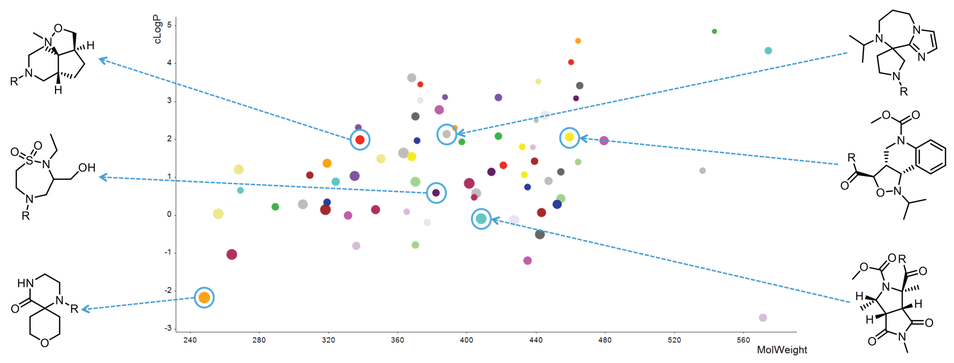

FIGURE 2 MW versus cLogP of the 71 clusters in the Sygnature set, coloured by cluster, sized by Fsp^3.

Built-in diversity

A benefit of the Sygnature set is that all libraries have been designed with lead-likeness in mind. The reagent sets of amines, carboxylic acids, sulfonyl chlorides etc. were carefully selected at the start of the project through a process of clustering to ensure diversity. These sets have been further refined by removing examples that frequently fail during synthesis thus improving the overall success rate of the libraries. The diversity is maintained through the purchase of new reagents. During the design of each library, reagents are also selected on strict MW and cLogP parameters for the final compounds. Any falling outside these limits are excluded from synthesis. Before full production of a library starts, a representative sample of final compounds is synthesised and tested for stability. Compound classes that show signs of instability are excluded from the library. Applying these criteria should deliver a high-quality screening set with lead-like properties. The Sygnature dataset was clustered using the chemical fingerprints and the default tree method within the program Molsoft L.L.C (v3.8-5). One cluster centre was selected from each cluster and a plot of molecular weight versus cLogP is shown in Figure 2.

Comparison with known drugs reveals high degree of sp^3 character

Properties of the Sygnature set were compared with those of known oral drugs. US Food and Drug Administration (FDA) approved compounds were selected as a comparator and a filter of MW > 100 and < 700 was applied to generate a list of 1,474 marketed drugs, here termed the Oral set.

A model to predict attrition in pre-clinical toxicology for basic oral drugs has shown shape terms such as sp^3 carbon count and flatness to be important. It showed that the rate of attrition due to toxicity was diminished by increasing carbon sp^3 count, whereas increased flatness or increased aromatic/aliphatic ratio promoted attrition.[2] This suggests that a higher Fraction of sp^3 (Fsp^3) is advantageous during development. The analysis in Figure 3 shows that there is indeed a notable difference in the degree of chirality and sp^3 character between the two sets. 66% of the Sygnature set has 2 or more chiral atoms compared to only 22% for the Oral set, and 62% of Sygnature set has an Fsp^3 of > 0.4 compared to 50% for the Oral set. This translates to a difference in molecular shape with the Sygnature set having a higher degree of 3D shape and chiral character.

A plot of MW versus cLogP (Figure 4) clearly shows that the Sygnature set falls within the boundaries of the Oral set and therefore known drug-like space. To examine the diversity of the Sygnature and Oral sets, a principal component analysis (PCA) was performed using chemical fingerprints (Figure 5). This shows that known drug space is well represented by a range of different libraries. In addition a significant proportion of compounds occupy new chemical space.

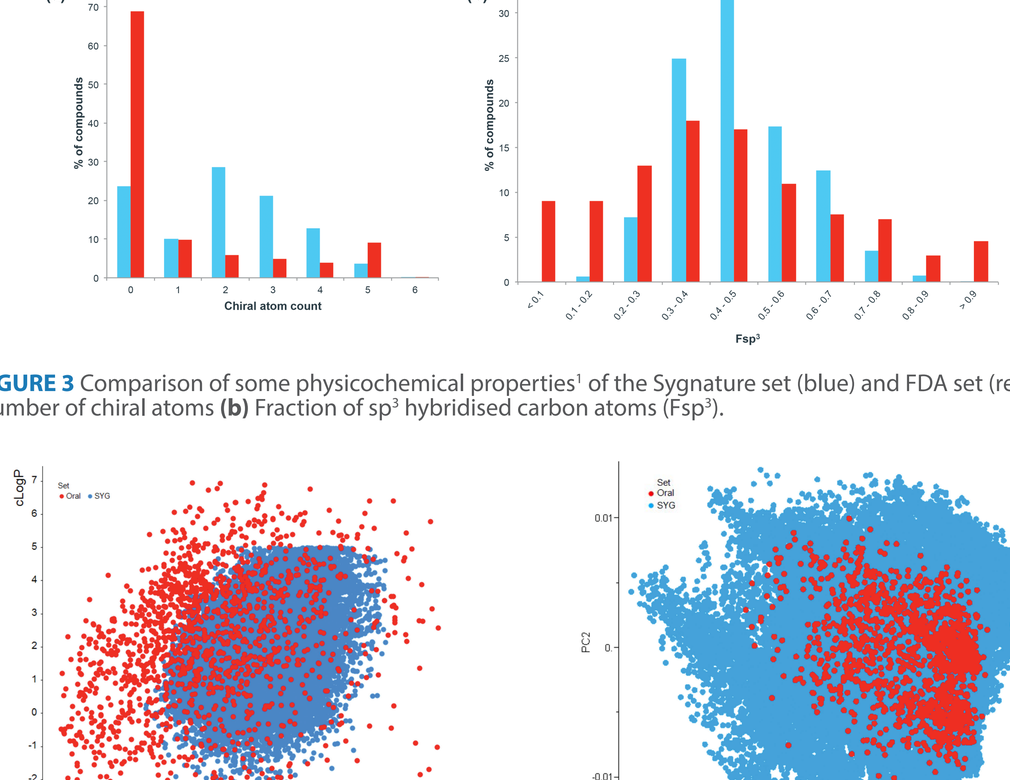

FIGURE 3 Comparison of some physicochemical properties[1] of the Sygnature set (blue) and FDA set (red). **(a)** Number of chiral atoms **(b)** Fraction of sp^3 hybridised carbon atoms (Fsp^3).

FIGURE 4 MW versus cLogP plot of Sygnature set (blue) and FDA set (red).

FIGURE 5 PCA of the Sygnature set (blue) and FDA set (red).[3]

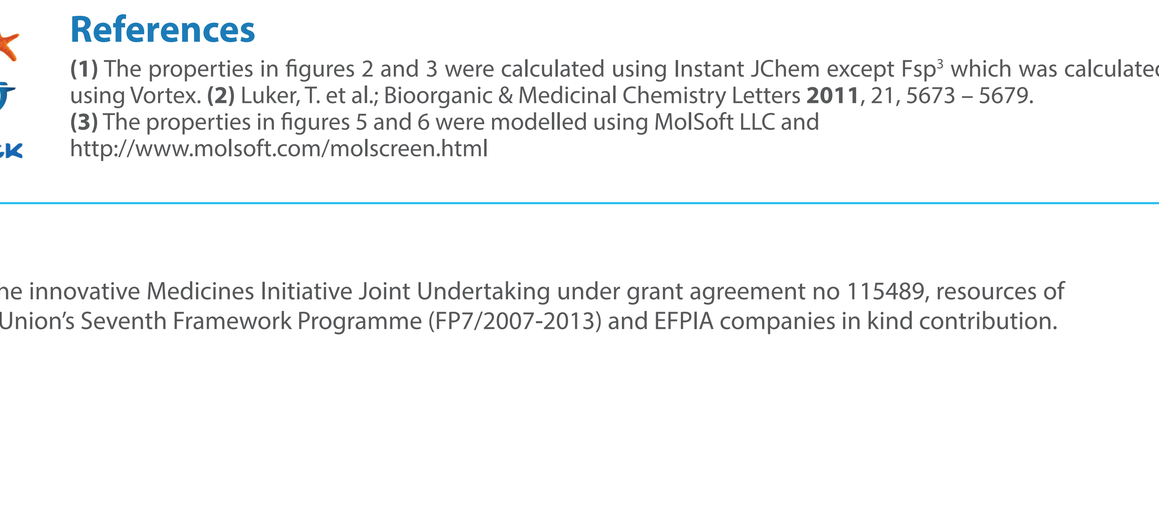

FIGURE 6 Some predicted drug properties of the Sygnature set.[3] **(a)** Drug-likeness **(b)** Solubility, Sygnature set (blue) FDA set (red) **(c)** PAINS liability **(d)** hERG liability **(e)** Permeability - PAMPA **(f)** PGP substrate.

Potential impact on medicinal chemistry

Common drug properties of the Sygnature set were examined by applying *in silico* models for drug-likeness, solubility, Pan Assay Interference Compounds (PAINS), hERG inhibition, PAMPA permeability and PGP substrate (Figure 6). Compounds predicted to be drug-like have a score > 0, which applies to 93% of the Sygnature set. This agrees well with previous observations that the set lies within rule of five parameters and within the MW and cLogP boundaries of known drugs. Solubility of the Sygnature set is similar to that of the FDA set. The remaining predicted properties suggest high compound quality with > 99% predicted not PAINS, > 99% hERG non-inhibitors, 98% permeable by PAMPA and 100% not PGP substrates. This gives confidence that these properties will translate into quality lead-like hits.

The ToxScore was also calculated and is based on the presence of known toxic groups, with a score ≥ 1 indicating likely toxicity based on substructure match. It is worth noting that ~20% of known drugs have a ToxScore > 1 and the Sygnature set compares well with 78% having a ToxScore of 0 and only 5% having a ToxScore > 1.

Conclusion

The 61 libraries synthesised for HTS by Sygnature Discovery have good calculated physical properties and obey well the rule of five. Compared to a set of FDA approved drugs the Sygnature set has a high degree of sp^3-rich, chiral compounds thus yielding a screening set with good 3D shape. The set conforms to MW versus cLogP limits for known drugs whilst exploring new chemical space with PCA analysis highlighting the diverse nature of these compounds. *In silico* modelling of drug properties gives confidence that these key characteristics will translate into high quality lead-like hits compared to classical screening collections. By keeping the final properties of compounds in mind when designing scaffolds and selecting reagents, a set of compounds with good predicted drug properties has been produced. The results of this analysis will enable further optimization of future libraries.

References

(1) The properties in figures 2 and 3 were calculated using Instant JChem except Fsp^3 which was calculated using Vortex. **(2)** Luker, T. et al.; Bioorganic & Medicinal Chemistry Letters **2011**, 21, 5673 – 5679.
(3) The properties in figures 5 and 6 were modelled using MolSoft LLC and http://www.molsoft.com/molscreen.html

The research leading to these results has received support from the innovative Medicines Initiative Joint Undertaking under grant agreement no 115489, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution.