The Selcia Fragment Library: Design, Analysis & Quality Control

Carol Austin[a], Simon Pettit[a], David Scowen[a], Jussi Kangasmetsa[a], Andrew Keats[a], Chris Swain[b], Lauren Freeman[c], Bob Carling[a], Ian Marshall[a], David O'Brien[a], Mike Jones[a], Martin Walker[a], Clive Cornell[a].

Contact: carol.austin@selcia.com

a. Discovery Division, Selcia Ltd., Fyfield Business and Research Park, Ongar, Essex, CM5 0GS, UK

b. Cambridge MedChem Consulting Ltd., http://www.cambridgemedchemconsulting.com, UK

c. University of Bath, Bath, BA2 7AY, UK

Introduction

The Selcia Fragment Library (SFL) was designed to be universal and compatible with our proprietary CEfrag™ screen, a novel fragment screening technique[1] based on capillary electrophoresis[2,3]. The first generation library was initially assembled using compounds from Selcia's compound collection and commercial compounds obtained from fragmentation of known biologically active compounds.

The compounds were selected based on 'rule-of-3' criteria[4], chemical attractiveness and potential for expandability (i.e. availability of near neighbours). Due to the high solution concentrations of compounds used during a fragment screen, it is important to ensure those compounds are soluble and of good quality. Hence, during the design of the SFL, close attention was made to quality control by calculating and measuring solubility and by performing detailed LC-MS and ^{1}H-NMR; those that did not meet the set criteria were excluded. The library has been supplemented with commercial and non-commercial compounds to increase the chemical diversity. Analysis has demonstrated that the majority of the compounds are singletons and that there is very little overlap between the Selcia Fragment Library and eight known commercially available fragment libraries.

Design of First Generation Selcia Fragment Library (SFL)

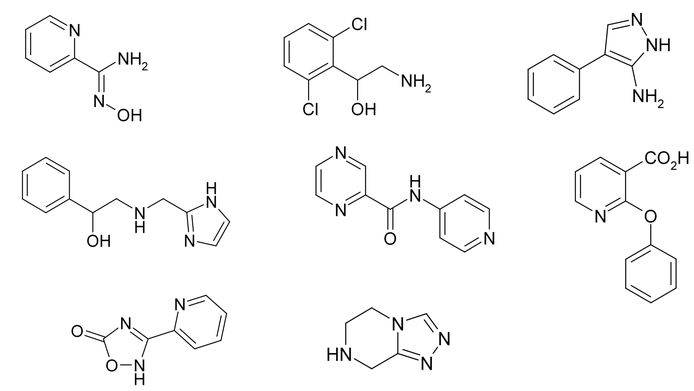
To obtain fragments for the first generation SFL, compounds from Selcia's internal compound collection were selected to meet the criteria outlined in Table 1. In addition, fragmentation of known biological actives was performed: fragments that were identified as being important in ligands binding to biological targets were identified using Binding DB (http://www.bindingdb.org), which is a publicly accessible database currently containing >20k experimentally determined binding affinities of protein ligand complexes, for 110 protein targets. Fragments were generated using Recap in MOE. Those fragments that were commercially available and met the 'rule-of-3'[4] and calculated solubility criteria (>1mM) were prioritised for inclusion.

To improve the chemical diversity, the library was supplemented with fragment classes poorly represented in the library. To improve the novelty of the compounds in the library, a custom synthesis programme of non-commercial compounds was also initiated; an example of some of the structures synthesised is shown in Figure 1.

Table 1: Library Design - Desired Properties of the Selcia Fragment Library

Characteristics	Criteria
Target classes	Universal
Molecular weight	< 300
cLogP	< 3
H-bond acceptors	< 5
H-bond donors	< 2
Calculated & measured solubility	> 1mM
Number of fragments	~1,500-2,500
Quality control	95%
Chemical tractability/expandability	Good
Overlap with commercial fragment libraries	Low

Figure 1: Examples of Compounds in the Selcia Fragment Library



Methods – Library Quality Control

1600 fragments, with calculated solubility >1mM (see below), were placed in tared tubes and weighed using a Bohdan weighing robot. The compounds were dissolved as 30mM stocks in DMSO and plated in 96-well plates using a Tecan liquid handling robot.

Calculated Solubility: The aqueous solubility of the parent compound (i.e. non-salt) was calculated using WSKOWWIN (part of EPI SuiteTM v4.0 available from http://www.epa.gov/opptintr/exposure/pubs/episuitedl.htm). Initially, a calculated solubility criteria of >1mM was used.

Measured Solubility: Solubility was measured using a turbidimetric method[5]. 30mM DMSO stocks of the fragments were diluted to 1mM in 35mM Hepes pH7.8 (3.3% final DMSO), in duplicate, in clear 96-well plates. The plates were sealed and shaken for 3h at 25°C. The absorbance at 650nm was measured using a SpectraMaxM5 (Molecular Devices). If the background subtracted A_{650nm} was ≥0.01 in both duplicates the compounds were scored as insoluble, those <0.01 were scored as soluble.

Reactive Compounds: All 1600 fragments were reviewed by Selcia's Medicinal Chemists and any 'reactive' compounds were removed (e.g. electrophiles).

LC-MS: LC-MS analysis of the library was performed on a Quattro LC (Micromass). 10µl of the compounds (3mM in DMSO) in 96-well plates was injected. Different mobile phases were used to analyse acidic and neutral compounds and UV (230nm), ELSD and MS (105-400 m/z) were measured. Purity was assessed by UV.

NMR Analysis: The library was analysed by ^{1}H-NMR using a Bruker *Avance* 500MHz system. The samples (5mM in DMSO with ~10% DMSO-d6) were transferred from 96-wellplate to the spectrometer using *BEST NMR* into a 120µl LC-Flow Probe (spectra acquired over 16 scans). The signals from DMSO and H_2O were suppressed using an automatically optimised WET solvent suppression sequence with ^{13}C decoupling. Spectra for each compound were visually assessed for compound integrity and % purity by two independent Medicinal Chemists.

Figure 2: Comparison of Calculated v Measured Solubility of Fragments

Calculated solubility (WSKOWWIN software, $Log_{10}[M]$) was plotted against the measured solubility ($Log_{10}[A_{650nm}]$). There was no significant correlation between calculated and measured solubility in aqueous buffer. This highlights the critical need to measure the solubility of a fragment library under the conditions used for the assay.

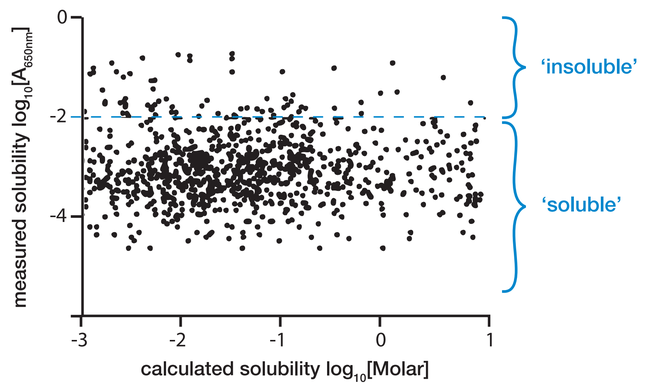


Table 2: Compounds that Failed Quality Criteria

The results of the quality control performed to select the final SFL are outlined below. Overall 24% of the library failed the QC, thus highlighting the importance of good quality control. These compounds have been removed from the library.

Criteria Tested	No. compounds	% Overall library
Potentially Reactive	63	3.9
Insoluble (measured)	52	3.3
<95% purity (LC/MS and ^{1}H-NMR)*	276	17.3
Total compounds that failed criteria	**386**	**24**

*~30% of those that failed QC were benzylamine fragments

Analysis of Library Characteristics

Those fragments that did not meet the quality criteria were culled from the library. This left a library of 1214 fragments which was analysed to determine its characteristics (i.e. clogP, MWt, polar surface area, H-bond donors, H-bond acceptors, heavy atom count and rotatable bond count). The analysis showed that the compounds generally conform to 'rule-of-3'[4] criteria (Figure 3).

Diversity analysis of the compounds demonstrated that the majority of the compounds were singletons (Figure 4).

Since there are several commercial fragment libraries available, we were interested in knowing whether our fragment library significantly differed to those commercially available. The structure of 8 commercially available fragment libraries was compared to the structures in the SFL. The greatest overlap was seen with the Maybridge library but this was <13% of the SFL (Table 3). Hence, there is very little overlap of the SFL with these commercially available fragment libraries.

Substructure searches, carried out using each of the fragments against a database of FDA approved drugs, demonstrated 20% of the SFL had substructures within known marketed drugs.

Figure 3: Selcia Fragment Library Characteristics

The characteristics (clogP, MWt, polar surface area, H-bond donors, H-bond acceptors, heavy atom count and rotatable bond count) of the 1214 SFL compounds (those that met the QC criteria) were calculated using ChemAxon tools, automated using applescript and shellscript. Graphs were created using Aabel.

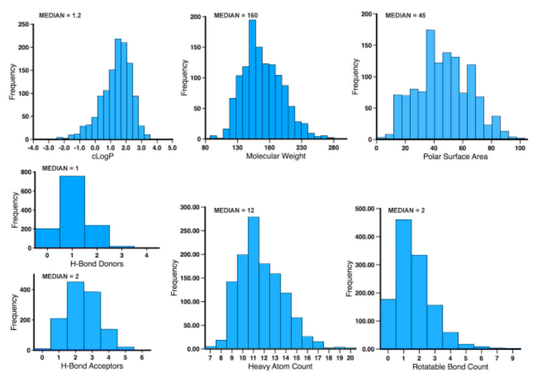


Figure 4: Diversity - Majority of Compounds are Singletons

Fingerprints based on MACCS descriptors were generated in MOE and subjected to cluster analysis using a Tanimoto coefficient 0.85. The number of compounds in each cluster was calculated and plotted using Aabel. There were no large clusters and the majority of compounds were singletons, indicating a diverse collection of fragments.

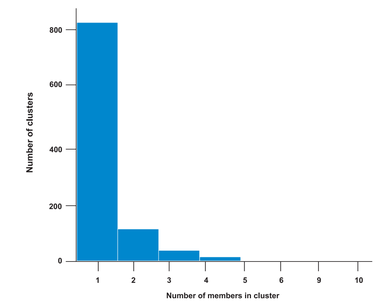


Table 3: There is Low Overlap of the Selcia Fragment Library with Commercially Available Fragment Libraries.

The SFL was compared with 8 commercially available fragment libraries. A SVL (Scientific Vector Language) script was used: db_nb_mols_incommon.svl (available from Chemical Computing Group), which calculated a matrix for the number of compounds each database has in common with each of the other databases based on structure (SMILES). Green indicates numbers of unique fragments in each library. Blue indicates numbers of Selcia compounds found in the other libraries. Red indicates % overlap of Selcia fragment library with indicated commercial fragment library.

Fragment Library	Selcia	Pyxis	Maybridge	Life Chemicals	Key Organics	InFarmatik	Enamine	Chembridge	Asinex	% of Selcia Library Overlap
Selcia	1214	0	148	0	49	6	17	21	2	100.0
Pyxis	0	312	0	0	0	0	0	0	0	0.0
Maybridge	148	0	1500	0	75	9	10	30	3	12.2
Life Chemicals	0	0	0	8087	5	0	3	8	8	0.0
Key Organics	49	0	75	5	6906	23	38	67	19	4.0
InFarmatik	6	0	9	0	23	1385	5	3	1	0.5
Enamine	17	0	10	3	38	5	1190	34	0	1.4
Chembridge	21	0	30	8	67	3	34	4570	1	1.7
Asinex	2	0	3	8	19	1	0	1	5142	0.2

Selcia Fragment Library Evolution

It is planned that the number of fragments in the SFL will be increased to ~1500 by Q1 2011. We aim to supplement the library with compounds representative of those previously shown to be active in published fragment screens. The novelty of the library will also increase with continuation of our custom synthesis program.

As our CEfrag™ screening programs progress, frequent hitters/problem compounds will be recognised and removed. Similarly, compounds failing longer stability QC (1 year in DMSO) will also be removed, thus further improving the quality of the Selcia Fragment Library.

Conclusion

The Selcia Fragment Library is designed for all target classes and contains diverse fragments that conform to 'rule-of-3' criteria[4]. The library samples have good solubility (>1mM, under assay conditions) and the purity has been experimentally verified for each compound (>95%).

The Selcia Fragment Library has little overlap with current commercially available fragment libraries. Selcia's proprietary CEfrag™ screen[1], together with the Selcia Fragment Library, represents a unique and powerful tool for fragment-based drug discovery.

References:
[1] Patent pending
[2] Pierceall *et al.* Affinity capillary electrophoresis analyses of protein-protein interactions in target-directed drug discovery. *Methods in Molecular Biology;* **261** (2004) 187-198
[3] Lewis *et al.* Affinity capillary electrophoresis for the screening of novel antimicrobial targets; *Journal of Biomolecular Screening;* **9** (2004) 303-308
[4] Congreve *et al.* A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today*; **8** (2003) 876-877
[5] Alsenz & Kansy. High throughput solubility measurement in drug discovery and development. *Advanced Drug Delivery Reviews*; **59** (2007) 546-567

Acknowledgements
Thanks to all Selcia's Medicinal Chemists who helped in the handling of compounds, review of compounds, ideas generation and analysis of NMR spectra