DESIGN AND SYNTHESIS OF THE COMPREHENSIVE FRAGMENT LIBRARY
A 3D ENABLED LIBRARY FOR MEDICINAL CHEMISTRY DISCOVERY

Warren S Wade[1], Todd Meyer[1], Peter Pallai[1], Paolo Tosco[2], James Zapf[3], Laura Lingardo[3], Gordon Alton[3]

[1]BioBlocks, Inc., [2]Cresset, [3]Visionary Pharmaceuticals

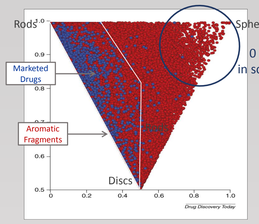
BioBlocks
A CHEMISTRY PARTNER IN DRUG DISCOVERY

INTRODUCTION

The Comprehensive Fragment Library (CFL) is a set of small, rigid, medicinally interesting fragments. This library originates from a starting set of >3 million potentially synthesizable virtual fragments designed from first principles and 3D enabled to maximize exploration of target interactions. Extensive clustering analysis allows broad coverage of chemistry space and provides an immediate follow-up strategy from any screening hit.

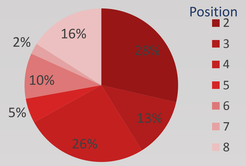
Supplementing Screening Decks with 3D compounds is an enormous task

AbbVie Screening Collection, 2007[1]



Even common 2D cores are distributed unevenly in the literature

193k Monosubstituted Quinolines in SciFinder



BUILDING THE VIRTUAL FRAGMENT SET

An exhaustive set of ring structures containing ≤18 heavy atoms with 1 handle atom was built in BIOVIA Pipeline Pilot with several inclusion criteria:

- ≥1 ring, including bridged, spiro and fused ring connections
- ≤ 3 unique rings
- ≤ 2 ring assemblies
- ≤ 2 rotatable bonds
- 1 atom has all rotatable bonds
- C,N,S,O atoms only
- ≤ 1 S

Rigid, partially aromatic structures are the highest value 3D enabling subset

Core structures decorated with handles chosen for synthesis potential

High Value	Moderate Value	Follow up only	Not chosen
Neutral, extendable	Halogen, high MW	Interaction with handle likely	Modification loses charge

The raw set of > 50 million structures with Me handles was filtered by medicinal chemistry criteria to produce a 7 million candidate fragment set

Example CFL Fragments	Excluded	Reason
		Two handles*
		Unstable
		Isolated double bond*
		Too many rotatable bonds*

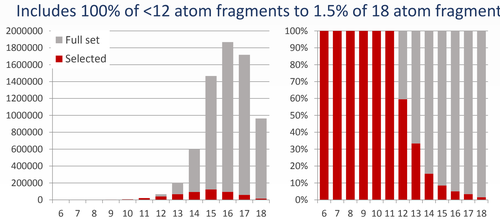
*Potential follow up for a CFL fragment hit

Practical computational limit is <1000k diastereomers

Requires >600 CPU-years for full 3D similarity matrix

A 580k subset was selected for clustering:

- Reduced mean of 16 heavy atoms to 14.6
- Includes representatives of all 9000 ring types
- Includes 100% of <12 atom fragments to 1.5% of 18 atom fragments

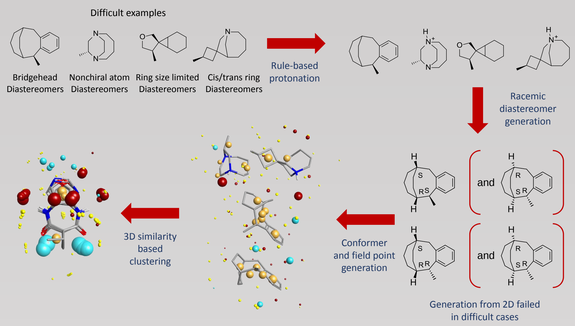


CLUSTERING

cresset

830k diastereomers were generated from the 580k 2D fragments

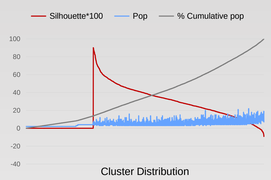
- RDKit was used to enumerate diastereomers, with manual correction
- Cresset's XedeX generates ≤5 conformations for each



2D and 3D clustering methods were investigated on a 20k distributed subset

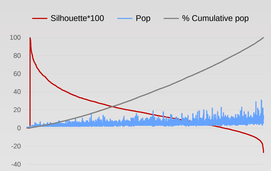
- 4k clusters using a parallel C++ implementation of the k-medoids/CLARANS algorithm[3]
- Cluster tightness and separation assessed by the silhouette metric[4]
  - 1.0: ideal; -1.0: misclustered

2D method: ECFP4



Average silhouette: 0.22

3D method: 0.75 fields, 0.25 shape

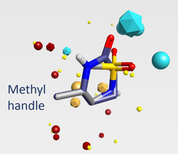


Average silhouette: 0.19

3D clustering method

Alignment to methyl handle:
- Rotate structures around the handle
- Determine maximum similarity

Similar compounds:
- Have common vectors
- Represent alternative sprouting choices

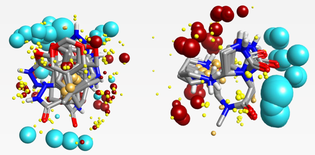


Maximum similarity: 0.78

3D clustering produced better clusters

Comparison of two 6-member clusters

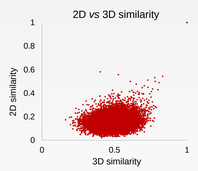
Silhouette: 0.509



2D Cluster

3D Cluster

2D and 3D are not correlated



3D clustering of the entire set was still computationally expensive

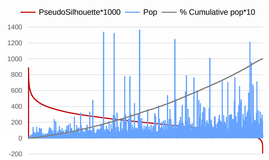
Initial clusters from a full similarity matrix of 150k compounds — 14 CPU-years

Examined 10k, 25k and 50k clusters

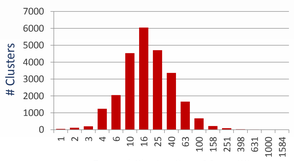
- 10k gave a well-distributed 150k compound set; not scalable to the entire library
- 50k gave cluster sizes too small to be stable

Remaining compounds placed in 25k clusters of most similar medoid — 20 CPU-years

- PseudoSilhouette calculated because full matrix unavailable



Mean intra-cluster similarity: 0.77



Binned Cluster Size, Mean 33

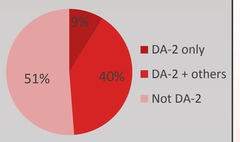
BUILDING THE PHYSICAL LIBRARY

Compounds prioritized by HBond pattern

CFL fragments with 1 likely interaction motif preferred

HBonds most common interactions with different targets
- Important for selectivity

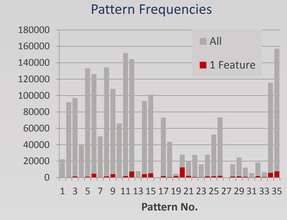
2D similarity measure
- Same patterns are distributed over clusters
- ~50% of clusters have simple, valuable patterns



HBond patterns were generated for all compounds
- Pipeline Pilot Molecular Pharmacophore Fingerprints Component
- HBA, HBD paired when separated by 1-6 bonds
- 35 core patterns including Named Functional Groups

Structure	Pattern Name	Special case of DA-2
	DA-2	
	Amide	DA-2
	Urea	DA-2,DA-2,DD-2
	Pyridine	Protonation state is handle dependent

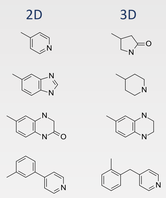
Pattern Frequencies



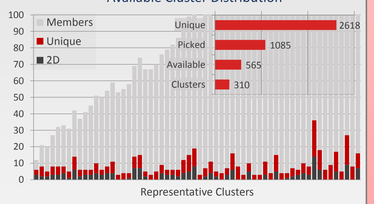
Available Chemical Directory™ (ACD) Search
- Handles of each CFL Candidate generated and searched individually in DiscoveryGate™
- Grouped by HBond pattern, starting with least complex: >850,000 queries
- Hits filtered for real vendors with a real price: 11,184 results (1.6%), 2,618 unique cores
- Manually reviewed for medicinal chemistry value and purchase

Available Compounds found in 1395 clusters (5.6% of total)
- 905 clusters have only 1 commercial member, 81% are 3D enabled
- Others evenly distributed 2D and 3D by PMI1,2
  - On rod/disk axis ≡ slope from 2-butyne or benzene near -1

Examples



Available Cluster Distribution



CFL contains the high value commercial set in <1 plate
Most CFL compounds are not commercial
Can be synthesized from available starting materials
Library members are novel to targets, novel for IP

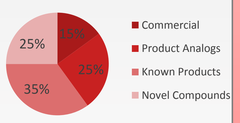


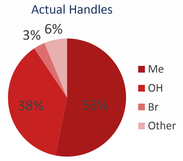
PLATE 1

Representative compounds picked from available clusters
- Compounds >0.8 similarity to medoid
- Ring systems varied over columns
- HBond patterns grouped in rows

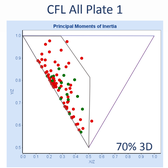
Plate assembly underway
- Range of vendors
- Synthesis when handle options limited
- Purities confirmed
- 200 mM DMSO solubility confirmed

	Plate 1	CFL Design
# Fragments	55	~1000
# Atoms	11.5	13
AlogP	1.0	1.0
HBD	0.91	0.6
HBA	1.80	2.2
Rot Bonds	0.91	0.5
%3D	70%	80%
%Unique Rings	55%	75%
% Commercial	80%	15%
% Handles	100%	100%

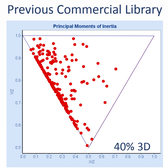
Actual Handles



CFL All Plate 1



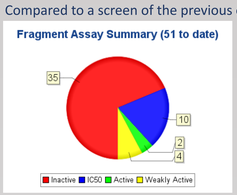
Previous Commercial Library

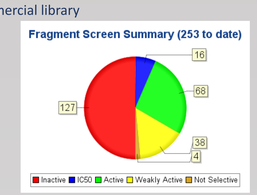


SCREENING AND FOLLOWUP

Visionary Pharmaceuticals

Library screened at 2 mM in a high concentration kinase inhibition assay
- Compared to a screen of the previous commercial library



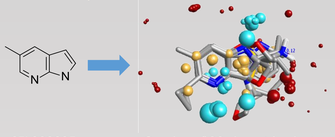


Lower hit rate from added topology
Higher % confirm in dose response
Follow up possible on weak, interesting compounds

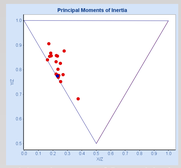
Privileged chemotypes present in both

Representative Hit

Included as a positive control for kinases

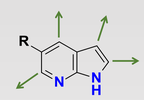


CFL000025 — Overlay of 20 closest in cluster



Members are 3D enabled

CFL clustering reveals many related compounds for rapid fragment SAR



Independent Analogs	CFL000025	10 Fragment Hits	Plate 1
Handle Analogs	**10**	100	550
Handle Vectors	4	47	257
Same 2D Scaffold	37	343	3308
Same HBond Pattern	274	12857	42667
Same 3D Cluster	47	573	6371
Total CFL Compounds	873	13676	50669
% Unique to Compound	100%	74%	26%
CFL Analogs	>30,000	>1 M	>25 M

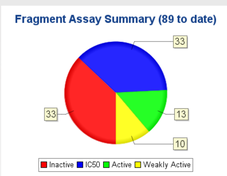
Leap-to-Lead™ generates fragment analogs

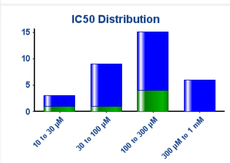
Compounds from the fragment hit column were assayed:
- New SAR identified from handle vectors
- New scaffolds identified from core alterations
- >3-fold enrichment of actives
- ~10-fold increase in best potency, increase in LE

Analogs include non-commercial compounds

IP identified despite well-known cores

One series was progressed to medicinal chemistry
Advanced to achieve µM potency in cell assays
Provisional filed, synthesis ongoing





Conclusions

Libraries of commercial compounds are insufficient to generate new chemical matter
The Comprehensive Fragment Library gives greater access to 3D-enabled interaction space
Preliminary CFL screening generates high quality hits through its coverage of chemistry space
Enhanced 3D enablement reveals new, unexpected chemotypes for important drug targets
The CFL gives high quality entry points to our proprietary Leap-to-Lead™ discovery platform

REFERENCES

[1]Akritopoulou-Zanze, I.; Metz, J.T.; Djuric, S.W. *Drug Discovery Today*, **2007**, 948-952.

[2]Metz, J.T. Principal Moments of Inertia Protocol, Pipeline Pilot Community.

[3]Ng, R. T.; Han, J. *IEEE T. Knowl. Data. En.* **2002**, *14*, 1003-1016.

[4]Rousseeuw, P. J. *J. Comput. Appl. Math.* **1987**, *20*, 53-65.

CONTACT

BioBlocks is developing the CFL for use in Lead Discovery Collaborations. While the library is not available for independent purchase, we welcome new collaborators. For additional information or to discuss using the CFL and the Leap-to-Lead™ platform in a drug discovery effort, please contact wwade@bioblocks.com or visit our website: www.bioblocks.com