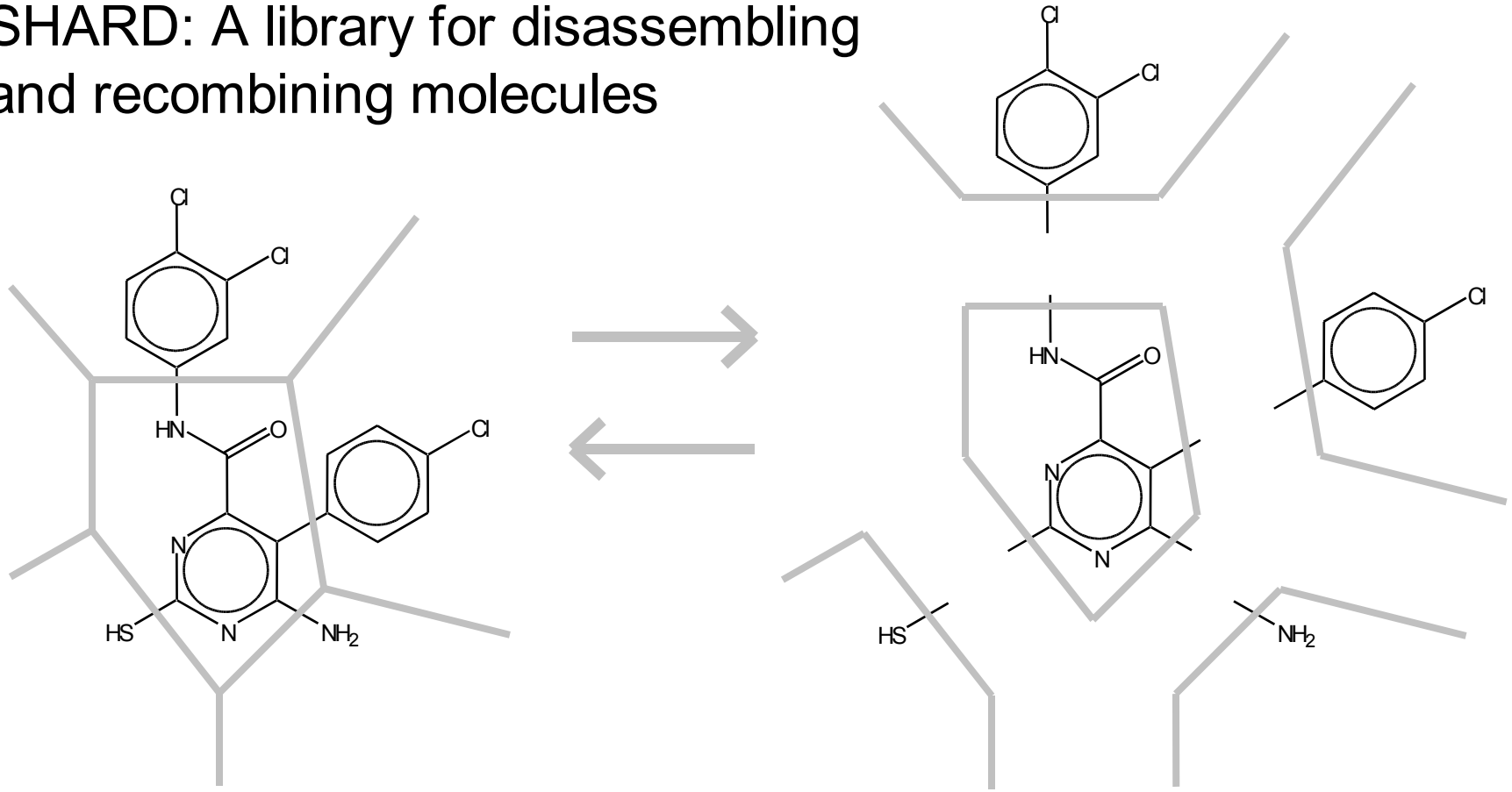
SHARD: A library for disassembling and recombining molecules



Given a SMILES structure SHARD returns

a core (perhaps defined by SMARTS)

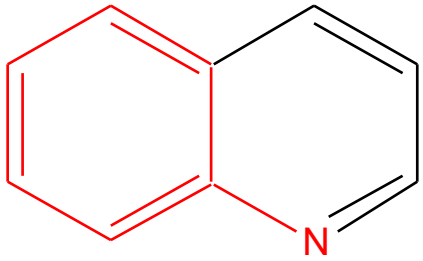
plus substituents (canonical, in SMARTS order)

SHARD can also put the pieces back together

How is “core” defined?

1) User input via SMARTS (like findsubs)

One difference: If a pattern matches atoms in a ring, the rest of the ring is included.



2) Predefined general definitions

A set of features defined via SMARTS (rings, HBA, HBD, etc.)

Plus the atoms that connect them

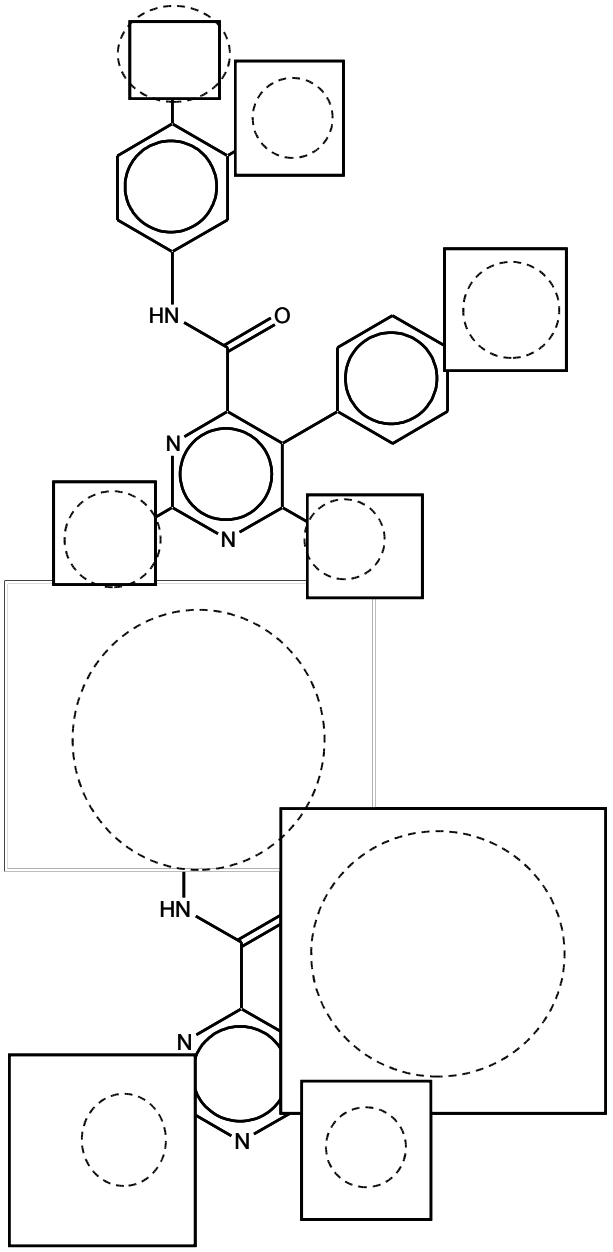
Existing libraries can be digested this way.

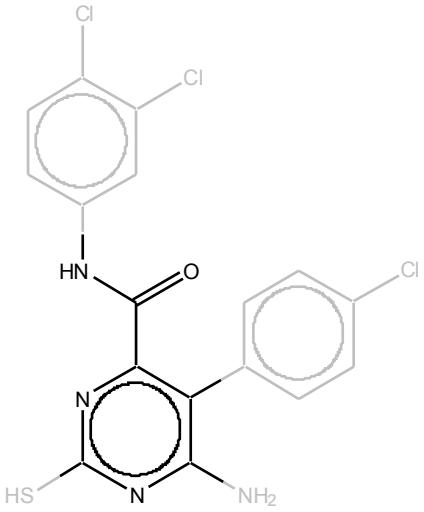
Examples

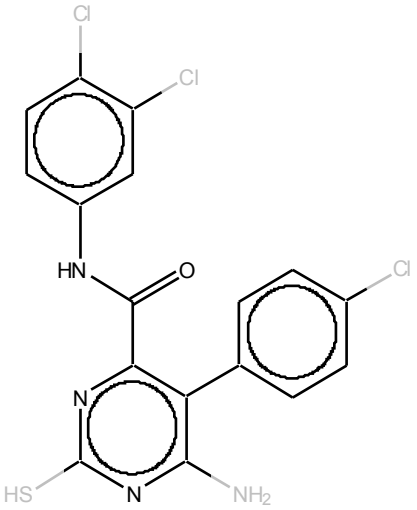
1) Rings and atoms connecting them (POPI - it eats the spinach)

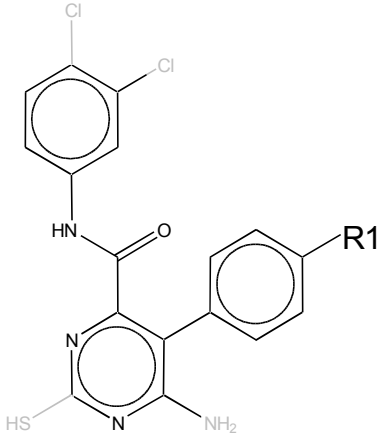
2) Inner features and atoms that connect them (PIPs)

* Atoms conjugated to the core may be included in core if desired




Cl
Cl
HN
O
Cl
N
HS
N
NH2


Cl
Cl
HN
O
Cl
N
HS
N
NH2


Cl
Cl
HN
O
R1
N
HS
N
NH2

Uses for SHARD

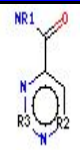
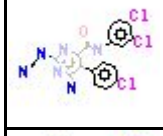
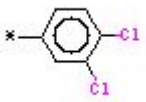
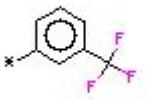
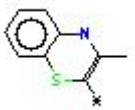
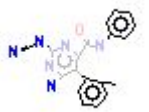
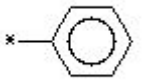
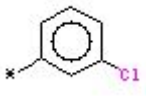
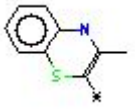
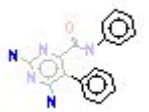
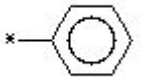
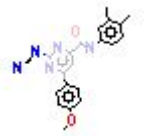
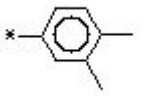
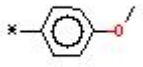
Automated SAR table generation

Deployed on the Neurogen intranet

SMARTS, POPI and PIP cores

Multiple core structures are allowed

Aligned core structures

Num A	Core A			
26				
ID	Core	R1	R2	R3
	A		N	
	A			
	A		N	
	A			
	A		N	N
	A			

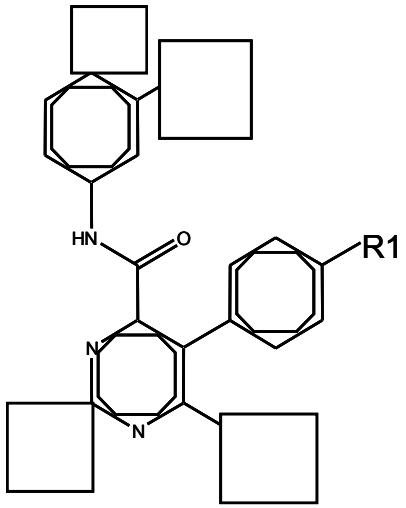
Uses for SHARD

Virtual Library Generation

Virtual libraries based on permutations of substitutions derived from singleton compounds

Based on existing compounds, so the libraries are “synthesizable” in some sense

Libraries can be generated from external substituent libraries



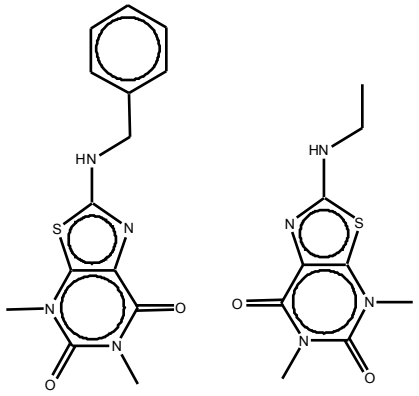
R1= Cl, Br, NH2, Me, Et, etc.

Uses for SHARD:

SERPENT: (Search for RePlacemENT)

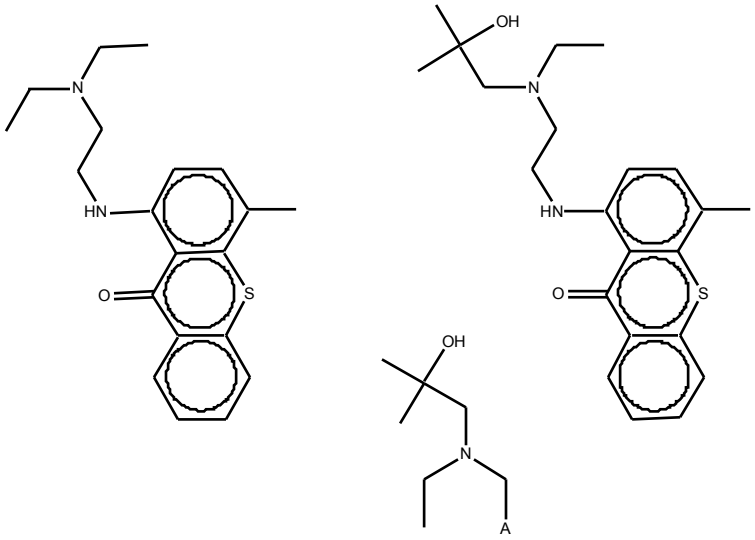
Show me all compounds where R2 replaces R1

Example: Methyl replaces Phenyl



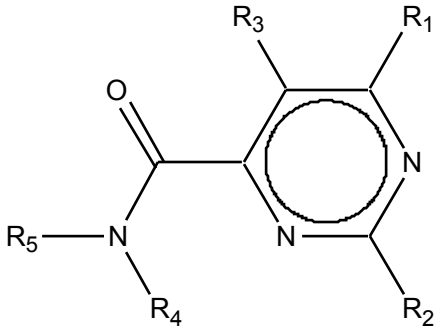
HYDRA: (Many headed SERPENT)

Show me all R2's which replace R1



Uses for SHARD:

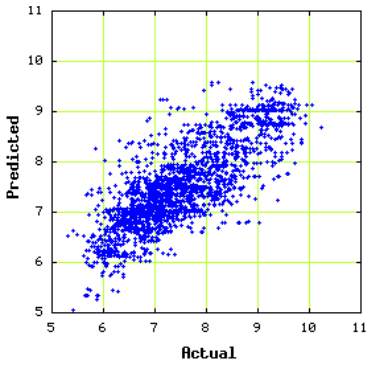
2.5D modeling – more than 2D less than 3D
Limited number of desciptors



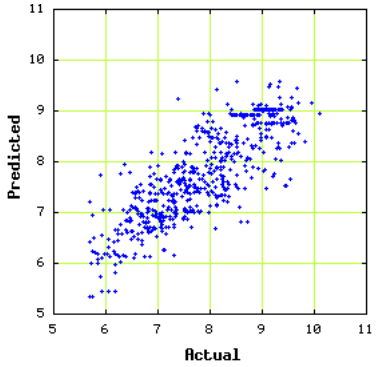
	Core	Properties for R1			Properties for R2			Properties for R3			
Cmpd A	1	P1(R1)	P2(R1)	P3(R1)	P1(R2)	P2(R2)	P3(R2)	P1(R3)	P2(R3)	P3(R3)	…
Cmpd B	2	P1(R1)	P2(R1)	P3(R1)	P1(R2)	P2(R2)	P3(R2)	P1(R3)	P2(R3)	P3(R3)	…
…	X	P1(R1)	P2(R1)	P3(R1)	P1(R2)	P2(R2)	P3(R2)	P1(R3)	P2(R3)	P3(R3)	…

Properties can include MW, size, shape, polarity, presence/absence, etc.

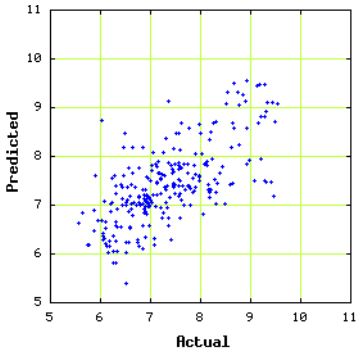
A note of caution:



Training set



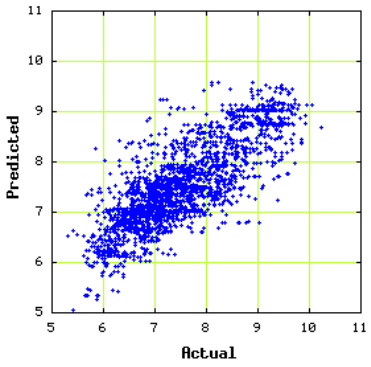
Validation Set



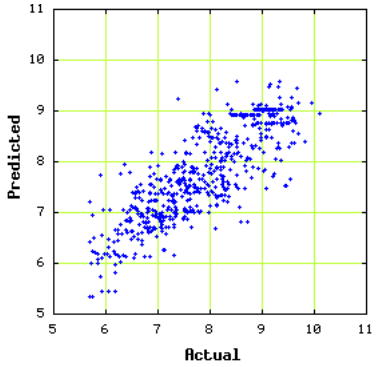
Test Set

Training Set (58.8%)	R_2=0.64
Validation Set (23.5%)	R_2=0.61
Test Set (17.6%)	R_2=0.53

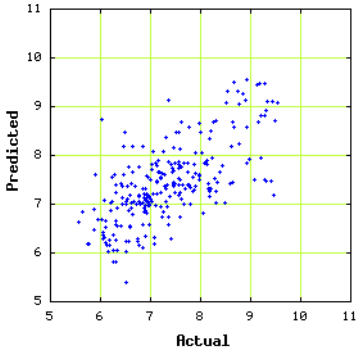
A note of caution:



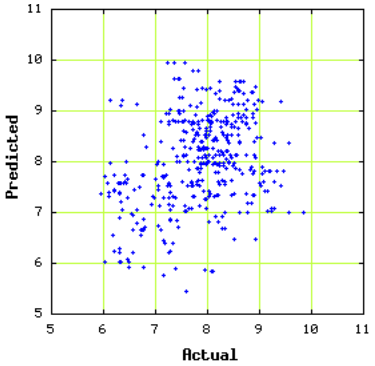
Training Set



Validation Set



Test Set



PrFiT Set

Training Set (58.8%)	R_2=0.64	PrFiT Set (15%) R_2=0.04
Validation Set (23.5%)	R_2=0.61	
Test Set (17.6%)	R_2=0.53	