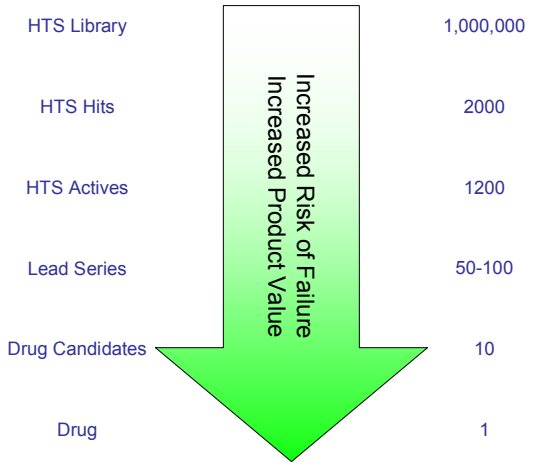
ChemModeling

Lead Generation

Jon Swanson
jon@chemmodeling.com
(636) 329-0300
ChemModeling, LLC
May 4, 2009

Drug Discovery is a Difficult Process



Oprea, Ed. ChemoInformatics in Drug Discovery, Wiley, 2005, p. 28

ChemModeling

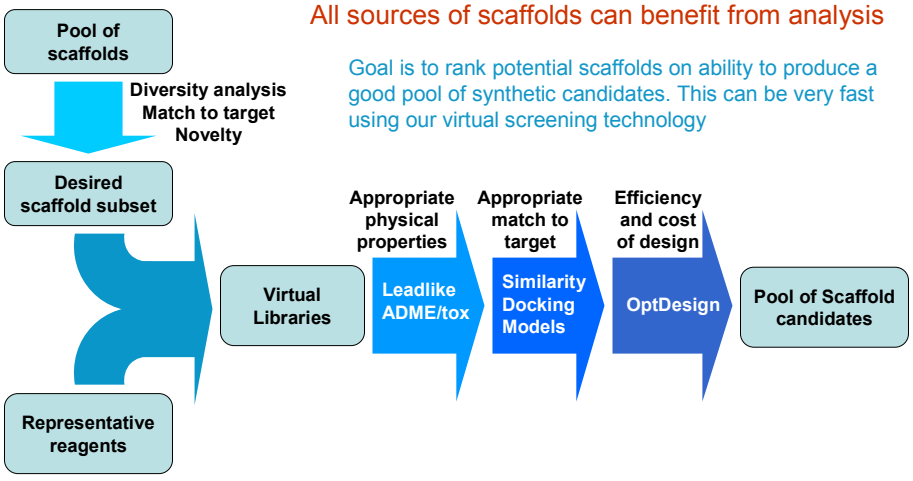
Knowledge-Driven Design Finds Optimal Candidates

- Where do we look?
  - Libraries of existing compounds
    - In-house collections
    - Vendor collections
  - Virtual Libraries
    - Suggestions from chemists
    - Constructed from “Ideas” databases
    - De Novo scaffold generation
- How do we look?
  - Properties specific to the target
  - Similarity to known active compounds
  - Match of pharmacophore model built from known actives
  - Relevant docking pose and reasonable score in crystal structure
- What do we look for?
  - Medicinal chemistry relevance
  - Optimal diversity/representativeness of compound set
  - Embedded SAR potential for quick optimization
  - Reasonable chemistry allowing rapid follow-up
  - High quality compounds (pure, well characterized, reliable vendor)

ChemModeling

3

For Virtual Libraries: First Select the Best Scaffolds



All sources of scaffolds can benefit from analysis

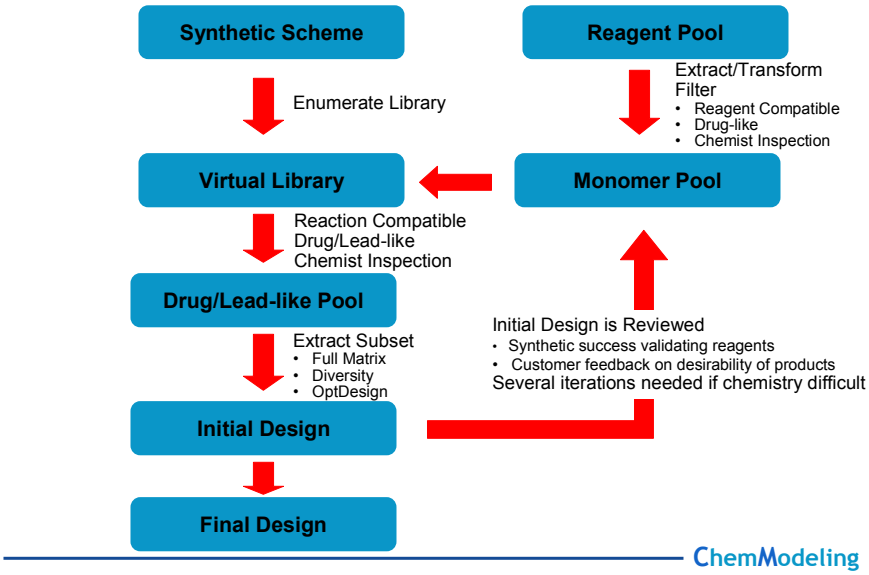
Goal is to rank potential scaffolds on ability to produce a good pool of synthetic candidates. This can be very fast using our virtual screening technology

Process most important for scaffolds with the least prior medicinal chemistry input

ChemModeling

4

Best Scaffolds are Used in a Full Library Design



ChemModeling

Considerations for Virtual Docking

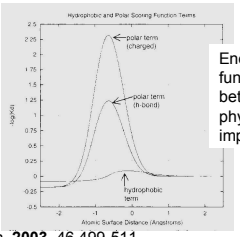
- You need to find the right conformation
  - Generate conformations up front
  - Incremental build-up
- You need a way to score these conformations
  - Molecular mechanics type functions
  - "Potential of Mean Force" methods
- You need to rank the results
  - Functions used during docking are necessarily fast and simple
  - Post docking, better methods can be used
- Ideally, protein flexibility can also be considered

FlexX Scoring Function

$$\Delta G = \Delta G_0 + \Delta G_{rot} N_{rot} + \Delta G_{hb} \sum_{h-bonds} f(\Delta R, \Delta\alpha) + \Delta G_{ion} \sum_{ionic} f(\Delta R, \Delta\alpha) + \Delta G_{aro} \sum_{aromatic} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} \sum_{lipo} f(\Delta R)$$

Strength of interactions function of the deviation from ideal

SurFlex Potentials



Energy of interaction function distance between atoms, physical effects are implicit.

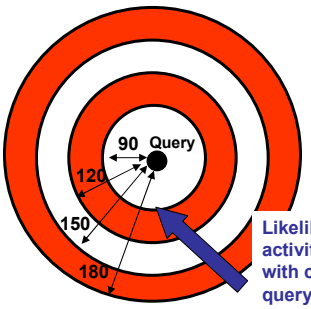
Rarey, et al J Mol Bio **1996**, 251, 470-480; Jain, J Med Chem, **2003**, 46,499-511

ChemModeling

Considerations for Similarity Searching

- You need a suitable function to assess similarity
  - Field based approach
  - Distance to critical features
  - Overlap of atoms/critical features
  - CoMFA model predictions
- You need get the right conformations
  - Rule-based construction
  - Incremental alignment of fragments
  - Generate a fixed set of conformations
- You need an efficient way to search large numbers of structures
  - Generate distances for components of a virtual library
  - Vector-based assessment of similarity from representative population

Distribution of topomer hits from BMS/Tripos validation



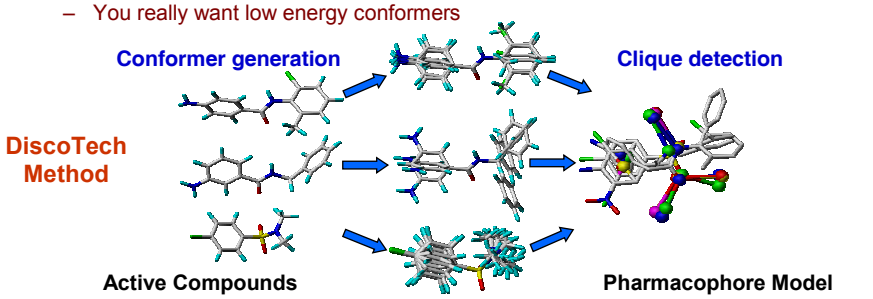
region	# tested	% active
< 90	4	0.50
90 - 120	61	0.08
120 - 150	176	0.00
150 - 180	156	0.00
>189	28	0.00

Cramer, R.D., Poss, M.A., Hermsmeier, M.A, Caulfield,T.J, Kowala,M.C., Valentine, M.T J Med Chem **1999**, 42, 3919-3933

ChemModeling

Considerations for Pharmacophore Modeling

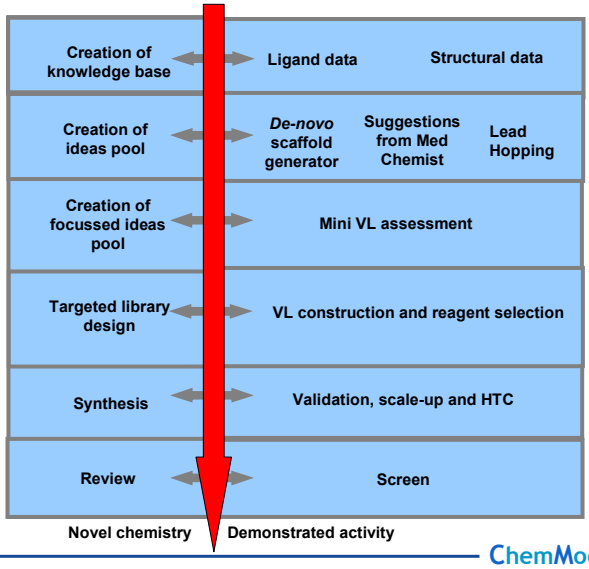
- You need to find the right conformations
  - Generate a fixed set of conformations
  - Modify conformations on the fly (typically using a genetic algorithm)
- You need to find the right model
  - Programs typically generate tens to hundred of alternates
  - Pick models that distinguish actives from in-actives
- Once you have the pharmacophore, you need to find matches
  - Typically done with 3D flexible searching
  - You really want low energy conformers



ChemModeling

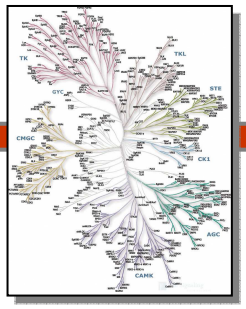
Example: Kinase Targeted Libraries

The complete process from collecting data to obtaining active compounds



ChemModeling

Creation of the Knowledge Base



Copyright Cell signalling

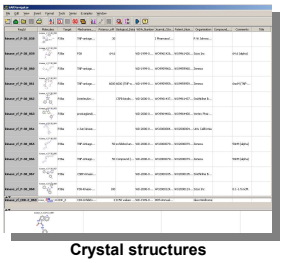
Which kinases do we target?

CDK2, P38a, FGFR, GSK3b, JnK3….

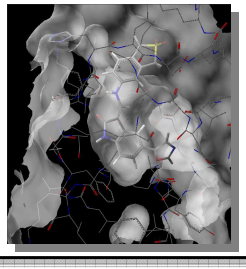
Selection criteria

- Occurrence
- Therapeutic relevance
- Availability of ligand /structural information
- Commercial screen
- Diversity

Ligand database



Crystal structures



ChemModeling

Creation of the Ideas Pool

100's of scaffolds

Creation of VLs: from drug-like reagent pool

4 boronic acids ⇨ **Kinase 104 scaffold** ⇦ **17 amines**

Ranked
- Docking score
- Visualisation
- Similarity
- Properties
- Synthetic feasibility

name	synthesis notes	size	property violations	topomer similar	(raw) flexs rank	(nor) flexs rank	(vl) flexx f rank	(ex) flexx rank
kinase_59	5	60	2	1	2	5	1	1
kinase_104	5	68	1	4	1.5	3	1.5	1
kinase_110*	4	561	3	0	1.5	10	1.5	5.5

Library	1a9u	1ke7	1m52	2fgi	Average Rank	Overall Rank
kinase104	4	2	1.5	1	2.12	1.5

Scaffolds selected for full design

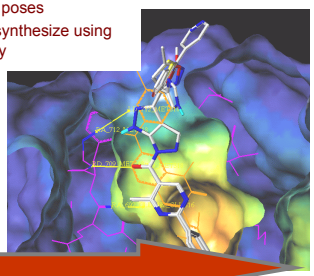
ChemModeling

Fast Docking Strategy

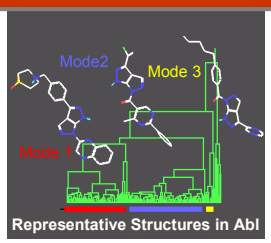
Dock a Diverse Subset of the Virtual Library
- Start with filtered virtual library
- Select by diversity or similarity

	Best	Worst	Avg
P38	-25.6	2.8	-11.6
CDK2	-26.7	2.7	-12.1
Abl	-32.7	-0.8	-17.2
FGFr	-21.9	-5.1	-14.4

Dock Complete Virtual Library
- Use identified base poses
- Select products to synthesize using scores and similarity



Identify Reasonable Binding Modes
- Cluster on RMS distance between docked structures
- Visually inspect examples from each mode



Representative Structures in Abl

ChemModeling

Results: Enhancement Rates – p38 MAP Kinase

Baseline - 10 Actives From 97 Screened for p38 = 10.3%

Subset	Subset size	No Actives	% Actives
All	97	10	10.3
Dock (< -20)	58	9	15.5
Topomer (< 120)	26	6	23.1
Intersection	14	5	35.7

Combination of docking & topomer searches = *~36% hit rate*

Good basis for creating designs from large virtual libraries

ChemModeling

Results for All Libraries

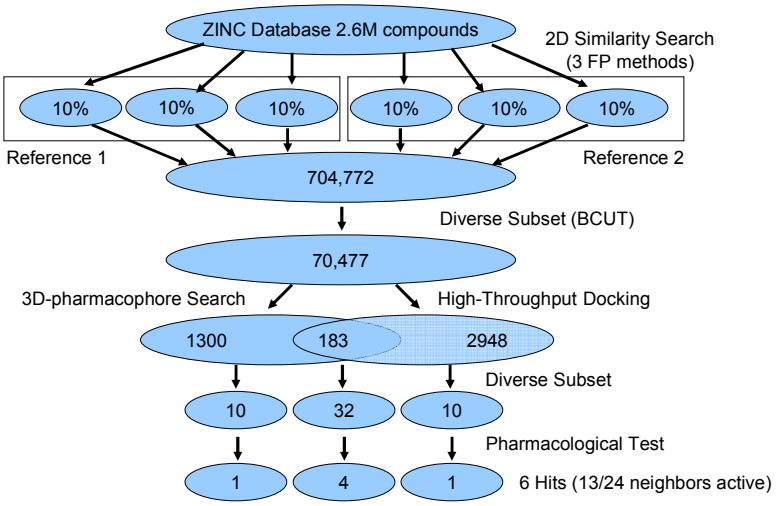
Library	# actives	# cmpds
1	6	15
2	3	22
3	3	11
4	1	22
5	1	28
6	2	19
7	0	28
8	3	10
9	7	12
10	12	20
11	3	16
12	0	27
13	4	10
14	1	4
15	10	16
16	0	7
17	9	11
18	6	6
Total	71	284
Hit Rate	25%	

- ~225 library ideas evaluated computationally
- ~48 library evaluated synthetically
- ~3800 compounds from 19 different scaffolds produced
- Assay performed by Upstate in the UK
  - Panel of 9 kinases
  - Measure % control at 10 uM ligand concentration
  - Results are the average of two assays
  - 10 uM ATP concentration
  - Active is defined as 50% of control or less
- 13 batches tested
- 284 compounds tested

- # of actives = # of compounds which show >50% inhibition @ 10uM in at least one assay.
- A compound displaying multiple kinase activity counts as one active.

ChemModeling

An Example of Computer-Aided Drug Discovery



Tikhonova, I.G;Sum, C.S; Neumann, S.; Engel, S.; Raaka, B.M.; Costanzi, S.; Gershengorn, M.C.. J. Med. Chem. 2008, 51, 625-633.

ChemModeling

15

Summary of the Complete Process

- Most compounds that enter the drug discovery process fail
  - Use target knowledge early to eliminate poor candidates
- Reduce the attrition rate through intelligent (informatics-based) application of appropriate tools
  - Efficient library design process
  - Property filters and predictions
  - Activity prediction tools (receptor and ligand based)
  - Chemistry expertise and knowledge
- Goal is to obtain lead compounds with
  - Appropriate biological activity
  - Appropriate target selectivity
  - Acceptable ADME-Tox profile
  - Patentable position
- All accomplished in as short a time as possible

ChemModeling

16