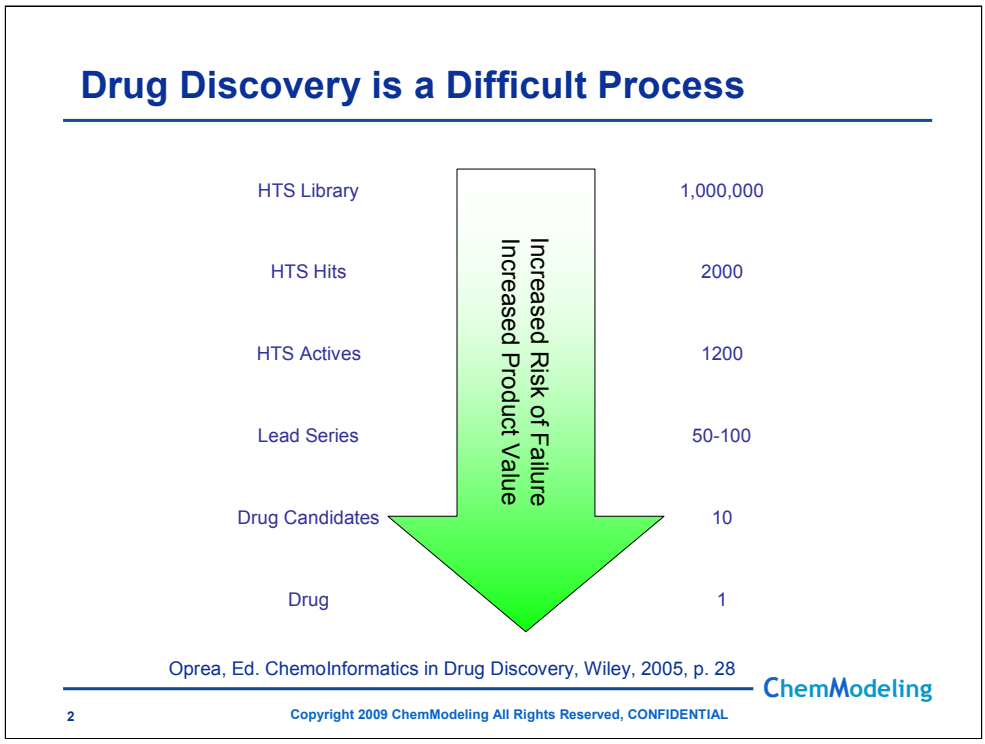
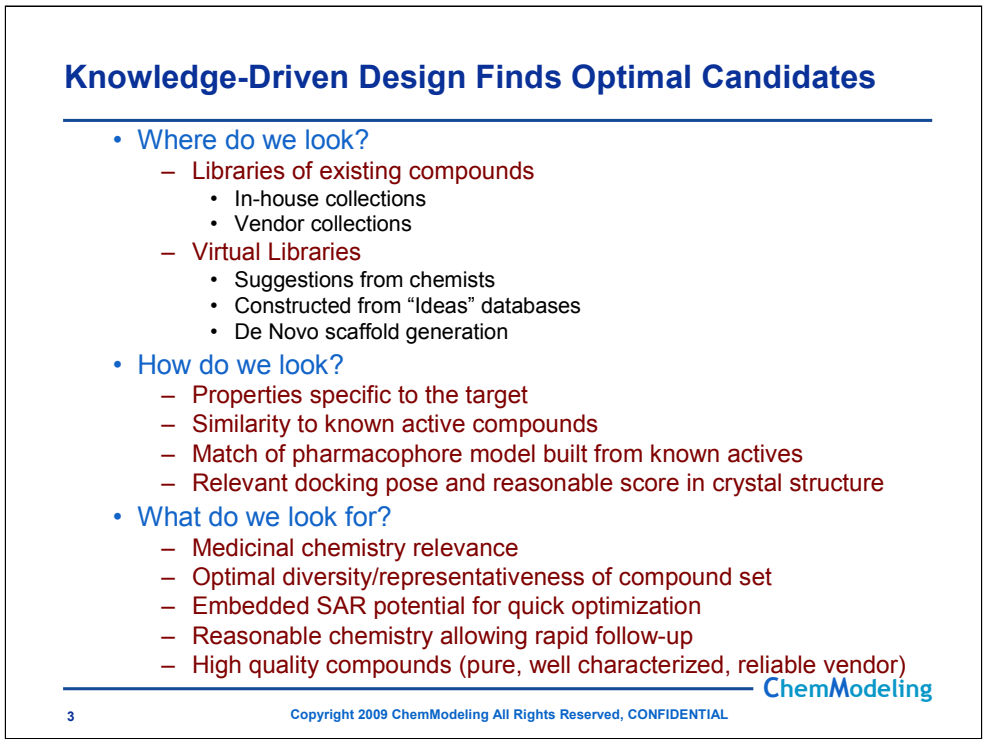
ChemModeling

Lead Generation

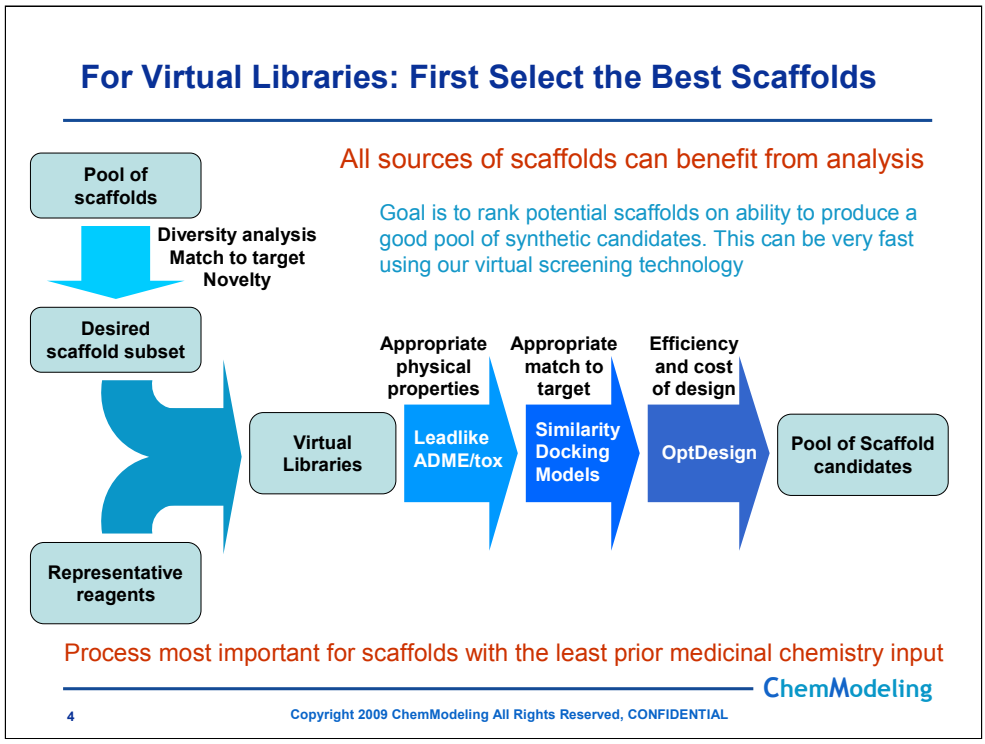
Jon Swanson
jon@chemmodeling.com
(636) 329-0300
ChemModeling, LLC
May 4, 2009



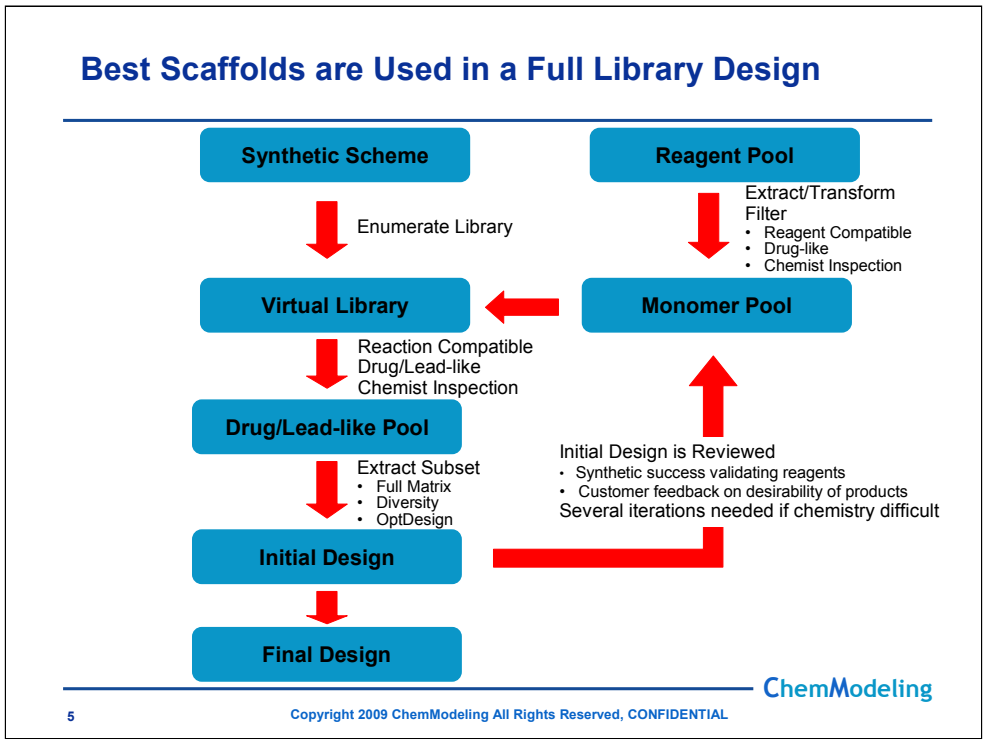
It is well-known that the number of compounds that become drugs is a small fraction of the compounds that are initially screened. Computational techniques are one method to eliminate poor candidates earlier in the drug-discovery process. These techniques include finding a better pool of compounds to screen initially, to identifying compounds expected to have poor specificity or ADME-tox profiles. We have a great deal of experience helping companies reduce costs by intelligent selection of compounds.



To start a lead discovery process one needs a source of screening compounds. We've had experience selecting compounds from existing vendor collections and designing custom libraries. The former case is more likely when little is known about the target as one has a greater diversity of compounds to sample from and the costs/compound are potentially lower. The latter case provides for novel chemical entities which are more focused to a specific target. We have used property profiling, similarity searching, virtual docking and pharmacophore searching to focus the selection. In the case of large vendor libraries, we will typically do diverse subset selection to reduce the pool that we select from. When working with our synthesis partners, we develop libraries that will provide meaningful SAR and allow easy follow-up lead optimization.



When making virtual libraries, good scaffolds are key. We've been involved in projects that represent a wide range of philosophies for scaffold selection. The simplest case is that the customer provides a synthetic protocol and no scaffold selection is needed. We've also been given rough templates as to what was desired and in conjunction with our chemists devised possible synthetic routes to scaffolds that met the template. We then did topomer distance clustering to pick a subset that were diverse with respect to each other. We have also used internal ideas databases as a starting point and done virtual screening or similarity scoring to select scaffolds. Our in-house scaffold generator is another source of ideas, which we virtually screen and have ranked by medicinal chemists.



Library design tends to be somewhat specific to the customer, as each has their own preferences and needs. This slide shows the general outline of the library design process. One aspect that distinguishes our approach is that we feel that library designs are most effective when developed along with input from the medicinal chemists. At each stage in the design we obtain feedback from the chemists, who help drive the design. We also see the design as an iterative process. For example, as synthetic validation occurs, significant changes to the planned reagent pool may be required.

Considerations for Virtual Docking

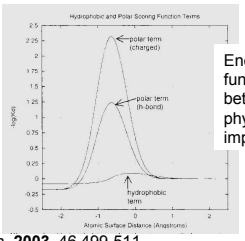
- You need to find the right conformation
  - Generate conformations up front
  - Incremental build-up
- You need a way to score these conformations
  - Molecular mechanics type functions
  - "Potential of Mean Force" methods
- You need to rank the results
  - Functions used during docking are necessarily 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

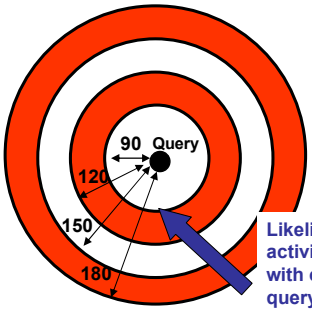
ChemModeling

If a crystal structure is available, virtual docking can be done to identify optimum compounds for a lead discovery library. There are many different ways of docking compounds in a crystal structure and many different scoring functions. Our approach is to look at known actives and evaluate the method and scoring function for the particular problem. If multiple scoring functions show selectivity, consensus scoring can be used.

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



Likelihood of activity increases with closeness to query

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

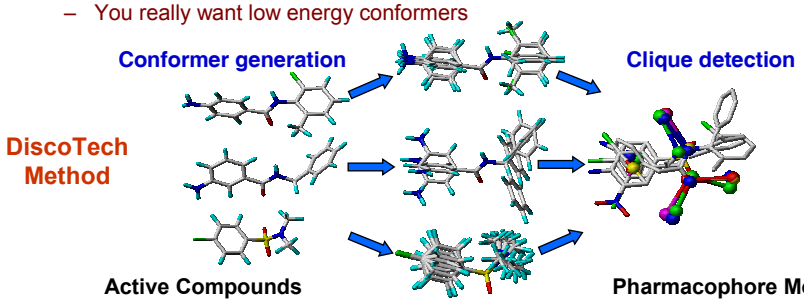
ChemModeling

7

As with scoring functions for docking, similarity methods should be evaluated for their relevance to a particular problem. There are several methods that we can use on very large datasets. The method illustrated above uses topomeric (shape-based) fields to assess similarity. The advantage of this method for virtual libraries is that only the fields for the scaffolds and synthon (not each individual product) need be calculated. Screening virtual libraries is thus A + B rather than an A x B in time, allowing the screening of huge virtual collections. If activity data is available, it may be possible to generate a CoMFA model to predict activity. This can be used with the topomeric fields to predict activities for large virtual libraries. The additional advantage here is that a CoMFA model will incorporate both positive and negative information.

Considerations for Pharmacophore Modeling

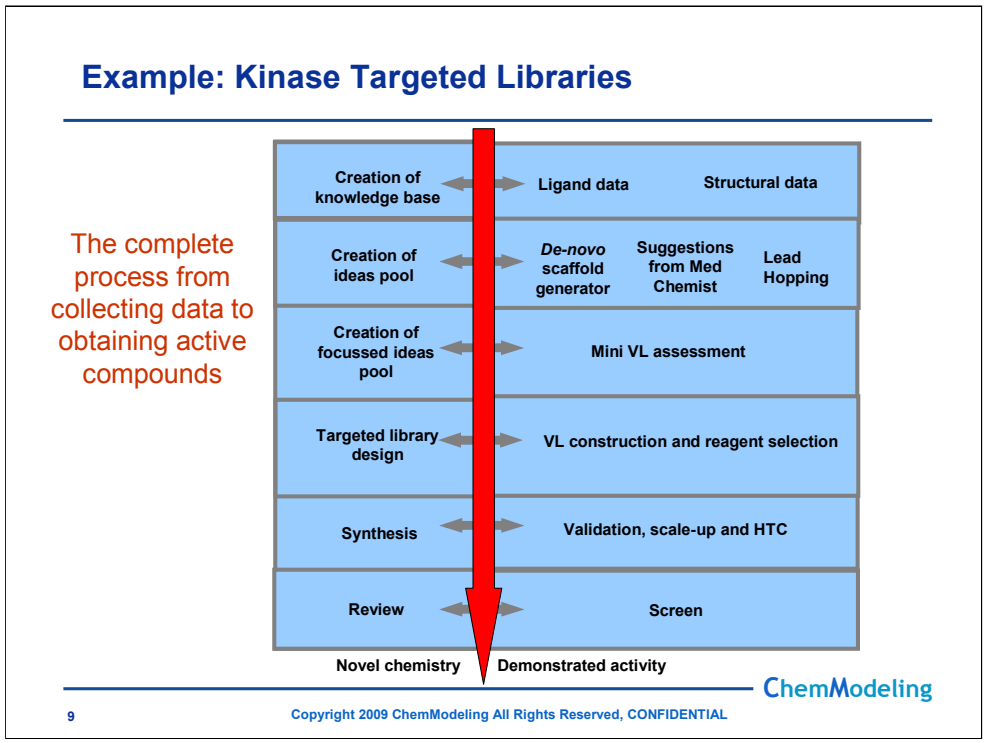
- You need to find the right conformations
  - Generate a fixed set of conformations
  - Modify conformations on the fly (typically using a genetic algorithm)
- You need to find the right model
  - Programs typically generate tens to hundred of alternates
  - Pick models that distinguish actives from 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

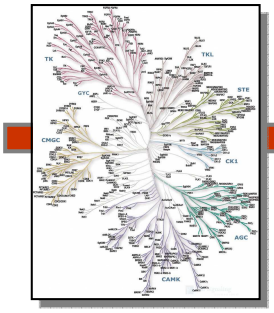
8

We have found pharmacophore modeling is a useful adjunct to similarity searching. The similarity searching is often quite good at identifying similar shapes. The results can then be pruned to those that also posses a relevant pharmacophore. We find one difficulty with pharmacophore modeling is the large number of potential models produced. It is this useful use highly active molecules to generate the models and then combine that with inactive compounds to vet the models. They must be able to distinguish active from inactive compounds to have value.



This shows a kinase targeted library protocol that we used when working as part of Tripos Discovery Research. The first step was to collate information about kinase inhibitors (crystal structures and active compounds). A pool of ideas was generated (actually part of a dynamic process of library building and evaluation). The quality of the ideas was evaluated by building small sublibaries and doing virtual docking and similarity comparisons to rank the scaffolds. The best scaffolds were selected and full library designs produced. These designs typically went through several iterations as the synthetic chemistry validation indicated what reagents would in practice. Finally, a subset of compounds were screened and the results fed back into the knowledge base.

Creation of the Knowledge Base



Copyright Cell signalling

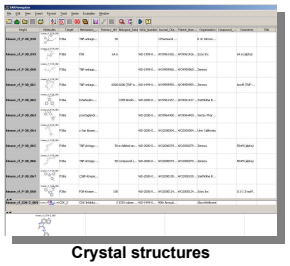
Which kinases do we target?

CDK2, P38a, FGFR, GSK3b, JnK3….

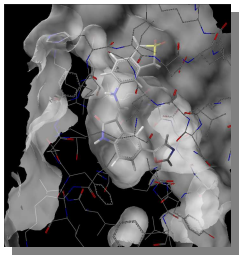
Selection criteria

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

Ligand database



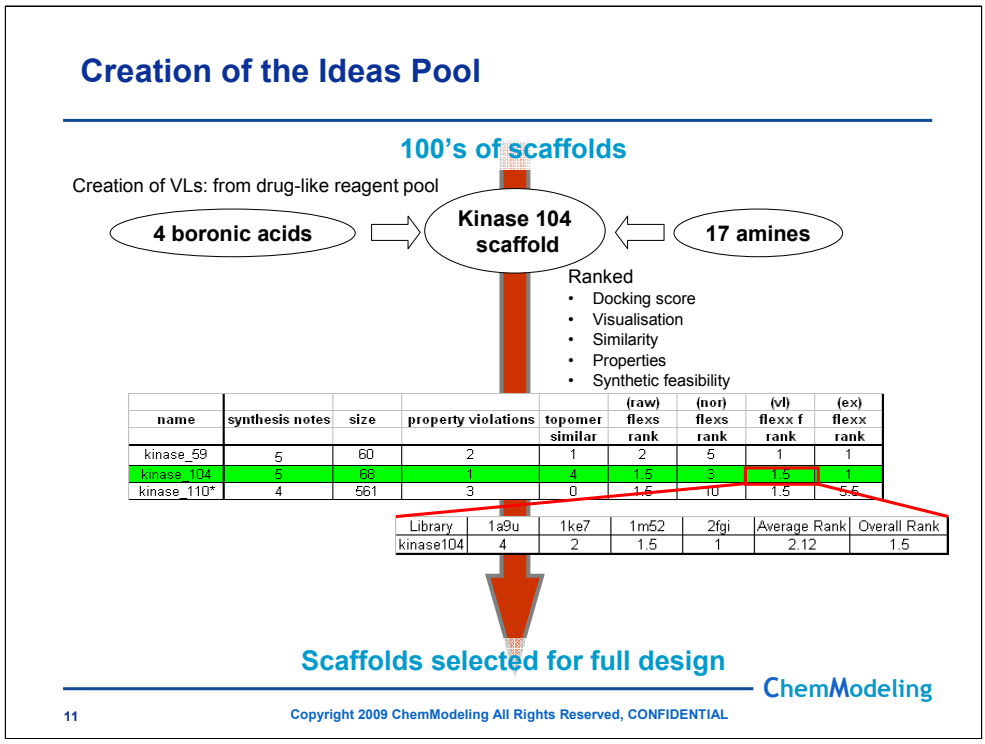
Crystal structures



ChemModeling

10

For the evaluation phase we chose four kinase crystal structures and a set of approximately 200 known kinases to use for virtual docking and similarity assessment.



The full evaluation consisted of an evaluation by the medicinal chemists as to the ease of synthesis, a count of the number of products that failed the Lipinski rules, the number of known kinases topomerically similar to compounds in the library, a virtual docking score (converted to a rank from 1 -10), and similarity ranking based in FlexS scores. In addition, the chemists were provided 3D models of well-docked representatives from the library and 3D overlays to the best scoring active compound. All this information was used to select scaffolds to progress to a full design and chemistry.

Fast Docking Strategy

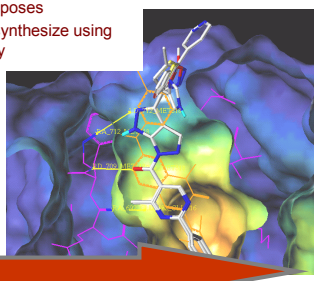
Dock a Diverse Subset of the Virtual Library

- 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

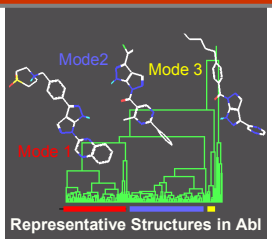
Dock Complete Virtual Library

- 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



ChemModeling

12

We developed a high speed process for doing the virtual docking, so that we can analyze all the compounds that were possible products of the virtual library. Typically, 50,000 or so products could be made from a given scaffold that meet reasonable physical properties for kinases. A subset would be chosen and docked. The scaffold was extracted and a hierarchical clustering done based on RMS distances between the scaffolds. These clustered into several docking modes. One or more modes seemed chemically reasonably. The scaffold was then held fixed, and the virtual library docked, which resulted in docking times of under a second per compound. Similarity was assessed using topomer scoring, which is inherently fast for combinatorial libraries.

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

Products that went into the final design were those with either a very good docking score (< -30) or both a good docking score and a small topomer distance to a known active. Analysis of p38 MAP targeted libraries showed that this did, in fact, produce a good enhancement.

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

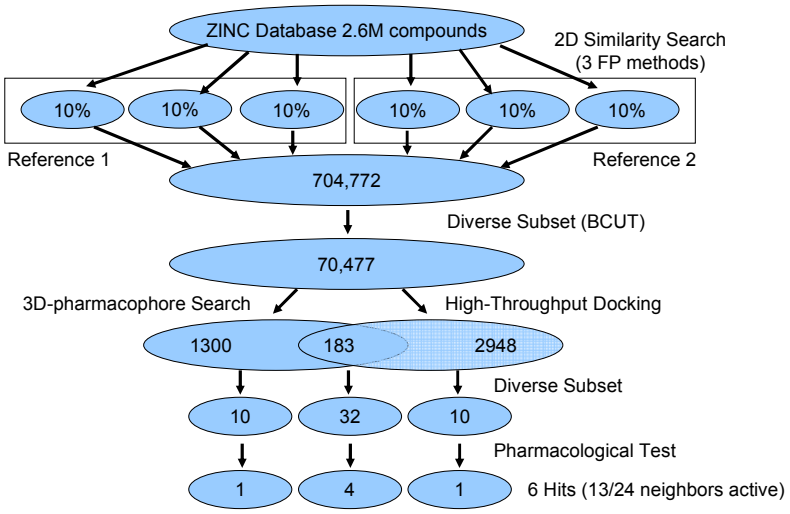
- # of actives = # of compounds which show >50% inhibition @ 10uM in at least one assay.
- A compound displaying multiple kinase activity counts as one active.

ChemModeling

14

A subset of the kinase-targeted libraries were tested and showed a 25% hit rate in one or more assays. In general, while no library was specific to a specific kinase target, individual compounds tended to be specific. We believe this is because the scaffold tended to target the ATP binding region and it was the specific sidechains that provided the specificity to a given kinase.

An Example of Computer-Aided Drug Discovery

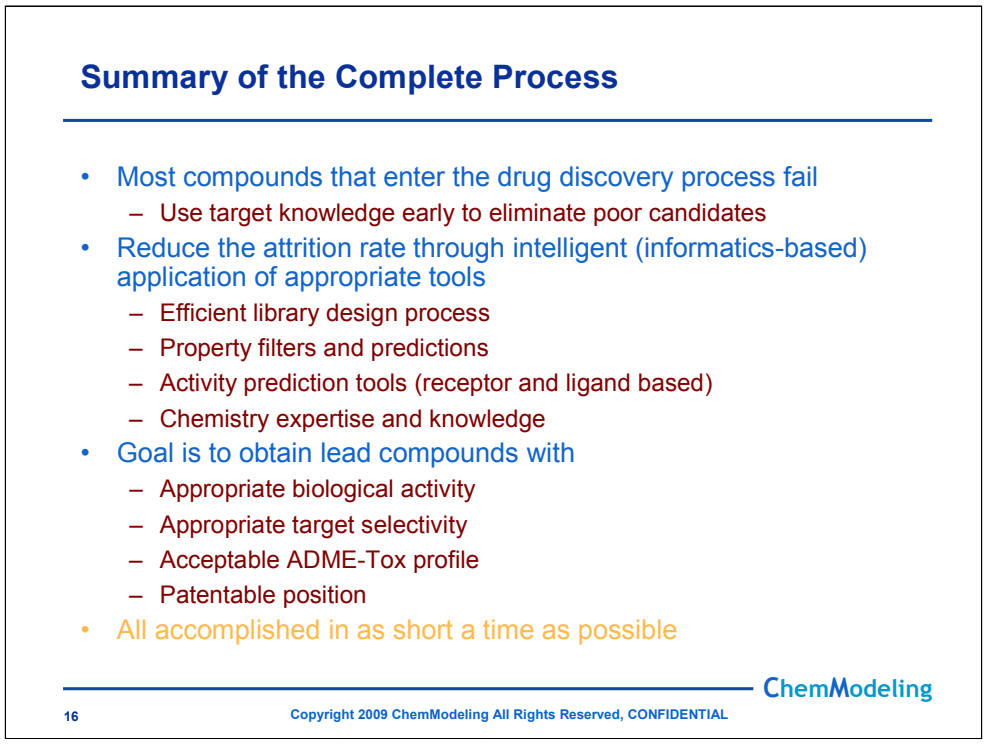


Tikhonova, I.G;Sum, C.S; Neumann, S.; Engel, S.; Raaka, B.M.; Costanzi, S.; Gershengorn, M.C.. J. Med. Chem. 2008, 51, 625-633.

ChemModeling

15

This slide shows a library design process from the literature. This is presented as an example here, as it starts from an existing compound collection, rather than from designed libraries. We have all the tools required to perform a similar experiment. Our process might be a little different. For example, we might include 2D substructure or topomer searching of the initial database. We can use 2D similarity as well as BCUTs for doing the diversity selection. We would also include an initial phase where we cleaned up the ZINC structures and filtered on molecular properties. The exact process would depend on the customer's preferences and the nature of the problem.



We have had considerable success designing libraries that customers were happy with. We plan the design process around the goals of the project and seek input at each step in the design process. In this way, the customer can be assured that the result corresponds with their requirements.