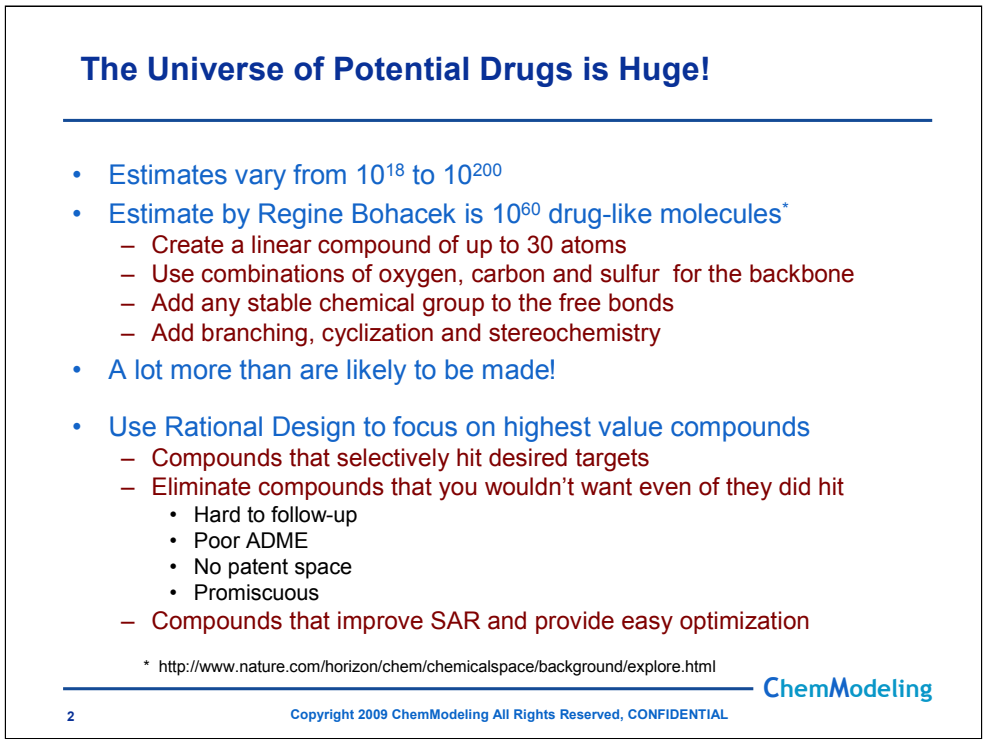
ChemModeling

Library Design

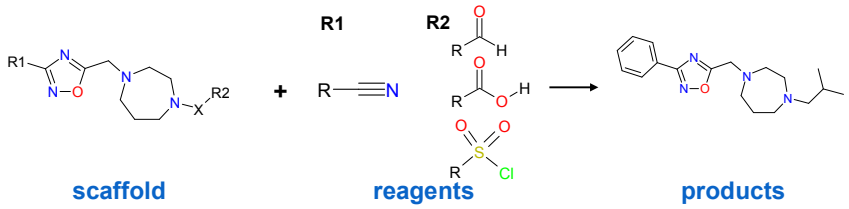
Jon Swanson
jon@chemmodeling.com
(636) 329-0300
ChemModeling, LLC
May 1, 2009



Companies we have worked with typically have collections of hundreds of thousands to millions of compounds. This is a small fraction of the compounds that can be imagined. We work closely with our clients to provide library designs that will provide the highest quality chemical entities and thus the greatest chance of success in the drug discovery process.

Library Designs are Combinatorial

Components of a Combinatorial Library



scaffold reagents products

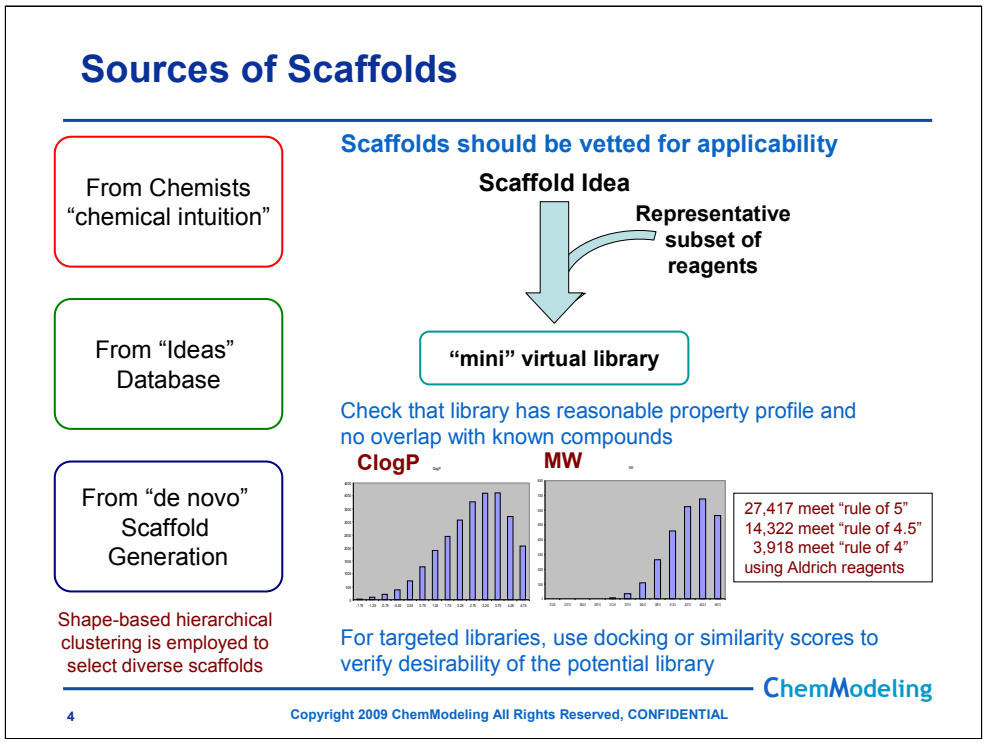
To Generate a Library-

- Identify scaffold and chemistry to be performed on that scaffold
- Create a reagent pool of acceptable monomers based on the chemistry
- Select a subset of products that spans the activity space
  - Diverse/Representative for general screening libraries
  - Use Docking/Similarity for targeted libraries
- Remove compounds that overlap in activity space with the existing collection
- Generate a final design of the desired size and density from the pool of desirable products

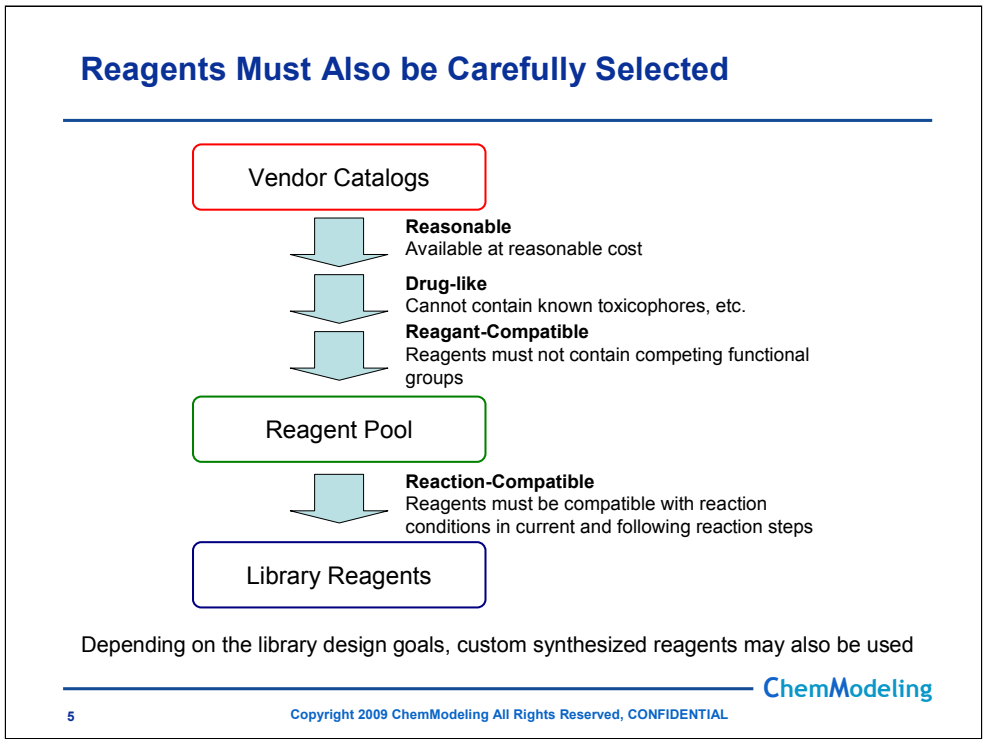
ChemModeling

3

Libraries designs are generated from a combinatorial library definition that we create based on the synthetic route to the desired products. The example shown comes from a synthetic route contained in our internal ideas database. Library designs can range from a few dozen products to several thousand, depending on the use to which the library will be put. Typically, library designs will be in the range of 200-300 compounds. If synthesis is also requested, we can provide alternate reagents as the synthesis progresses to ensure a library of the desired size.



We have had a wide range of sources for the scaffolds that are at the heart of any combinatorial library. The simplest case is when the customer provides the synthetic route. We have also worked on projects where the customer provided rough templates of what they were interested in. The scaffolds that were finally selected were based on a hierarchical analysis of scaffold diversity, novelty assessment, the ability of the medicinal chemists we were working with to define a synthetic route, and the final approval by the customer of the specific scaffold. We have also been involved in intermediate type situations where we chose a subset of scaffolds from a larger set provided by the customer. We have also selected scaffolds from our ideas database using similarity methods to find those most likely to be active in a target specified by the customer.



Reagents need to be processed to be used in combinatorial design. The reagents are converted to synthons which are then combined with scaffolds to produce combinatorial SLNs (cSLNS), a shorthand for complete libraries. We maintain a number of synthon collections. If the synthesis is to be done by one of our partners, we will use their list of available reagents. If the customer is planning to do the synthesis themselves or outsource the synthesis, we will generate a set of synthons that reflect their reagent sources.

Products Must Have Reasonable Properties

Lipinski's "Rule of 5" is the best known filtering criteria

Poor absorption or permeation of an orally administered drug is more likely to occur if any two of these criteria are violated:

- Molecular weight is greater than 500
- Lipophilicity is high (ClogP is greater than 5)
- Number of Hydrogen bond donors is greater than 5
- Number of Hydrogen bond acceptors is greater than 10

There are MANY others

=> Rules need to be tailored to specific customers needs

Properties of Oral Drugs Categorized by Gene Family

	90% MW	90% ClogP	90% HBD	90% HBA	90% Rbonds
Aminergic GPCRs	460	5.6	2	6	8
Ion Channels	430	4.7	3	6	7
Nuclear Hormone Receptors	495	7.3	2	6	10
Peptide GPCRs	752	6.5	8	10	17
Phospho-diesterases	465	5.2	2	8	9
Protein Kinases	505	5.7	4	7	9
Serine Proteases	572	4.8	4	8	12

Hopkins, et al, Nature Biotechnology **2006,** 7, 805-815

ChemModeling

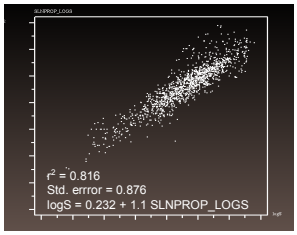
6

The exact property filters depend on the customer and the use for which the library is intended. In some cases we have generated populations with several different filtering criteria so that the customer can see how the pool of possible products is affected by the choice of filters.

Also Consider Other Factors Relevant to Drug Interactions

- Fragment Based Filters (applied to reagents and products)
  - Unwanted
  - Unstable
  - Toxic groups
- Similarity/Dissimilarity to Known Targets
- Custom Scoring Functions

ESOL – Estimated Aqueous Solubility



Plot of ESOL predicted solubility implemented in slnProperty versus the experimental logS values for compounds used as a training set for ALOGPS program from the ALOGPS website.

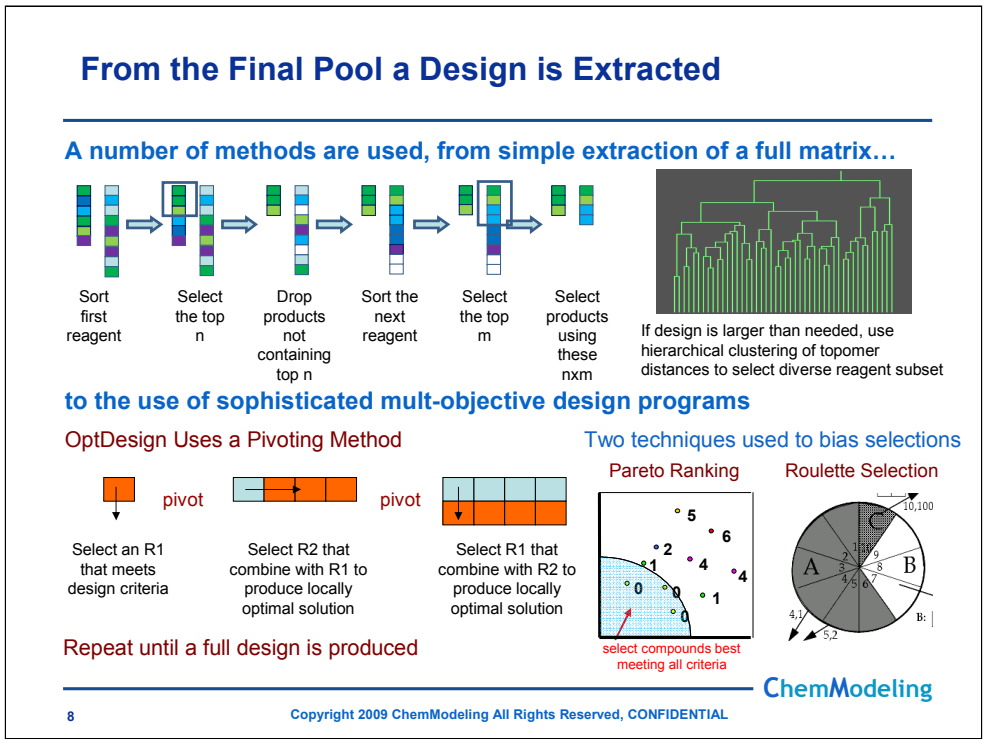
log(S) = 0.16 – 0.63 * ClogP
– 0.0062 * MW + 0.066 * RotBonds
– 0.74 AromaticFraction

AromaticFraction is fraction of heavy atoms in aromatic 6-membered rings

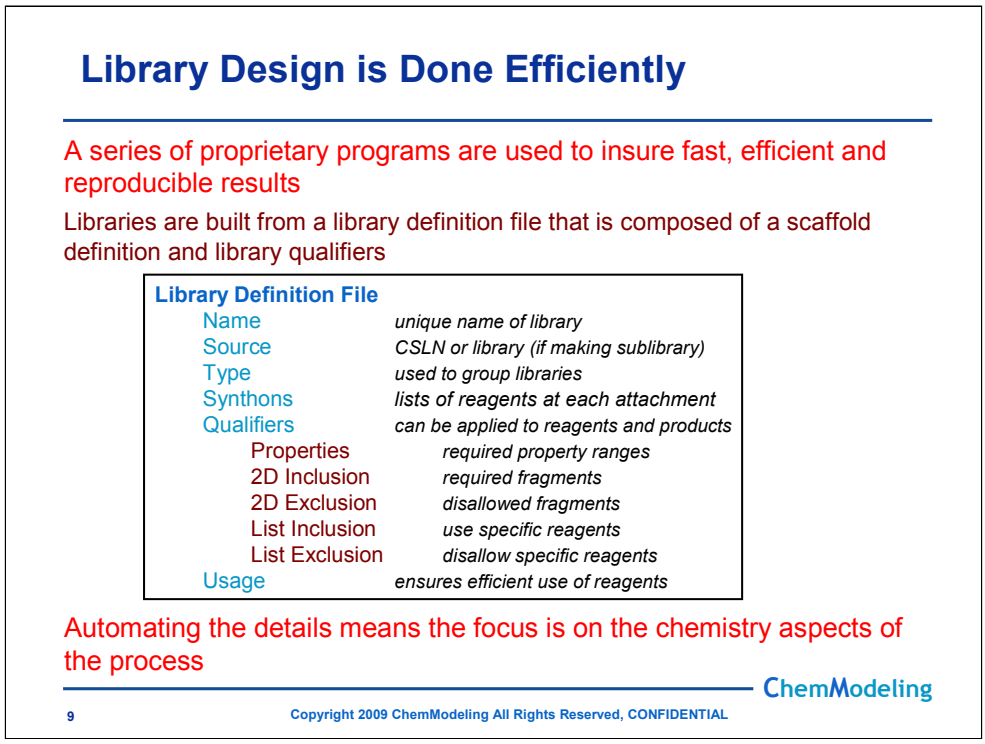
Delaney, J. S. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1000 – 1005.

ChemModeling

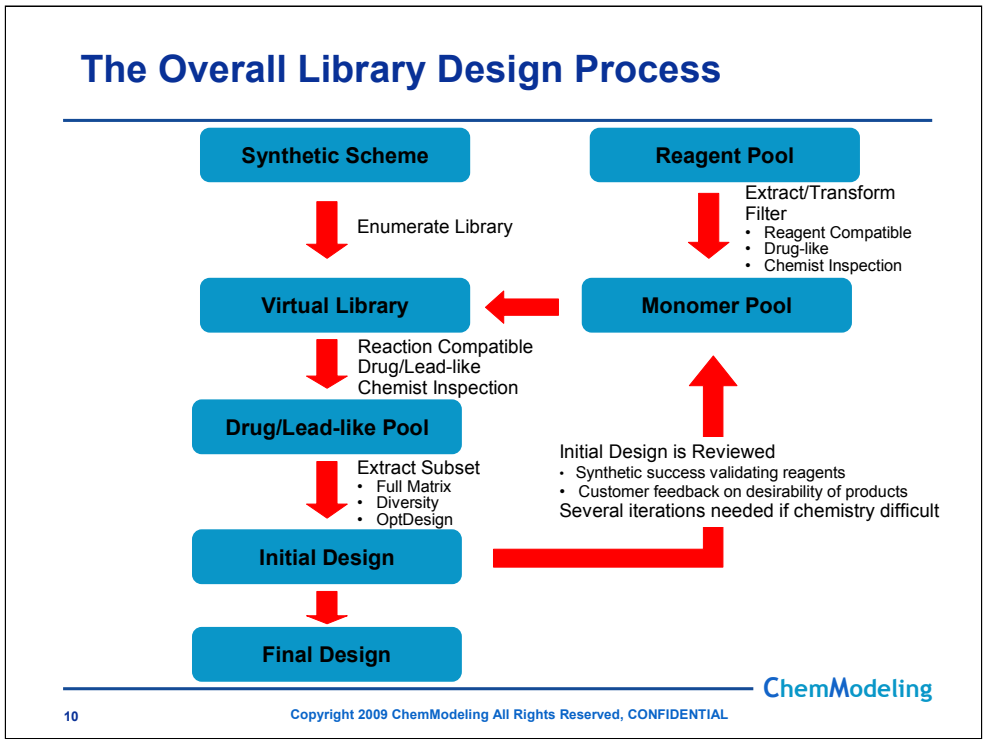
We have developed an extensive set of “undesirable” filters as a result of years of doing library design. We review these with the customer, who is free to add additional filters. We also have a fairly sophisticated scoring function built into our filtering program. It can take any combination of property or fragment filters and create a “score” value that can be filtered on. Either raw values (as in the ESOL implementation) or limits (as one might use to implement the Lipinski rules- i.e., at most two violations) can be encoded.



We have developed a number of tools ourselves, as well as using Tripos tools for extracting a subset of a library, The simplest case is when the customer wants a full or nearly full matrix. The prior filtering steps have already ensured that the compounds remaining are of high value. We can then go through and extract full matrices. Usually there is no single solution to generating a full matrix, so we pick one of right “shape” for the final design the customer is looking for and then use shape-based topomer clustering to select a final set of diverse reagents from those that make up the full matrix solution. At other times we have been requested to optimize the design with respect to a range of requirements, such physical properties, intra-library diversity, similarity to an external data set and bias reagent selection toward reagents with historically good success rates. For this OptDesign was used. When multiple objectives need to be satisfied, compounds are ranked ordered in each objective. The Pareto function ranks each compound by the number of compounds by which it is dominated. This allows selection of compounds that are good in all criteria. With the roulette wheel selection, reagents are assigned to a class, based on some attribute (e.g., availability), and each class is assigned an integer weight. Candidates are chosen on the basis of two numbers- one determines the class from which the candidate will be drawn, and the other determines which member will be chosen.



We have been designing libraries for a long time and have evolved a fairly efficient system for defining, enumerating and filtering virtual libraries. This lets us focus on more important facets of the process, such as making sure the reagents are compatible with the chemistry that will be performed and that the filters will be effective in targeting the type of products desired by the customer.



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.

Summary of the Complete Process

- Design is an Iterative Process
  - What pleases the chemist often doesn't please the computer and vice versa
  - All designs are a balance of competing requirements
- Design is a Step-wise Process
  - A design should start with a good choice of scaffold
  - The preliminary design provides feedback as to the final library size and properties
  - It may not be advisable to proceed to the final design in every case
  - Modification of scaffold and/or reagents may in some cases greatly improve the design
- Decisions that need to be made
  - Source of the synthetic route
  - Source of the reagent lists
  - Exact values for the design constraints
  - Acceptable number of iterations at each stage and criteria to continue forward

ChemModeling

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.