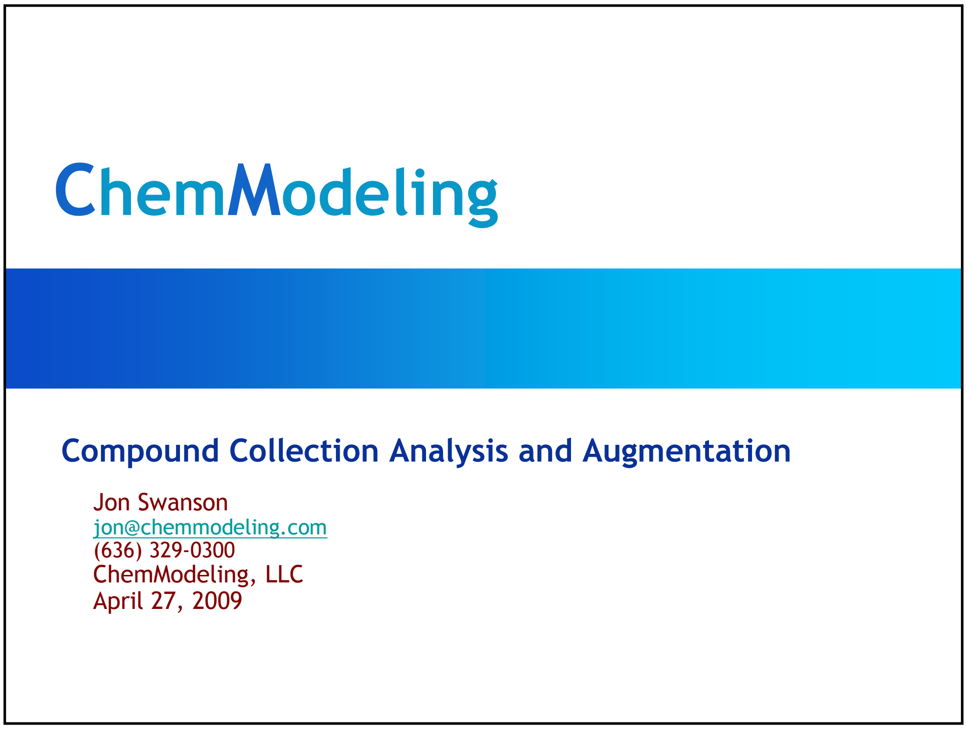
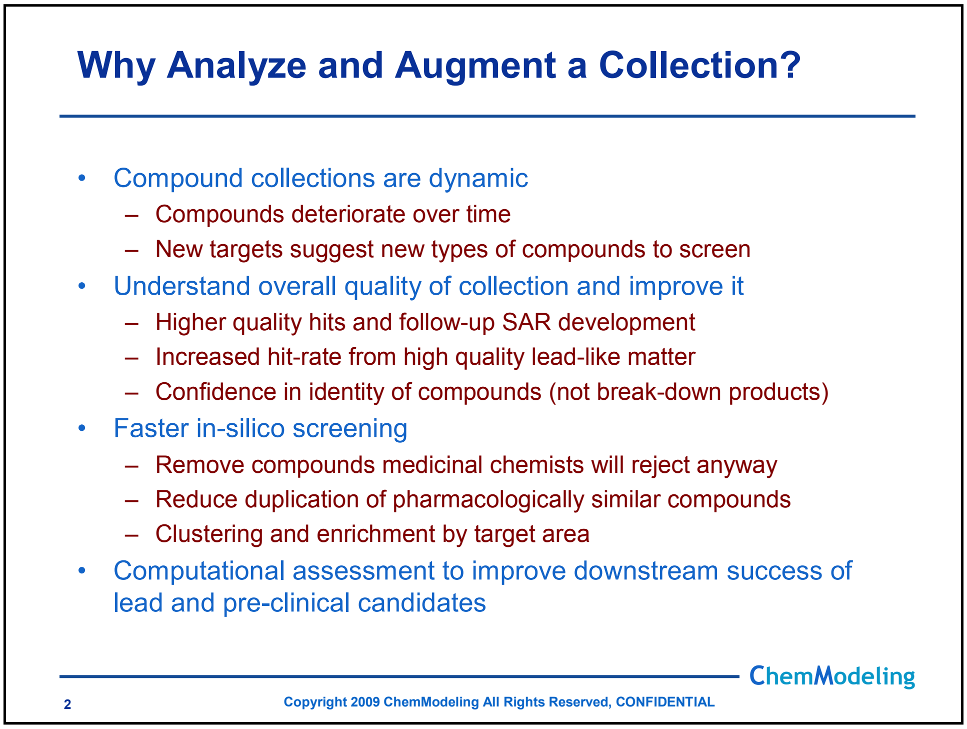


ChemModeling

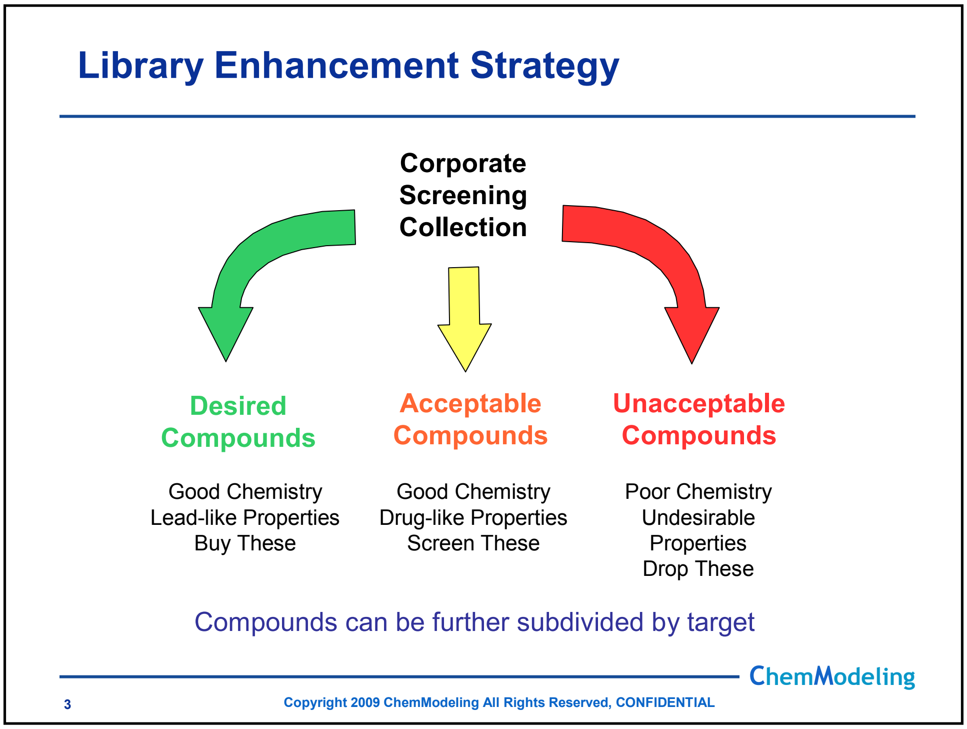
Compound Collection Analysis and Augmentation

Jon Swanson
jon@chemmodeling.com
(636) 329-0300
ChemModeling, LLC
April 27, 2009



Why Analyze and Augment a Collection?

- Compound collections are dynamic
  - Compounds deteriorate over time
  - New targets suggest new types of compounds to screen
- Understand overall quality of collection and improve it
  - Higher quality hits and follow-up SAR development
  - Increased hit-rate from high quality lead-like matter
  - Confidence in identity of compounds (not break-down products)
- Faster in-silico screening
  - Remove compounds medicinal chemists will reject anyway
  - Reduce duplication of pharmacologically similar compounds
  - Clustering and enrichment by target area
- Computational assessment to improve downstream success of lead and pre-clinical candidates



Library Enhancement Strategy

Desired Compounds

Good Chemistry
Lead-like Properties
Buy These

Acceptable Compounds

Good Chemistry
Drug-like Properties
Screen These

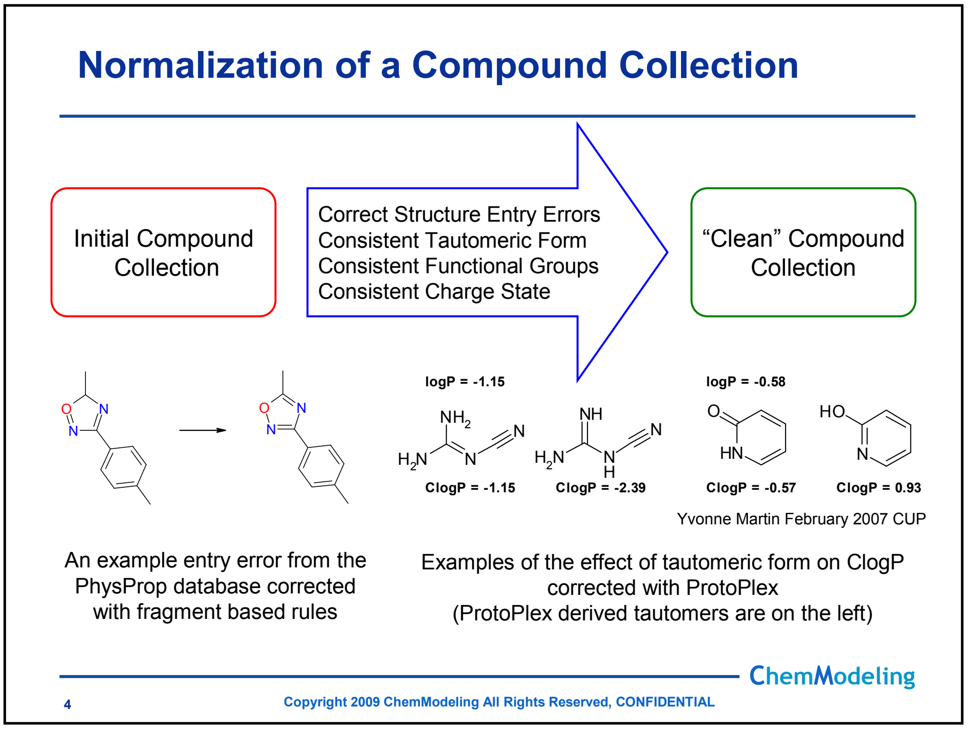
Unacceptable Compounds

Poor Chemistry
Undesirable Properties
Drop These

Compounds can be further subdivided by target

ChemModeling

3



Normalization of a Compound Collection

Initial Compound Collection

Correct Structure Entry Errors
Consistent Tautomeric Form
Consistent Functional Groups
Consistent Charge State

“Clean” Compound Collection

Yvonne Martin February 2007 CUP

An example entry error from the PhysProp database corrected with fragment based rules

Examples of the effect of tautomeric form on ClogP corrected with ProtoPlex
(ProtoPlex derived tautomers are on the left)

ChemModeling

4

Toward a Lead-like or Targeted Subset

Compounds in a screening set should have drug-like or lead-like 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

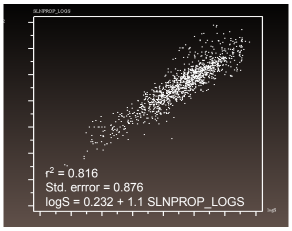
	90% MW	90% ClogP	90% HBD	90% HBA	90% Rbonds
Aminergic GPCRs	460	5.6	2	6	8
Ion Channels	430	4.7	3	6	7
Nuclear Hormone Receptors	495	7.3	2	6	10
Peptide GPCRs	752	6.5	8	10	17
Phospho-diesterases	465	5.2	2	8	9
Protein Kinases	505	5.7	4	7	9
Serine Proteases	572	4.8	4	8	12

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

Other Factors are also Important

- Fragment Based Filters
  - Unwanted
  - Unstable
  - Toxic groups
- Similarity/Dissimilarity to Known Targets
- Custom Scoring Functions

ESOL – Estimated Aqueous Solubility

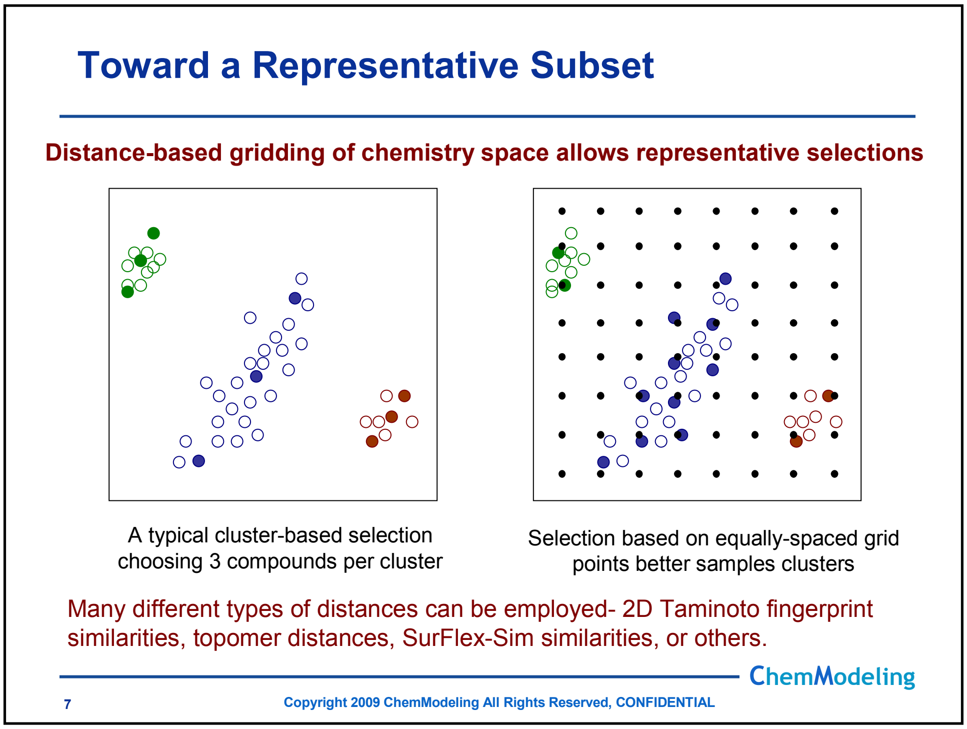


Plot of ESOL predicted solubility implemented in slnProperty versus the experimental logS values for compounds used as a training set for ALOGPS program from the ALOGPS website.

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.



Toward a Representative Subset

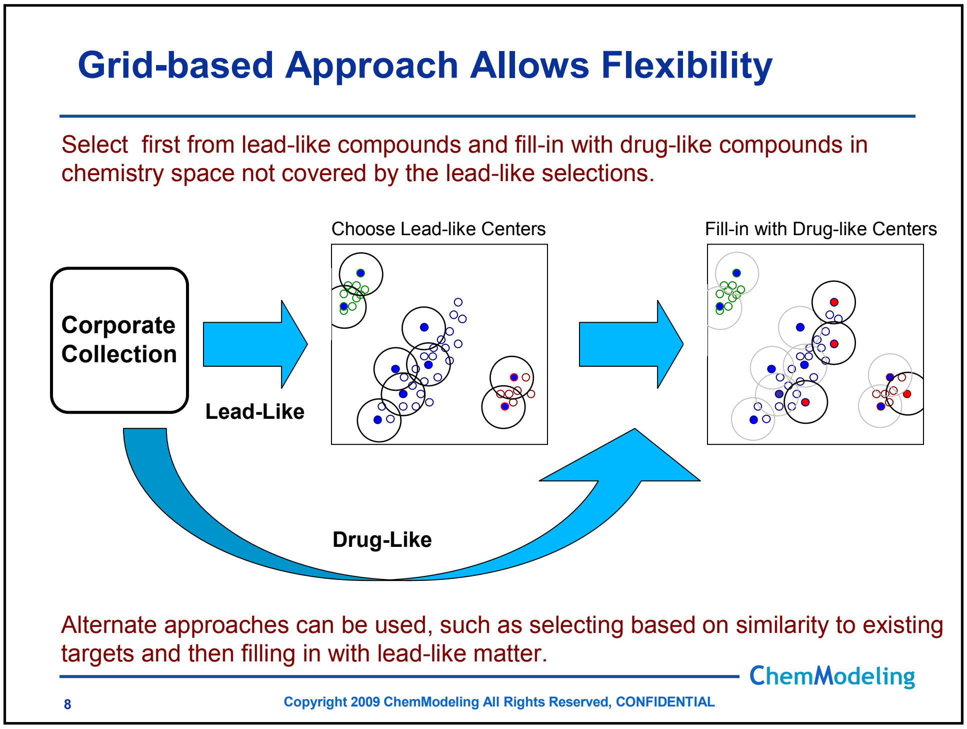
Distance-based gridding of chemistry space allows representative selections

A typical cluster-based selection choosing 3 compounds per cluster

Selection based on equally-spaced grid points better samples clusters

Many different types of distances can be employed- 2D Taminoto fingerprint similarities, topomer distances, SurFlex-Sim similarities, or others.

ChemModeling



Grid-based Approach Allows Flexibility

Select first from lead-like compounds and fill-in with drug-like compounds in chemistry space not covered by the lead-like selections.

Corporate Collection

Lead-Like

Choose Lead-like Centers

Fill-in with Drug-like Centers

Drug-Like

Alternate approaches can be used, such as selecting based on similarity to existing targets and then filling in with lead-like matter.

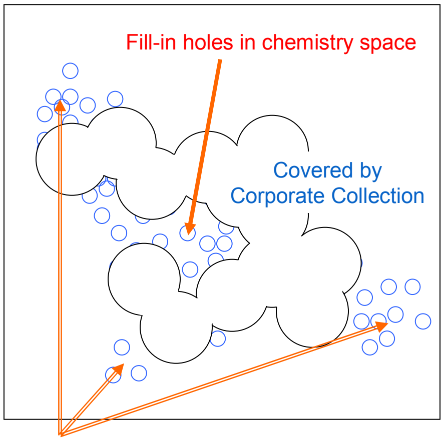
ChemModeling

Augmenting a Compound Collection

- Process vendor collection in same manner as corporate collection
- Produce a lead-like subset
- Compare corporate collection to vendor collection
  - Eliminate any vendor compounds that are within specified cut-off distance of corporate collection
- Cluster remaining lead-like, novel subset
  - Grid spacing for vendor collection often looser than for corporate collection
  - Can also fill-in clusters with low occupancy of corporate compounds
- Select compounds from clusters based on client preferences
  - Preferred vendors
  - Best properties
  - Best price
  - Purity

ChemModeling

Augmentation Can Be Tailored


Fill-in holes in chemistry space
Covered by Corporate Collection
Include areas not covered by original collection

- Select sequentially
  - Preferred Vendors
  - Preferred Targets
- Select based on target
  - Similarity to known actives
  - Privileged substructures
  - Meet pharmacophore model
  - Meet SAR model
- Select based on properties
  - Preferred vendors
  - Best properties
  - Best price
  - Purity

ChemModeling