Building compound archives for the future

Richard Lewis, Joakim Deheuvels, Peter Ertl, Bernard Pirard, Finton Sirockin, Nik Stiefl

Strasbourg Summer School - Strasbourg, 2016

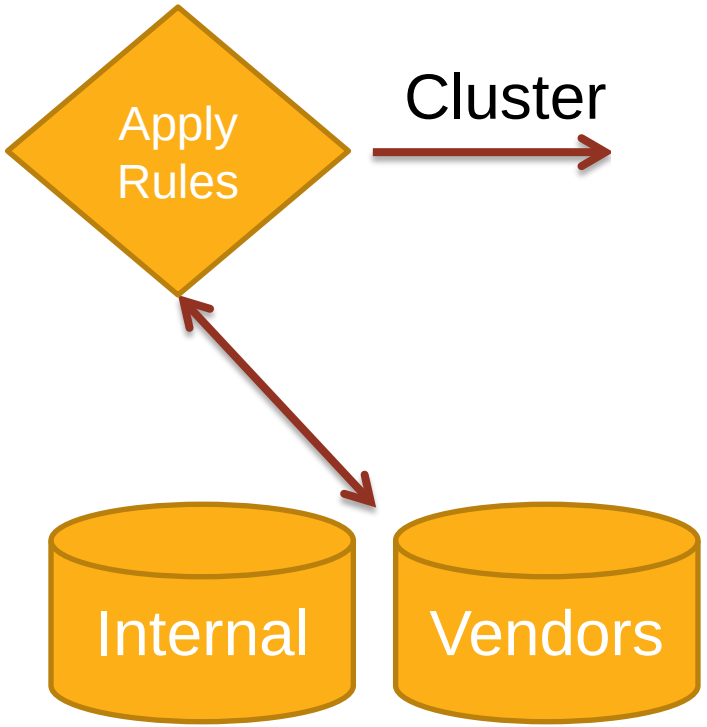
NOVARTIS

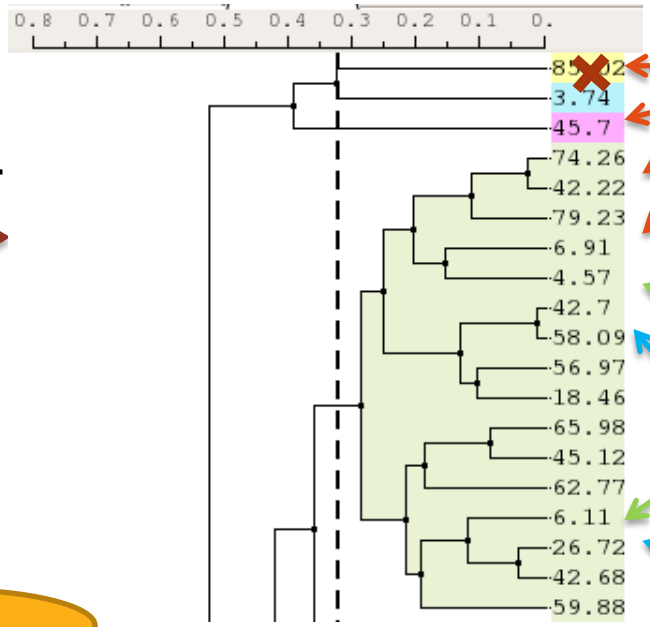
Overview

- Imagine you are building a new screening deck to provide hits for ***new*** biological assays
  - Horvath et al (2014), Design of a General-Purpose European Compound Screening Library for EU-OPENSCREEN ChemMedChem, 9, 2309-2326
- Imagine you are not constrained to vendor catalogues
- Will the targets of the future be covered by the compound libraries of today?
  - A critical review of past strategies
- Understanding protein pockets
- Future directions

Archive selection process

General process





Diversity

Price

Chemical intuition

NOVARTIS

How we designed libraries or bought compounds in the past - Filters

Ugliness

Lilly Score

Rules for Identifying Potentially Reactive or Promiscuous Compounds
R. Bruns and I. Watson *J. Med. Chem.*, **2012**, *55* (22), pp 9763–9772

GW, AMGEN, RPR filters

Aggregation (Shoichet)

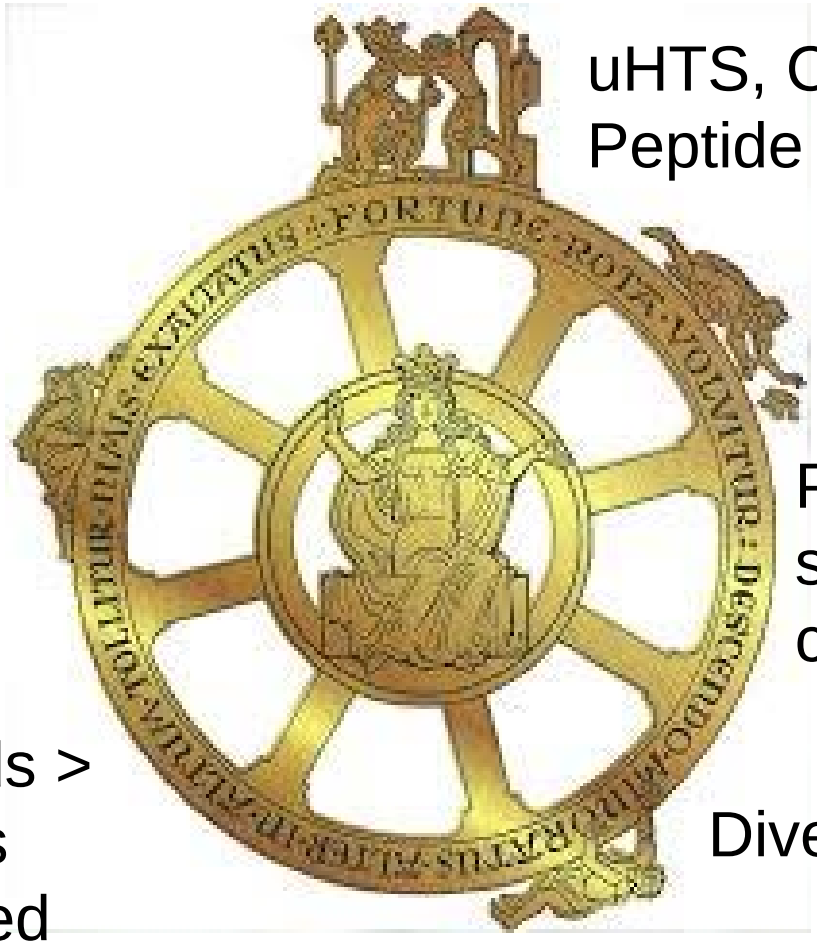
PAINS

Cousin filters

- Don’t make/buy ugly compounds unless following a hit

NOVARTIS

Compound qualities and quantities



uHTS, Comb Chem,
Peptide libraries

HTS –
screen
entire deck

P'Chem profiles bad,
signal:noise bad,
costs high

Compounds >
assay slots
Hand-picked

Diversity sets

- What you ask for is what you get

Presentation Title | Date | Subject | Business Use Only

Physicochemical space

Rule-of-5 limits for oral bioavailability

DPD filters for CMC – like profiles

High LogP strongly linked to promiscuity

'By eye' filters for cellular assays

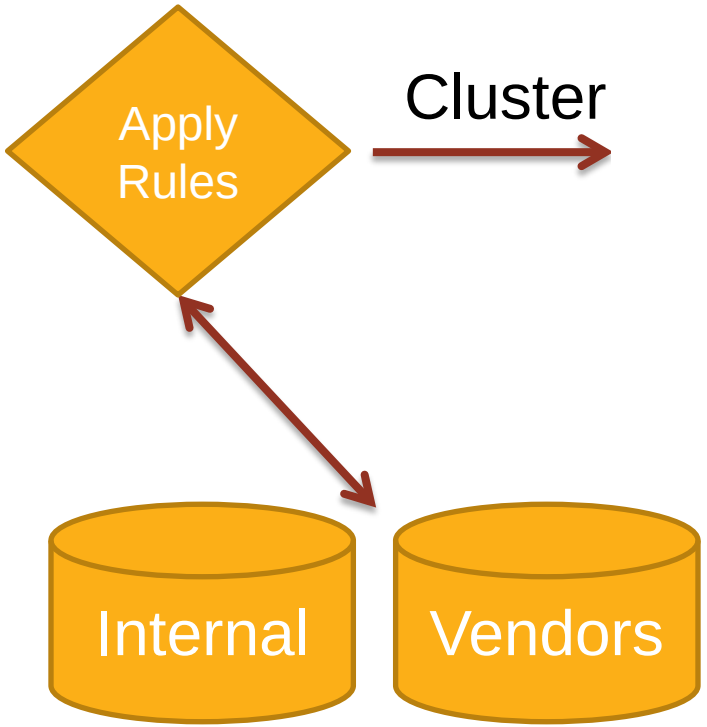
High LogP not strongly linked to promiscuity

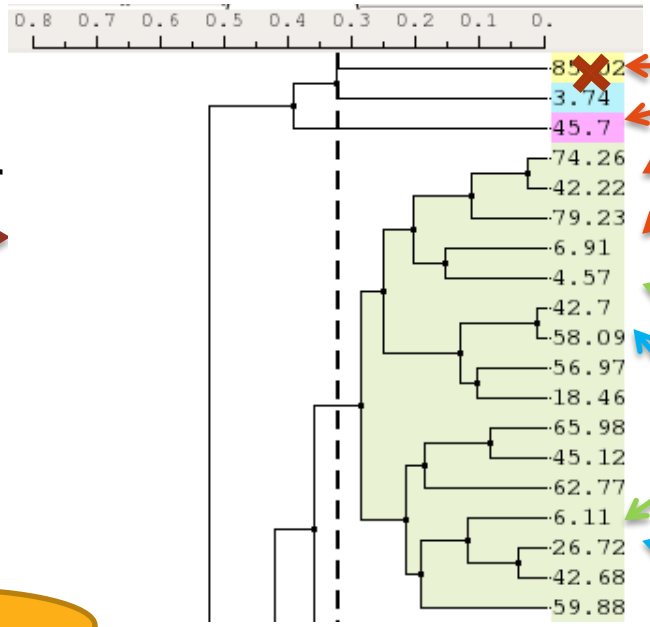
- Everything in moderation but limits can be broken

Presentation Title | Date | Subject | Business Use Only

Archive selection process

General process





Diversity

Price

Chemical intuition

NOVARTIS

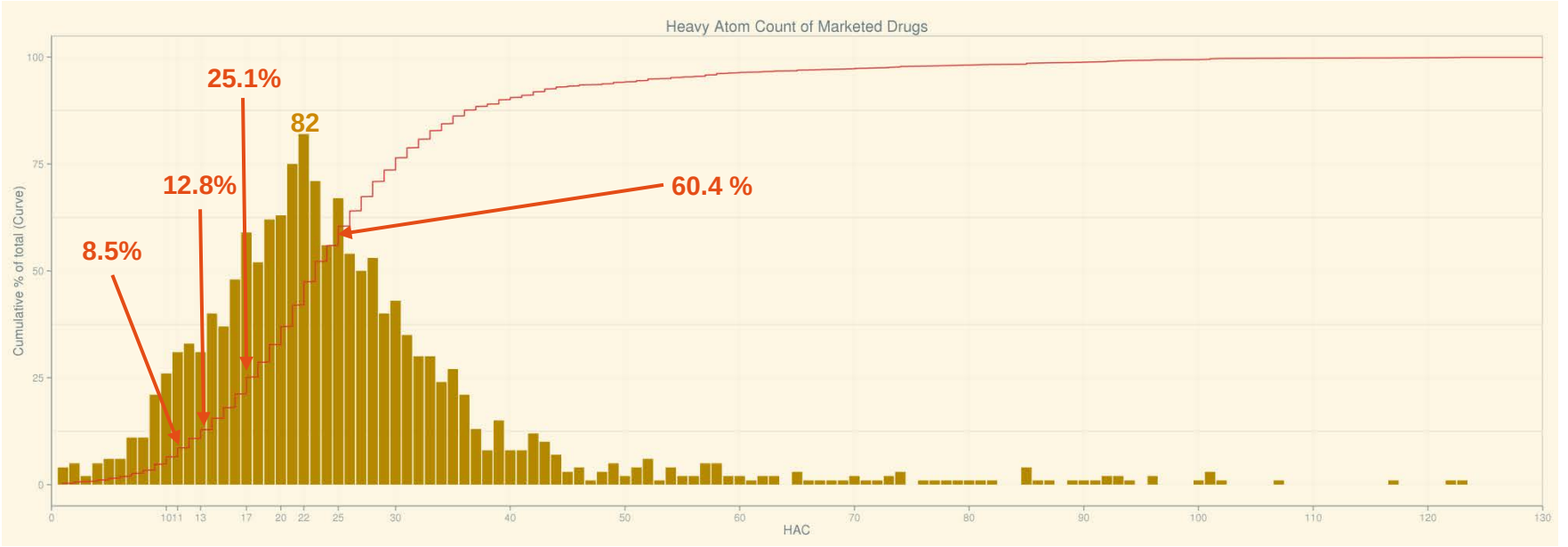
Why are we still failing to find hits?

- Chemical Space is too vast
  - We cannot make everything we can design
  - We cannot store everything we could make
  - We cannot screen everything we could store
  - Current archives represent mostly what we know about (old LO series, known targets)

Dimensions - I

How many molecules can be made with X ''heavy'' atoms ?

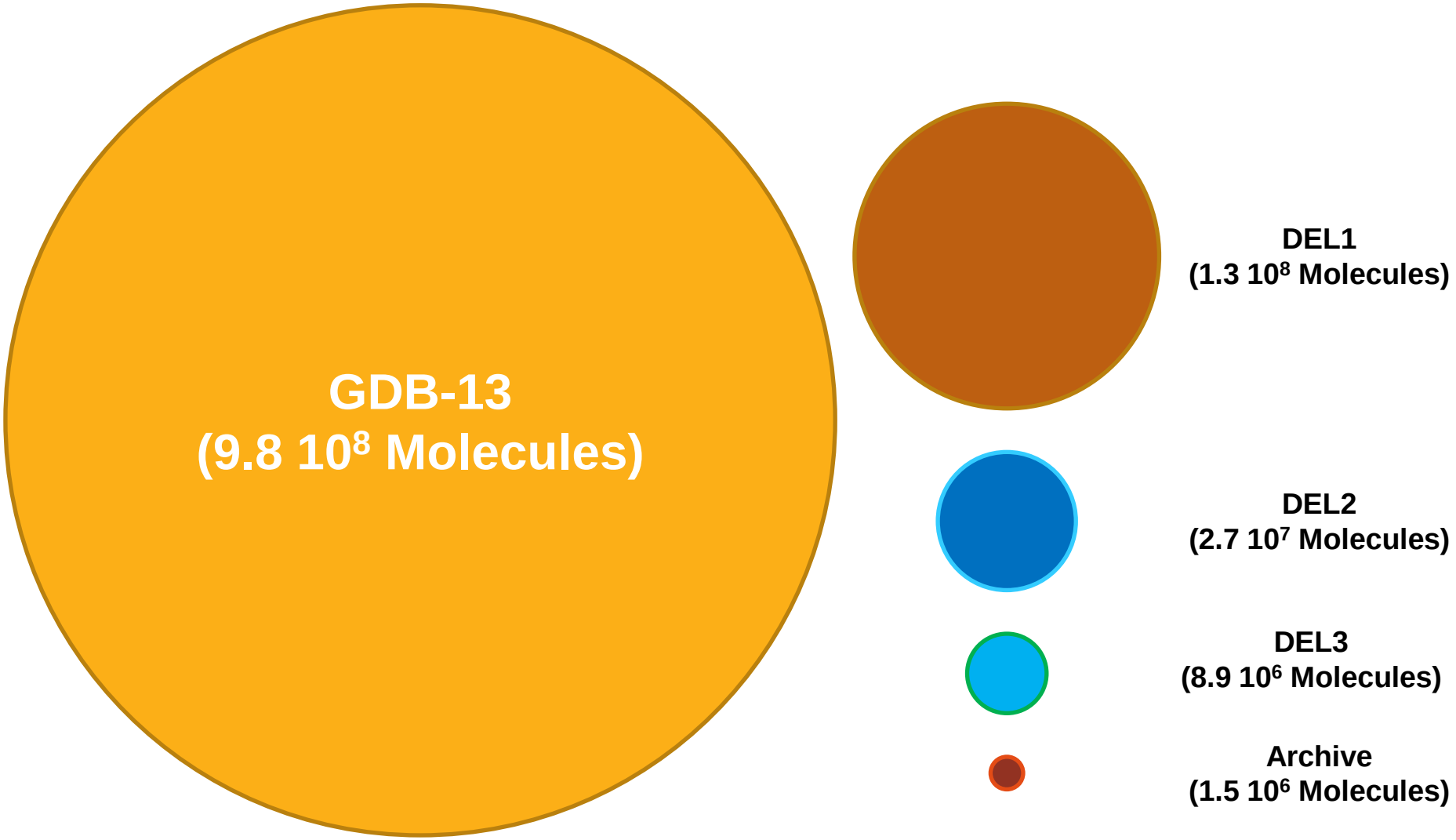
GDB-11 : 2.6 10^{7} molecules with 11 non-H atoms [1] (MW ~ 160)
GDB-13 : 9.8 10^{8} molecules with 13 non-H atoms [2]
GDB-17 : 1.7 10^{11} molecules with 17 non-H atoms [3]
Extrapolation : 1.0 10^{27} molecules with 25 non-H atoms (MW ~ 400)



[1] Fink, T.; Reymond, J.-L. Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery. *J. Chem. Inf. Model.* **2007**, *47* (2), 342–353.
[2] Blum, L. C.; van Deursen, R.; Reymond, J.-L. Visualisation and Subsets of the Chemical Universe Database GDB-13 for Virtual Screening. *Journal of Computer-Aided Molecular Design* **2011**, *25* (7), 637–647
[3] Ruddigkeit, L.; Blum, L. C.; Reymond, J.-L. Visualization and Virtual Screening of the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2013**, *53* (1), 56–65.

Dimensions - II

GDB-13 vs. Novartis Archive vs. In-House DELibraries



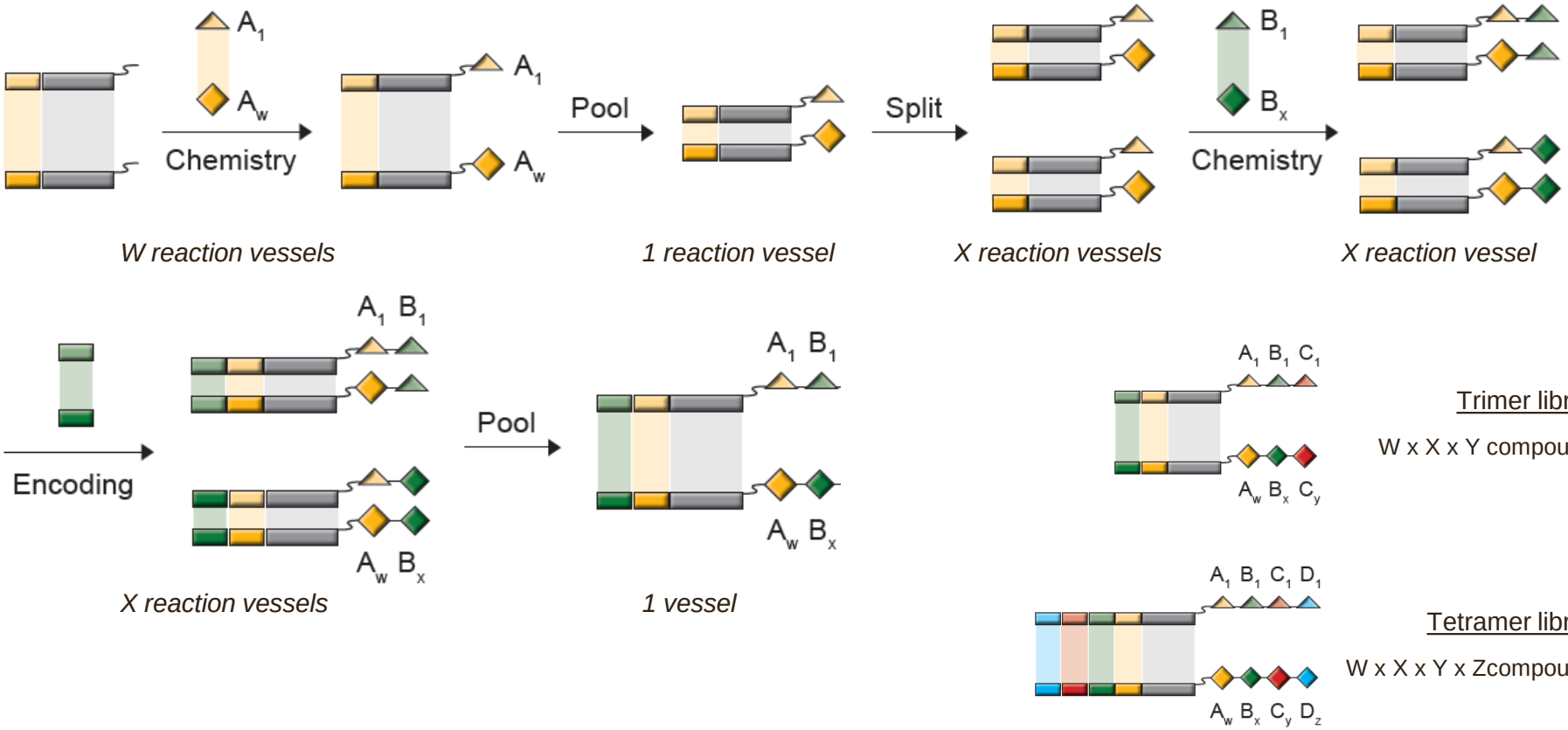
NOVARTIS

Chemical Space approaches

- 2D similarity metrics (based on substructural features) were designed to retrieve similar molecules quickly, not to assess dissimilarity
  - Manley PW, Stiefl N, Cowan-Jacob SW, Kaufman S, Mestan J, Wartmann M, Wiesmann M, Woodman R, Gallagher N., Bioorg Med Chem, 2010;18(19):6977-86. **Structural resemblances and comparisons of the relative pharmacological properties of imatinib and nilotinib.**
- Overwhelmed by the granularity of chemical space
- Driven by what we know about current structures, not future targets
- No real guide to novelty – libraries driven by synthetic accessibility

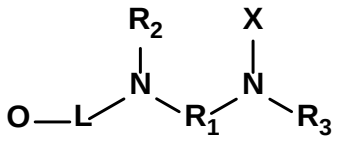
DEL technology uses DNA oligonucleotides to record the combinatorial synthesis of organic molecules...

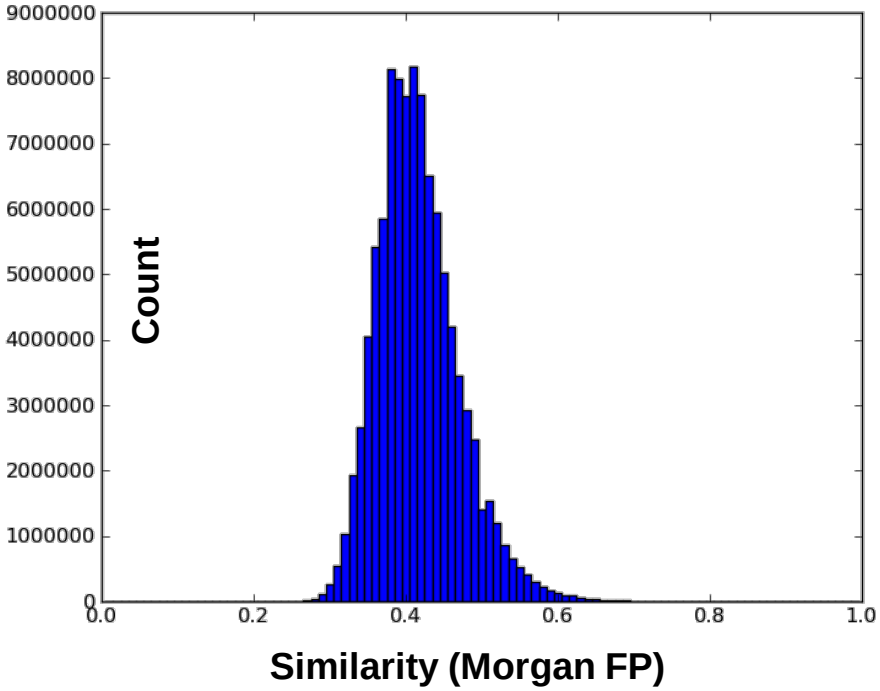
- Dimer library, W x X compounds
  - Pos 1: W building blocks, A_1 to A_W
  - Pos 2: X building blocks, B_1 to B_X



How much does a DEL library contribute?

DEL3 sample similarity distribution vs. NCA



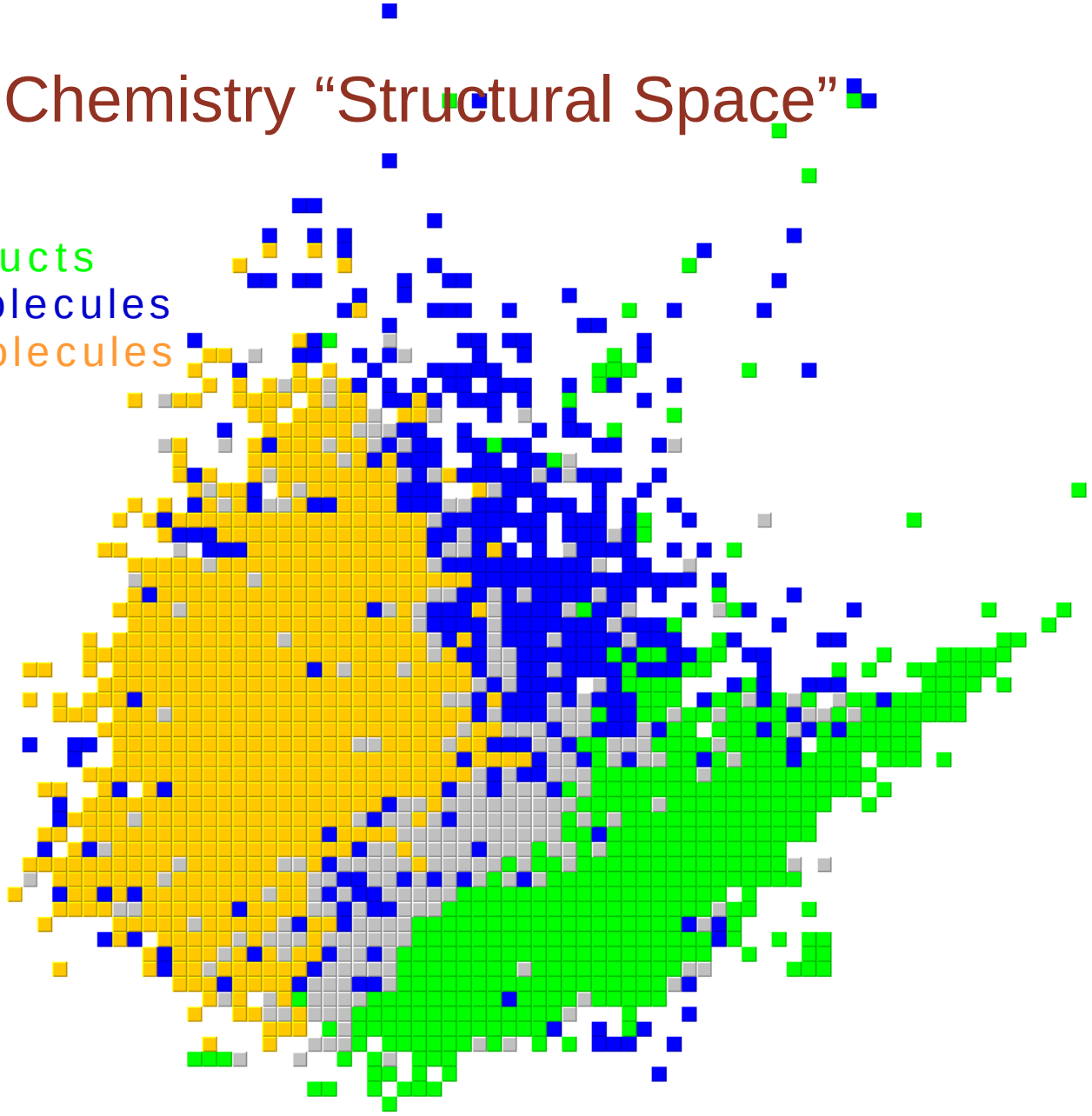


Biological Space - an experimental signpost

- We can classify molecules by their biological fingerprints
  - Petrone, Paula M., Benjamin Simms, Florian Nigsch, Eugen Lounkine, Peter Kutchukian, Allen Cornett, Zhan Deng, John W. Davies, Jeremy L. Jenkins, and Meir Glick. "Rethinking molecular similarity: comparing compounds on the basis of biological activity." *ACS chemical biology* 7, no. 8 (2012): 1399-1409.
  - Can identify which targets/combinations/profiles are well populated
  - And which are not
- Build Bayesian activity models to score new compounds
- Use de novo methods to refine promising proposals
  - Ertl, P., & Lewis, R. (2012). IADE: a system for intelligent automatic design of bioisosteric analogs. *Journal of computer-aided molecular design*, *26*(11), 1207-1215.
- But again is limited to what we have assayed
  - Models score for ‘kinase-like’ not ‘new-target like’

Map of Chemistry “Structural Space”

- Natural Products
- Bioactive molecules
- Synthetic molecules

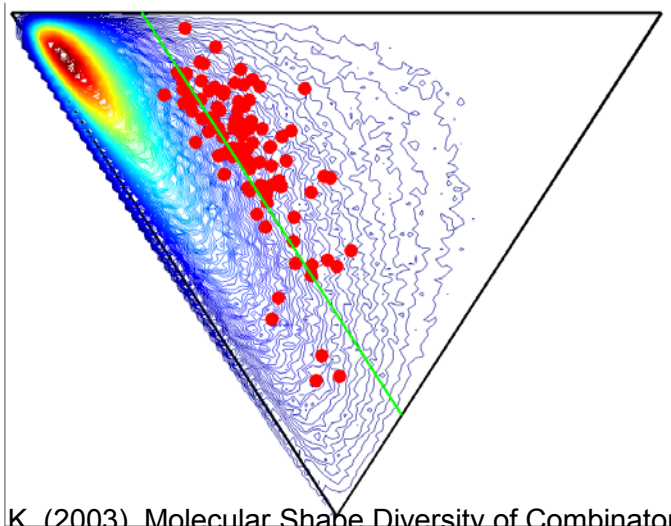


Dark Matter

- Molecules that have shown no activity across many screens
- Useful Pharmacophore but physicochemical properties prevent assay?
  - Signal in SAR model
  - Mark as extrema in physicochemical space
- Pharmacophore that is not sensible in biological space
- Is dark matter clustered with active space?

Shape

- Moments of inertia → shape descriptors



Sauer, W. H. B., & Schwarz, M. K. (2003). Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *Journal of Chemical Information and Computer Sciences, 43*(3), 987-1003.

- But what evidence do we have that diversity in this shape space has any relevance to the future composition of libraries
  - Other than to avoid areas already very heavily sampled?
- This only make sense if one follows the distribution of the space of binding sites – work in progress

Shape 2

- Fraction sp3 has become a popular war cry
  - No more flat heterocycles, long live DOS
  - Even though the analysis was flawed
    - PW Kenny, CA Montanari (2013) Inflation of correlation in the pursuit of drug-likeness JCAMD 27 (1), 1-13
- An in-house analysis did indicate one area where the archive was deficient
- It was also an area where no drugs were found
- Compounds with long linear alkyl chains

Enrichment Drivers: Design vs Accessibility

- There is a slow change from synthetic accessibility being the main driver
  - Metric of cost/compound
  - Acceptance that new chemistry will not be cheap/simple
- Goal of design
  - Reduce the number of targets for which no chemical matter is found
  - Do it in the most compound-efficient way
    - Not by creating mega-archives

So how do we perform such a design?

Filling the holes

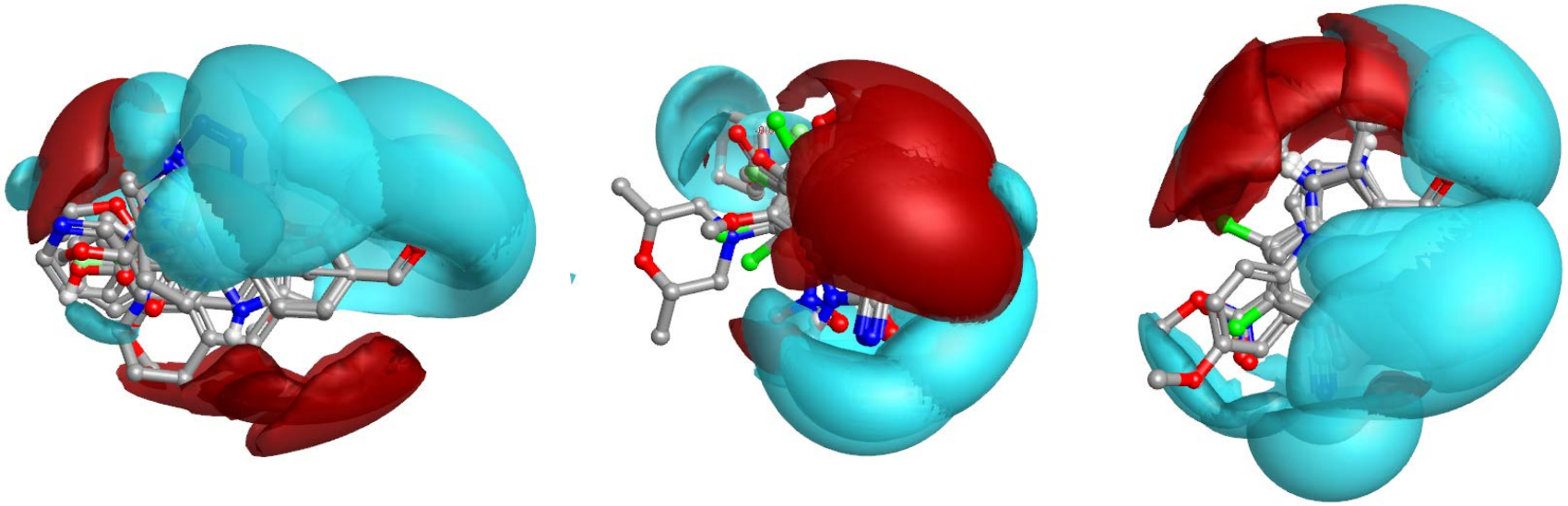
- We have illustrated various metrics
  - Ugliness, 2D fingerprints, P'Chem space, Shape, bioFPs
- You can identify holes in the space with any current compound library
- Does it make sense to fill the holes?
- Or are the holes there for a reason?
- How do you go about it algorithmically?

DELs: Maximise pharmacophore coverage and attractiveness?

Clustering capping groups that distribute similarly PH4 in surrounding space

- Pharmacophore clustering of all candidates using an algorithm developed in collaboration with Cresset

3 example clusters from the processing of 1832 aldehydes currently in stock



NOVARTIS

Frequency Fingerprints

von Korff, M.; Freyss, J.; Sander, T. Flexophore, a New Versatile 3D Pharmacophore Descriptor That Considers Molecular Flexibility. J. Chem. Inf. Model. 2008, 48 (4), 797–810.

- Capture the feature, its geometry and its relative abundance in bins
- Allows operations on large datasets where N*M comparisons would be infeasible
  - Union (for library proposal)
  - Difference (to exclude populated areas/dark matter)
  - Intersection (to compare designs)
  - Histogram comparisons for distance metrics

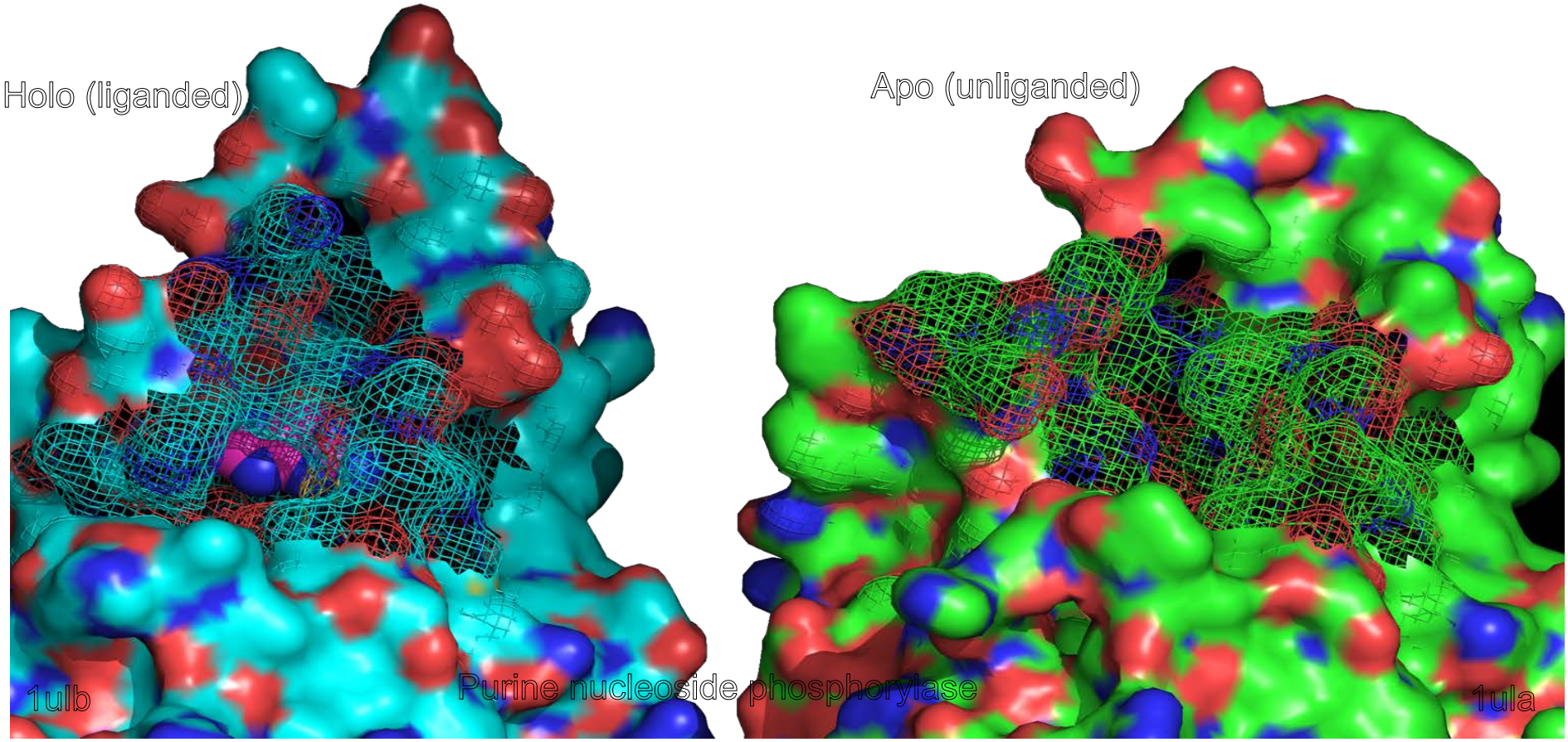
Basic Assumptions

- Ligands exert their effects by
  - Binding to targets
    - (Assuming that they can get to their targets)
- Ligand sites (binding pockets) can be identified
  - Assume that there is a finite repertoire of pockets
  - Druggability vs ligandability of pockets
    - We concentrate on ligandability
      - How much $DG_{binding}$ is possible
    - Build in druglikeness during optimisation
- Build a database of ligand pockets
- Identify which pocket families have cognate ligands, which do not.

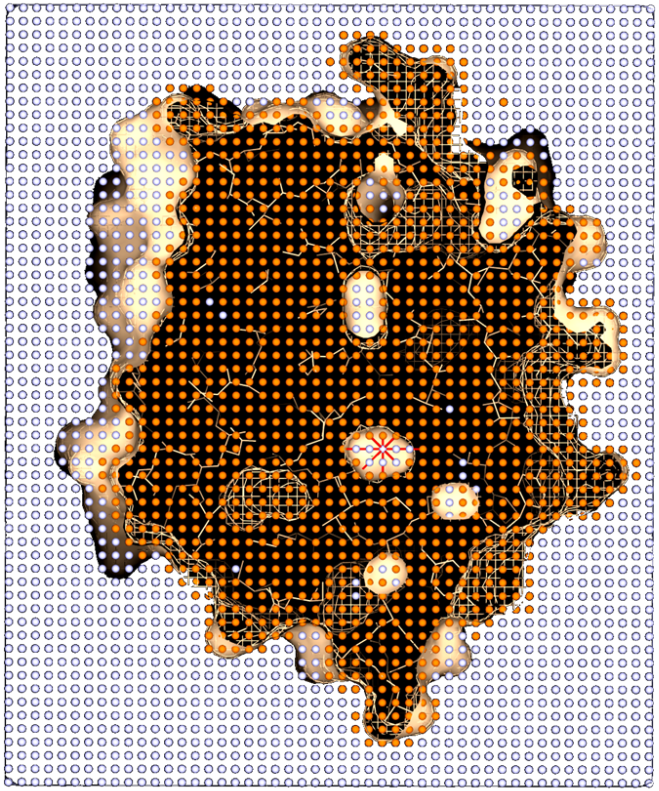
Protein pockets

- A less granular universe?
  - Are the number of protein pockets/pharmacopores more limited in scope than chemical space?
    - Abdullah Kahraman, Richard J. Morris, Roman A. Laskowski, Janet M. Thornton (2007) Shape Variation in Protein Binding Pockets and their Ligands J. Mol Bio, 368, 283-301
    - Mason, J. S., Morize, I., Menard, P. R., Cheney, D. L., Hulme, C., & Labaudiniere, R. F. (1999). New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *Journal of medicinal chemistry*, *42*(17), 3251-3264.
  - What architectures are well-covered?
  - What pockets/cryptic pockets are poorly covered?
  - Can we distinguish pocket space from dark matter space?

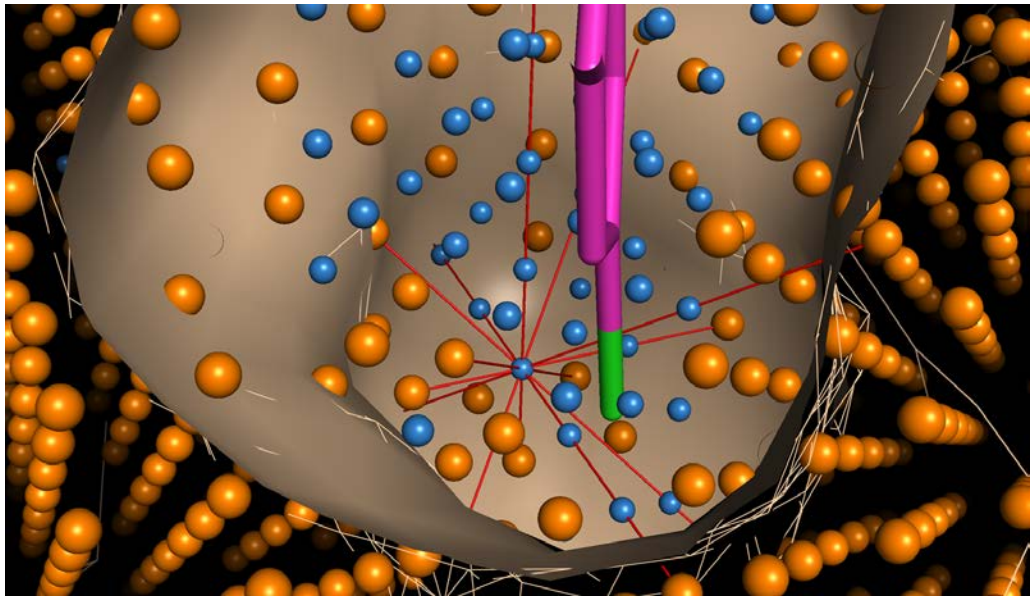
Protein cavities to Fingerprints



Ligsite



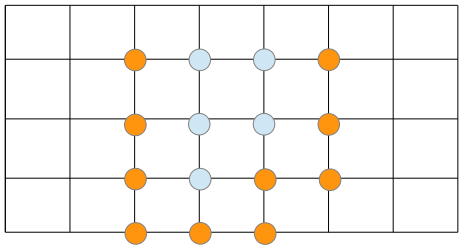
7 axes: X, Y, Z + 4 cubic diagonals

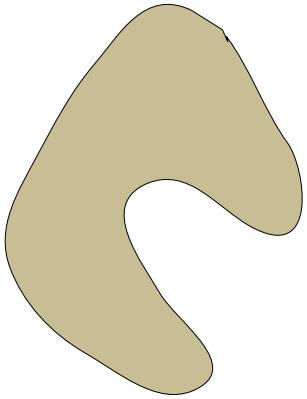


Protein – Solvent - Protein

Identification :
Modified LIGSITE algorithm

Grid orientation



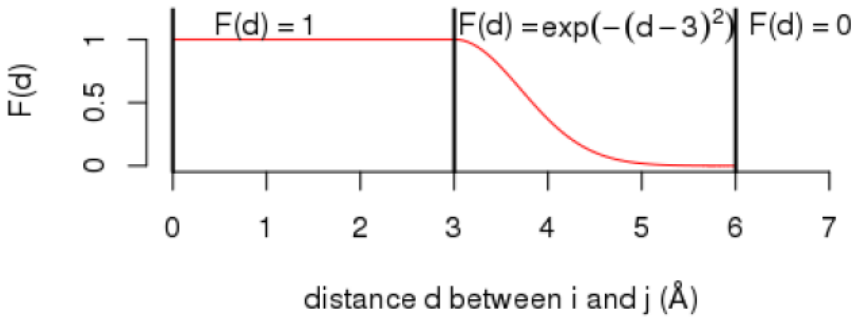


P-S-P ==> P-S < 5.5 Å

14 levels of buriedness

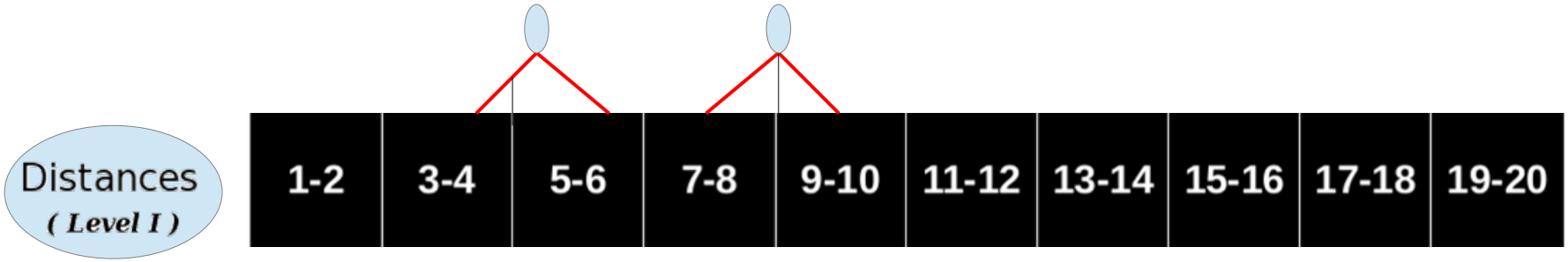
Buriedness > 8 => pocket

$$PW(i) = \sum_{j=1}^{Nprobes} Buriedness_j * F(d_{ij})$$



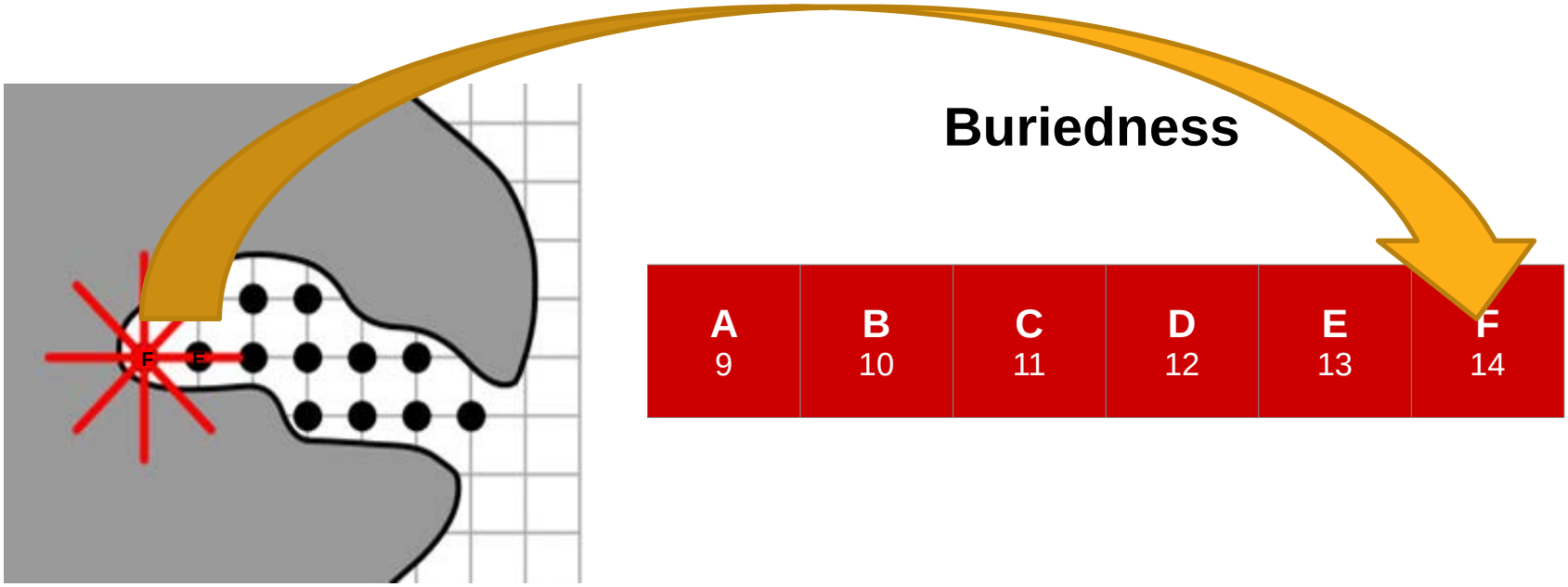
NOVARTIS

Fingerprint



10 distance bins with fuzzy membership

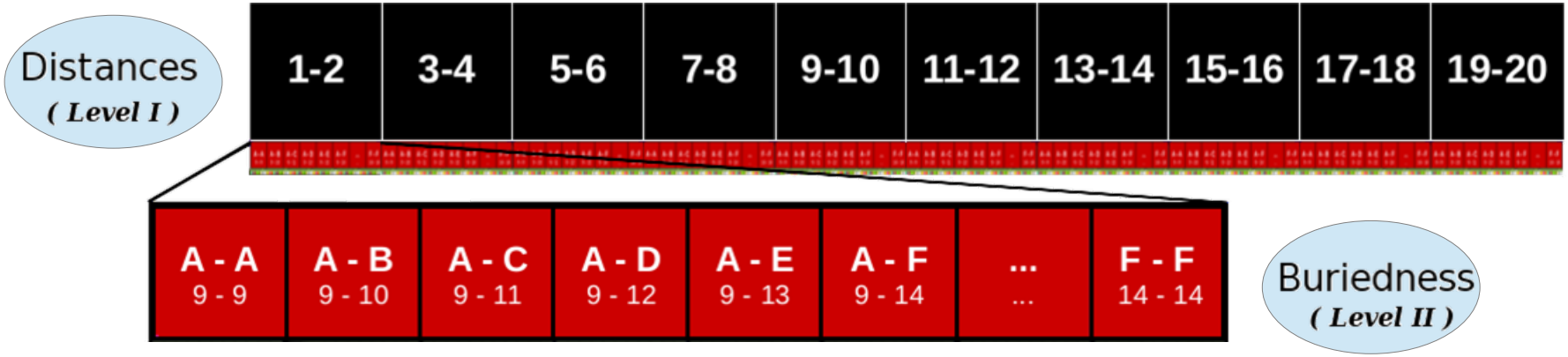
Description : Geometrical features



Weisel et al. Chemistry Central Journal 2007 1:7

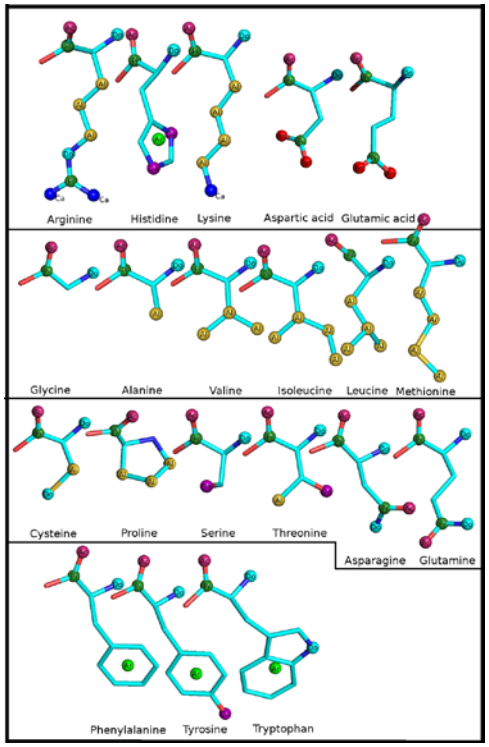
Fingerprint

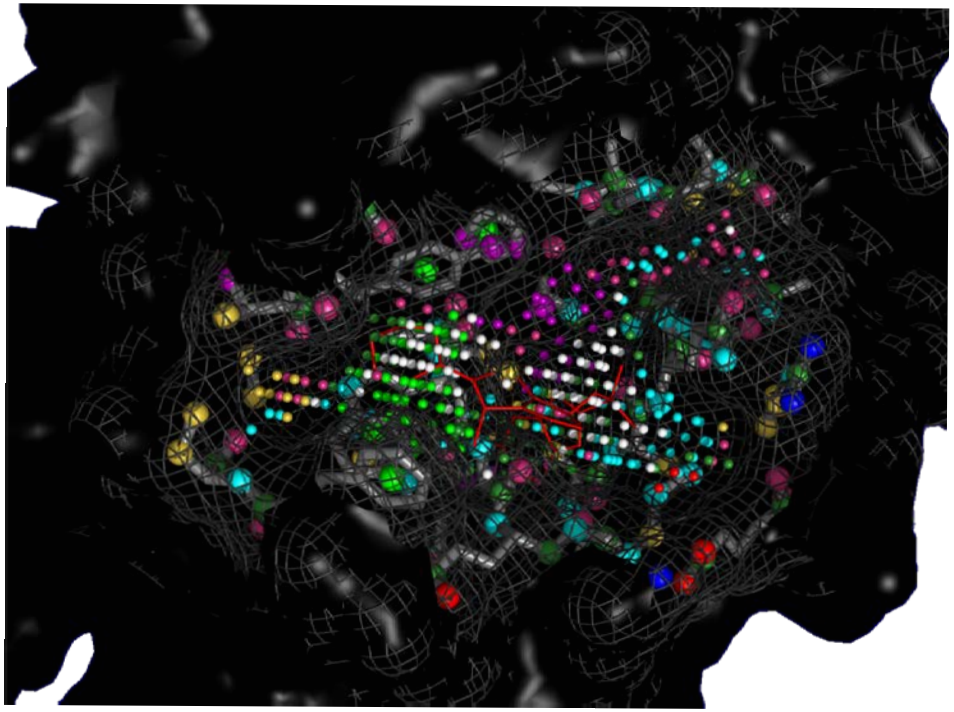
Distance + buriedness



10 X 21 bins

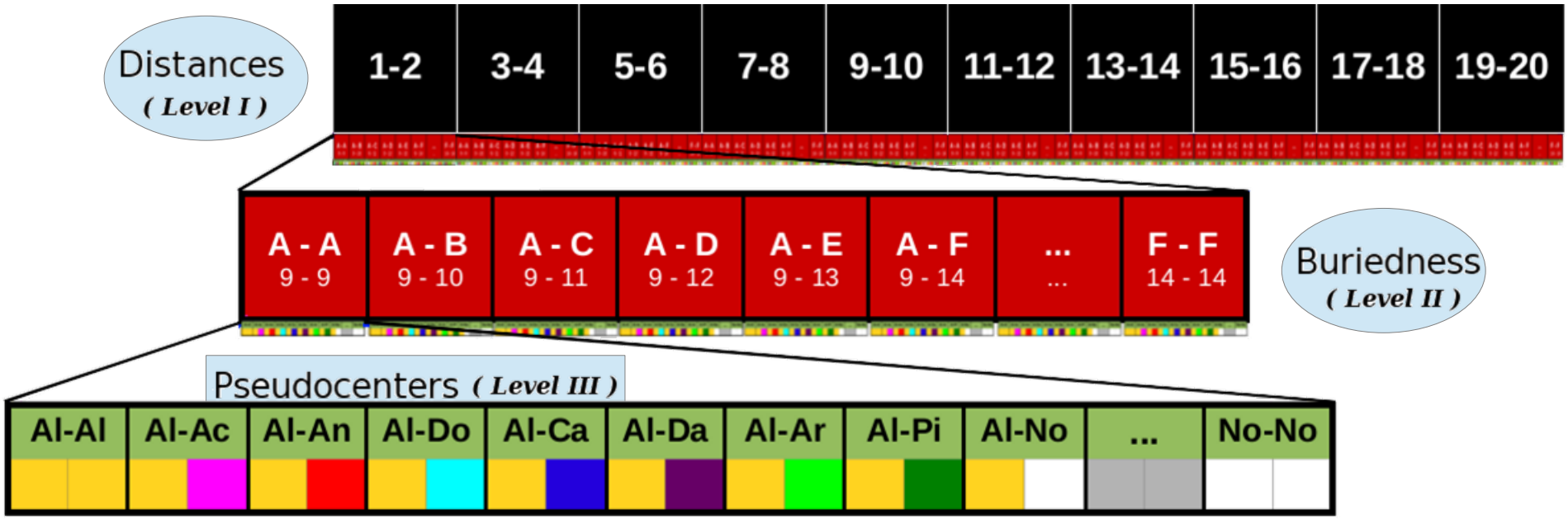
Description : Physico-chemical features





Pseudocenter types for the 20 standard amino acids

Complete pocket Fingerprint

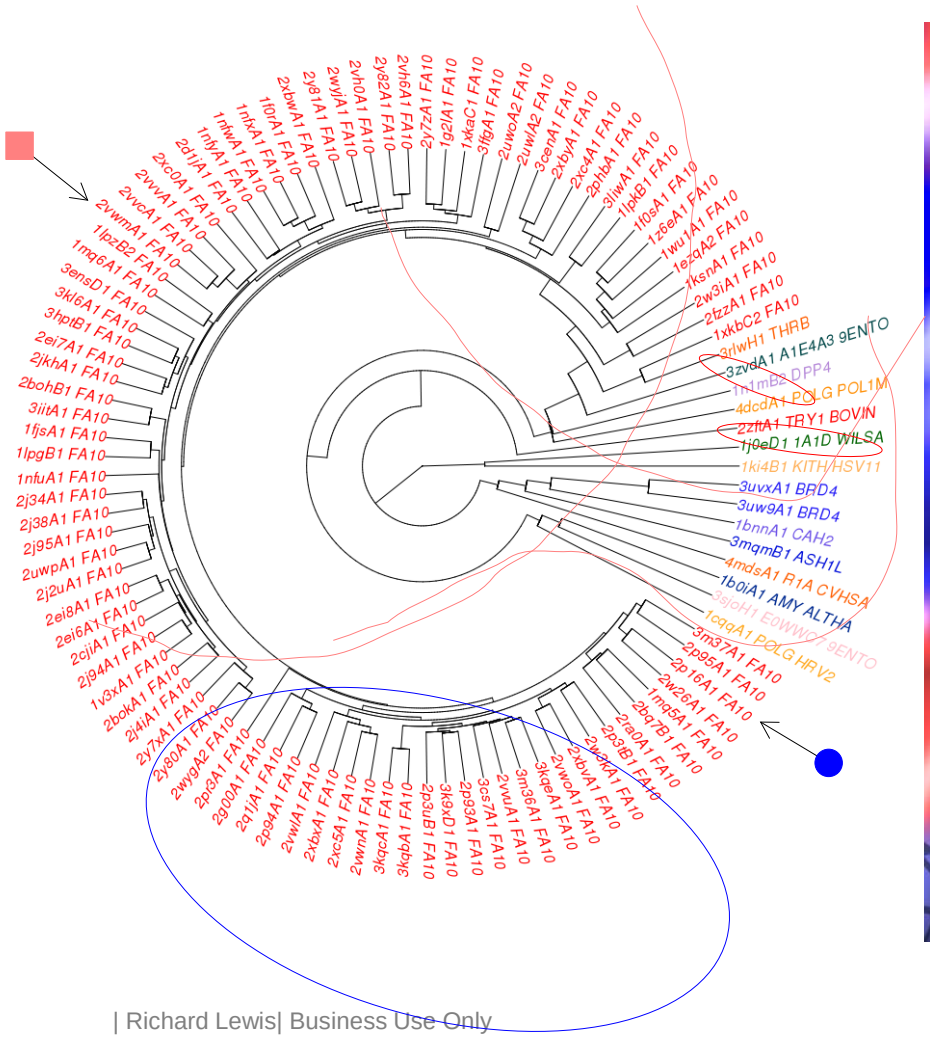


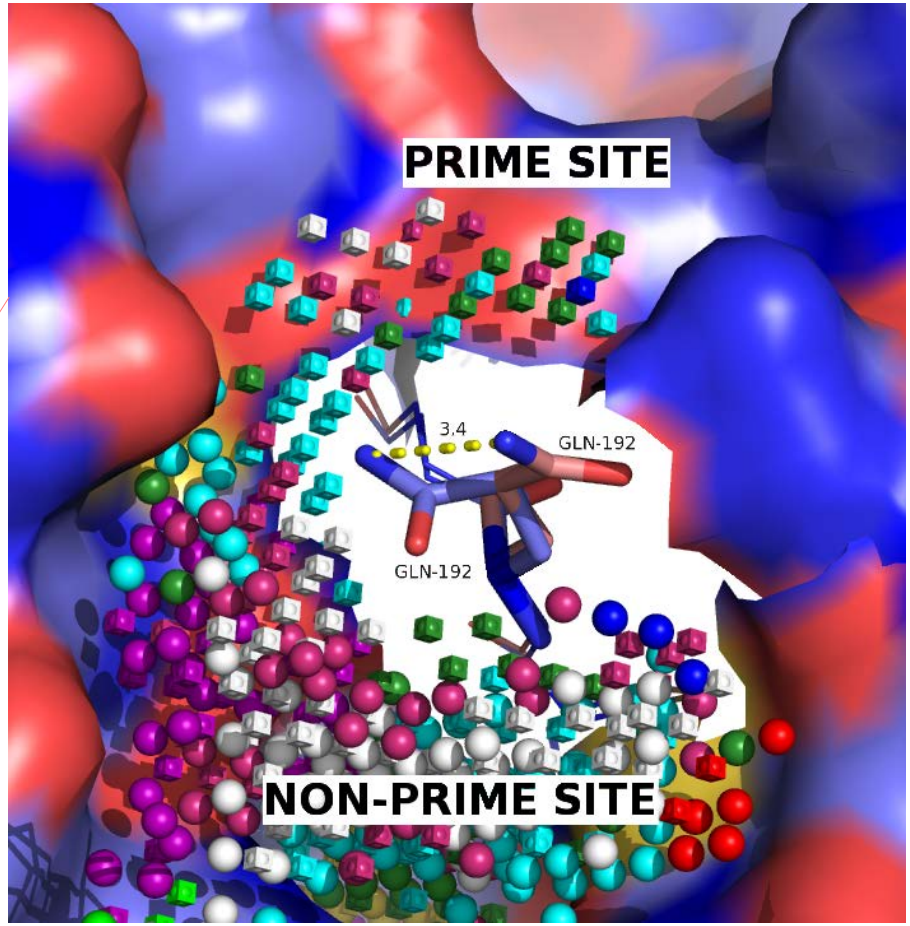
10 X 21 X 45 = 9450 bits

Similar to the 4-centre pharmacophore key but less sparse

Validation : FXa, a S1 serine protease

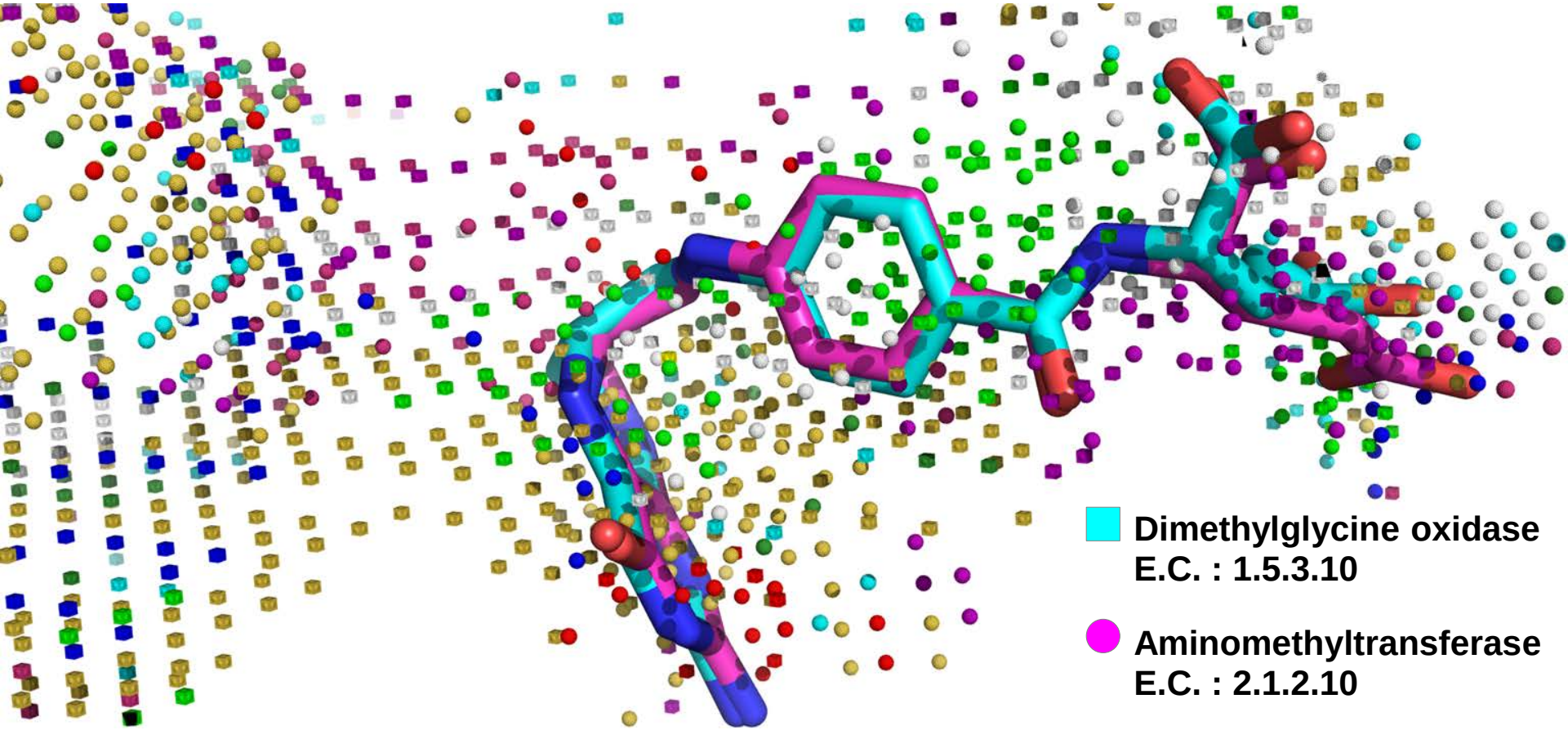
Does the FP retrieve pockets with similar function?





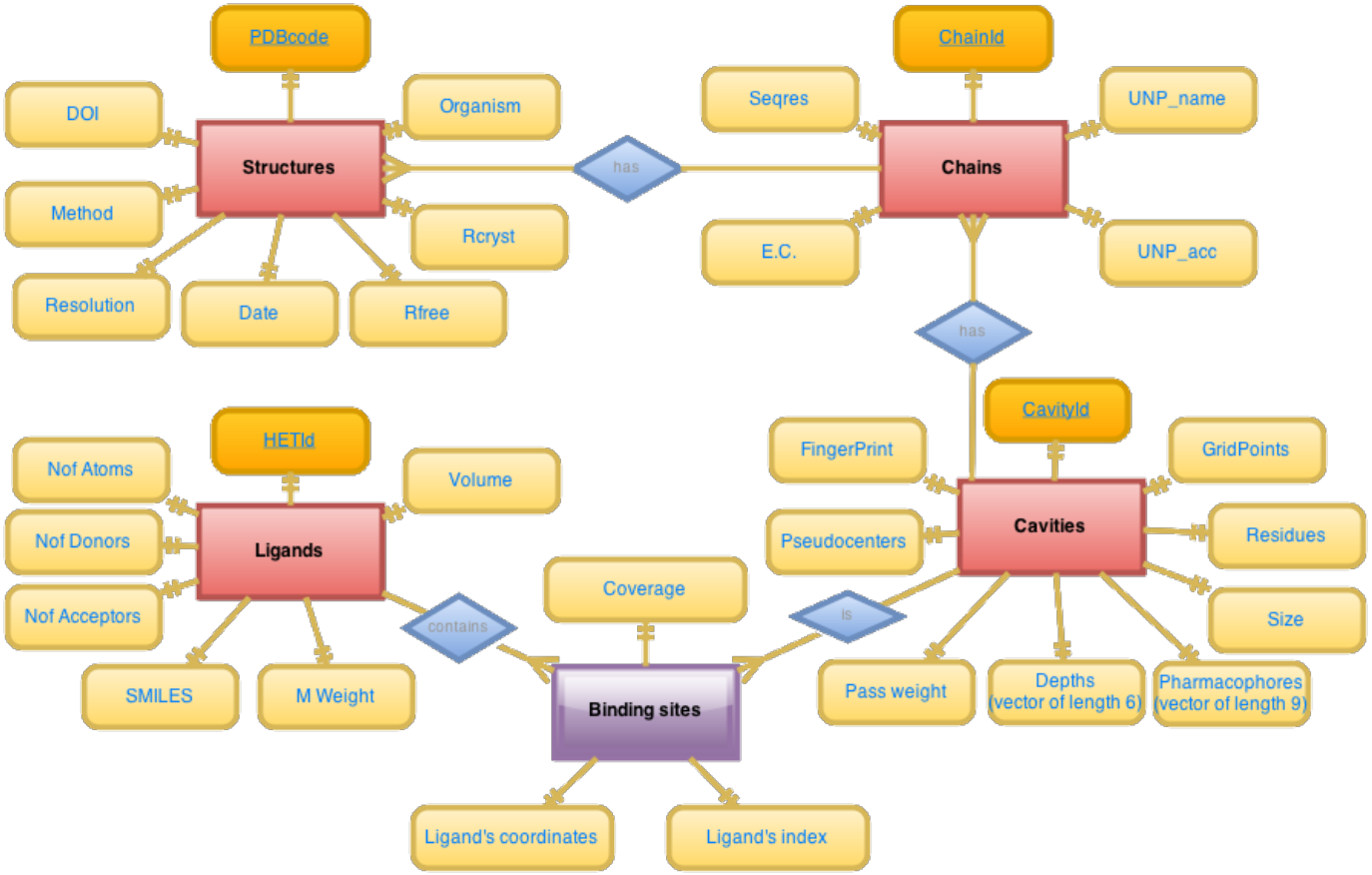
NOVARTIS

Are pockets conserved between unrelated proteins?



NOVARTIS

Pocket database satistics



- 86'279 structures
- 811'922 cavities

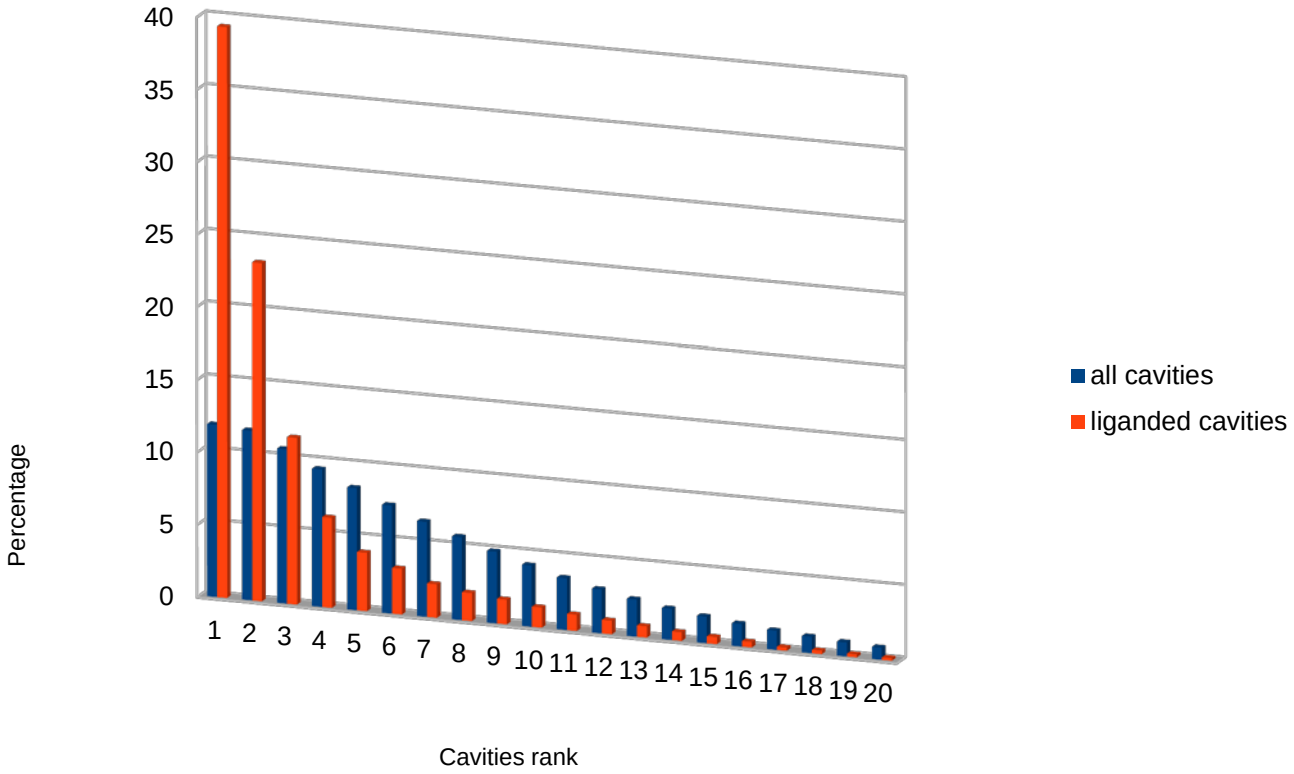
- 2 hours to return best hits from fingerprint screening

NOVARTIS

Database : statistics

Average pocket size (number of grid points):

all 175 **liganded 350**



NOVARTIS

Work in progress

- We have extended the shape analysis map
  - There are regions of our map that are populated by new chemistries (macrocycles)
  - And regions where we have nothing
  - Looking at the inertial axes of pockets for Sauer shape plot
- Looking to map pocket and ligand space together on a single common fingerprint
  - Weighting for partial matches
  - Allowance for pharmacophore colouring
  - Already using the shapes of empty pockets to search for ligands
- Use the pocket fingerprint to drive the search for poorly populated pockets
  - Correlate with hit rates in HTS?
- Sensitivity to conformation
  - Initial work suggests reasonable robustness

Conclusions

- Early enrichment efforts were driven by abstract considerations of diversity
  - Anchored in what we knew, not in what we were trying to discover
- Diversity through Chemical space to too vast to cover in any archive
- Pocket space offers a more tractable and relevant space
- We have succeeded in the first steps of producing a viable FP with enough granularity to be useful
- Each new x-ray structure will give us a map of terra incognita for pocket-based measures of enrichment and diversity

Acknowledgements

- CADD colleagues at Novartis Basel
- Structural biologists everywhere