cresset

smarter chemistry | smarter decisions™

Upcoming Features and New Horizons

Mark Mackey

What's new at Cresset

- New features
  - XED FF3
  - Activity cliff and disparity analysis
  - GPU acceleration
- New Horizons
  - Protein similarity
  - Off-target effects

XED FF3

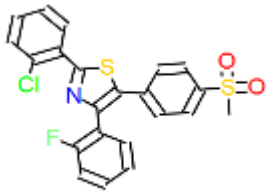
XED FF3

- Evolution, not revolution
- Huckel-based parameter scaling (since 1994!)
- Nitrogens
  - No separate types for Ntri and Nsp3
  - Parameters are scaled based on calculated bond orders
- Halogens
  - Improved fields around Cl, Br, I
- Additional elements
  - Limited support for B, Si

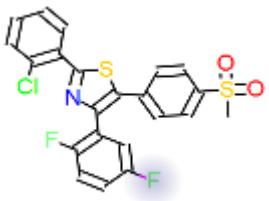
Disparity Analysis

Disparity Analysis

- Useful tool to investigate SAR
  - Which structural changes have had the biggest effect on activity?
  - Disparity = (Δactivity)/(1-similarity)



8



4

High similarity

High Δactivity

=> High disparity

Disparity Analysis

- Useful tool to investigate SAR
  - Which structural changes have had the biggest effect on activity?
  - Disparity = (Δactivity)/(1-similarity)
- Widely used technique
  - Reinvented more recently as “activity cliffs”
- Always been looked at from a 2D point of view
  - MCS or 2D similarity

3D Disparity Analysis

- Issue – what is the 3D similarity of a pair of molecules?
  - Context-specific – aligned to common reference
- Calculate similarity matrix in field (or field+shape) space
  - $O(N^2)$, but fast enough to be usable on 100's of mols
- Demonstration this afternoon

Disparity analysis

> Appearing in Forge and Torch in Q3

> Still under development

> More features to come

GPU acceleration

GPU processing

- Graphics cards are now extremely powerful (but limited) parallel processors
- Harness this to accelerate field comparisons
- Using OpenCL
  - Industry standard
  - Multi-vendor
  - Available on CPU as well

Complicated algorithms…

$$S = \alpha \sum_{fields} w_f \frac{\begin{array}{r} w_{A\to B} \sum_{fp_A} \left(\sqrt{scale\left(size(fp_A)\right) \times F_B\left(position(fp_A)\right)} - constraint(fp_A) \right) \\ + w_{B\to A} \sum_{fp_B} \left(\sqrt{scale(size)(fp_B) \times F_A\left(position(fp_B)\right)} - constraint(fpB) \right) \end{array}}{\begin{array}{r} \sum_{fp_A} \sqrt{scale\left(size(fp_A)\right) \times F_A\left(position(fp_A)\right)} \\ + \sum_{fp_B} \sqrt{scale(size)(fp_B) \times F_B\left(position(fp_B)\right)} \end{array}}$$

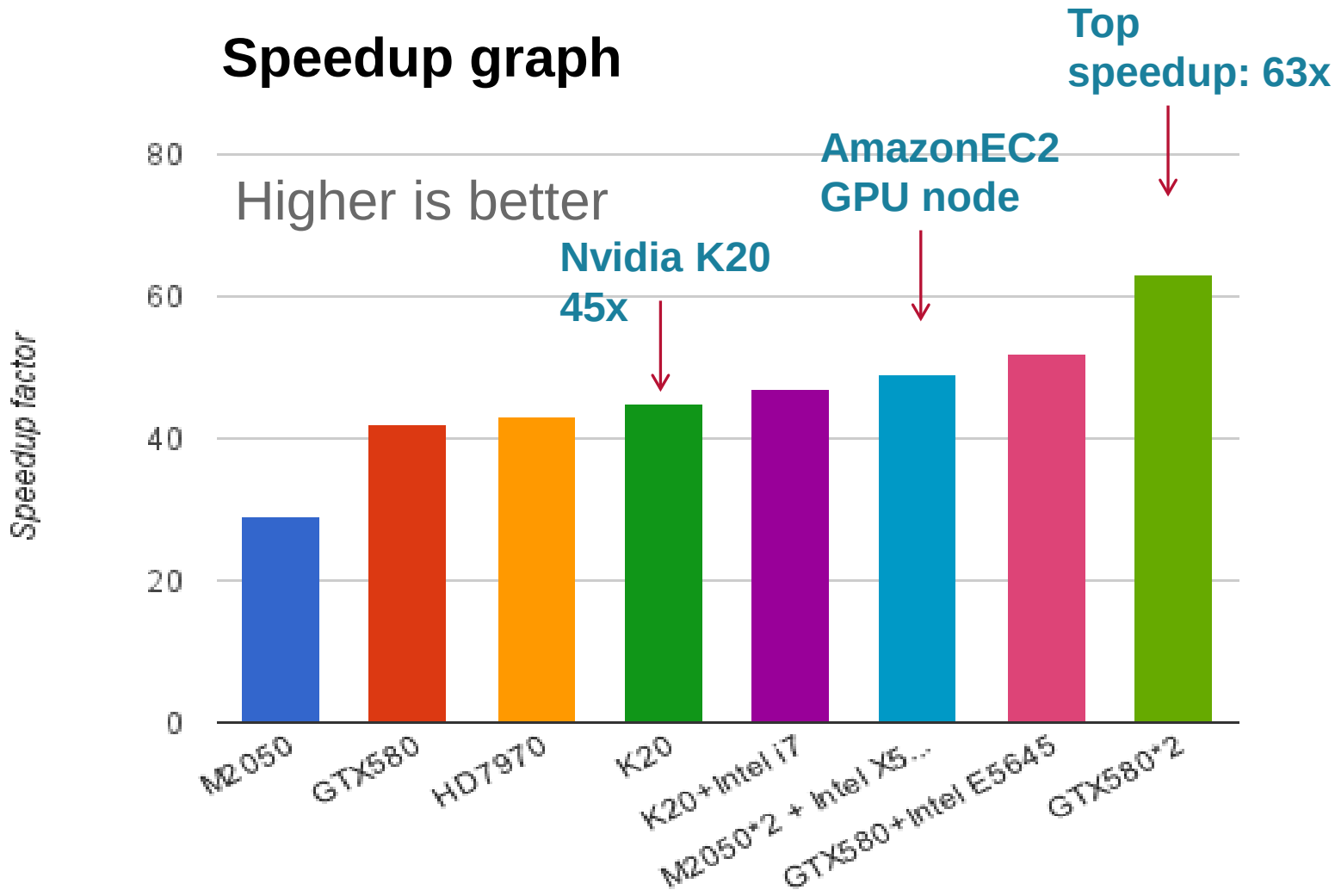
$$+ (1-\alpha) \frac{2V_{AB} - \tau \times constraint(AB)}{V_{AA} + V_{BB}}$$

(and I've left some bits out)

blazeV10 GPU benchmark

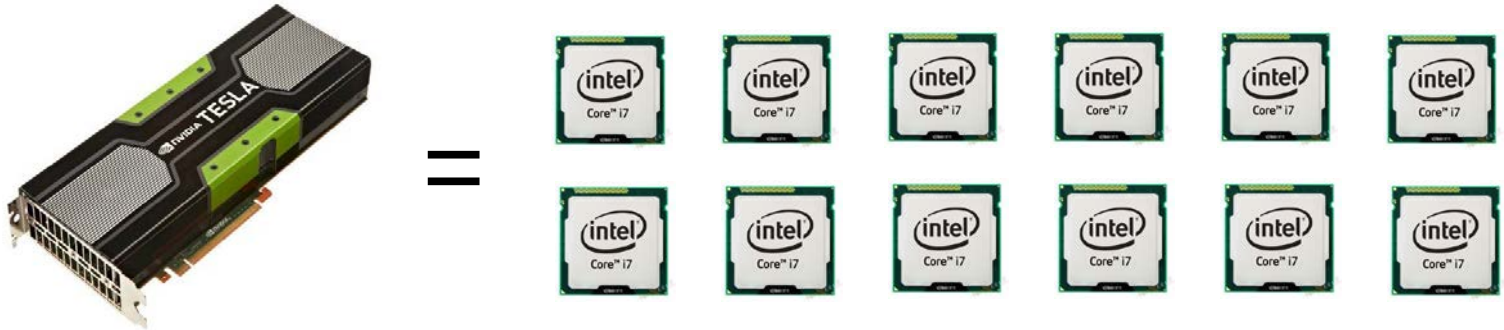
- ~1000 molecules – 80k conformations
- Standard instance: 12 conformations processed per second on a single core of Intel® Core i7-3770 CPU @ 3.40GHz
- CPUs:
  - Intel® Core i7-3770 CPU @ 3.40GHz (4 cores - 4 threads)
  - Intel® Xeon CPU X5570 @ 2.93GHz (4 cores – 8 threads)
  - Intel® Xeon CPU E5645 @ 2.40GHz (6 cores - 12 threads)
- GPGPUs:
  - NVIDIA GTX580
  - AMD HD7970
- HPC GPUs:
  - NVIDIA M2050
  - NVIDIA K20
- CPUs and GPUs will work together, it's a heterogeneous world!

blazeV10 GPU benchmark



Saving you space and money

> **Less hardware**: blazeV10 GPU is ~45 times faster on a K20 than blazeV10 on a single core of Intel i7 CPU!



> **It's cheaper**: for a $**2.10/hour GPU instance on AmazonEC2 you can process 2m conformations, whereas you can only process 1.3m conformations with 14 dual-core High-CPU Medium instances.**

> **It's greener**: one GPU consumes ~400watts, one quad-core CPU workstation consumes ~200watts: we achieve 5x performance per watt = 5x less gCO2 consumed per answer.

Plus new science!

- Diversity analysis in 3D with fields
  - $O(N^2)$ calculation – not currently cost-effective on large data sets
- Screening virtual databases
- Flexible overlays

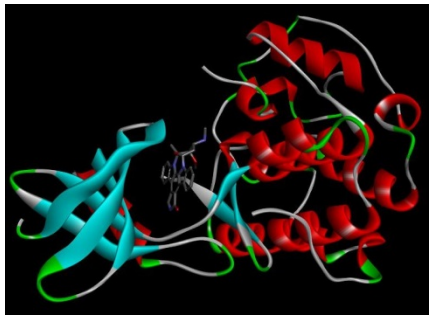
Protein Similarity

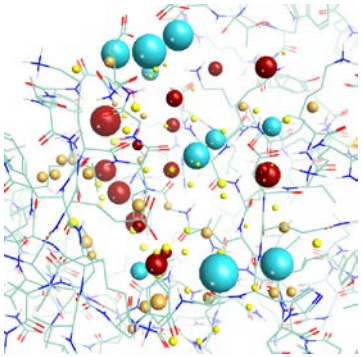
Fields on Proteins

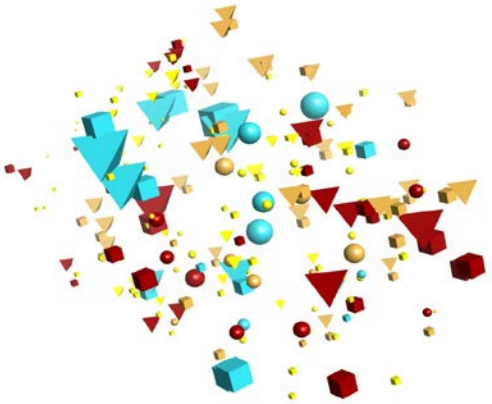
- Can we do (sensible) analysis of proteins in field space?
  - Issues of charge, solvation, calculation time, etc
- Pilot project: active site similarity metrics
  - How similar are two proteins?
  - Which proteins should you be routinely counter-screening against?

Protein Fields

Reference Protein 1AQ1



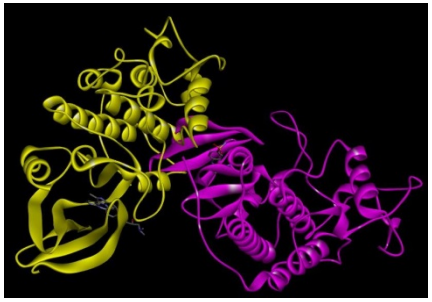


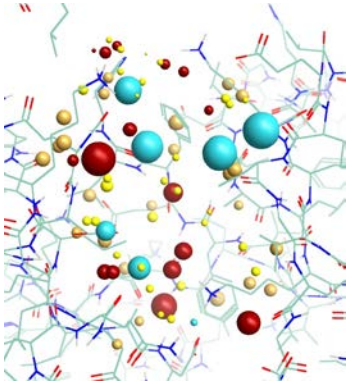


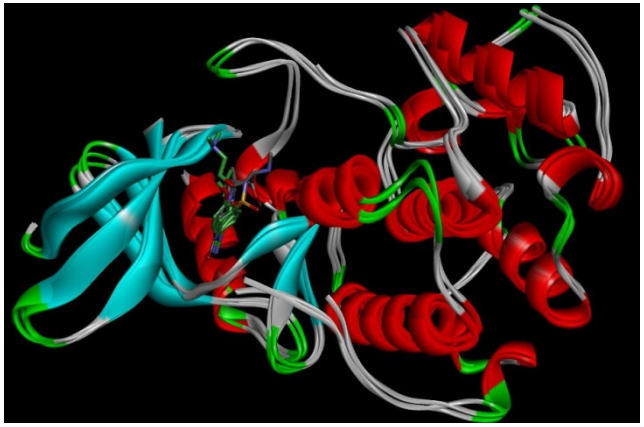
Generate Protein Fields

Align Field Points

Target Proteins 1JSV & 1OIR







Active site similarity

Can we align the active sites of proteins bound to the same ligand?

Can we align the active sites of the same protein bound to different ligands?

Can we distinguish functionally-related proteins from the rest?

Can we differentiate proteins involved in known OTEs from the mass?

Off-target effects

Can we predict OTEs via field similarity?

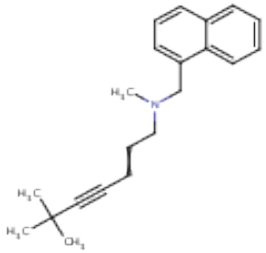
> Several methods looking at this in 2D (eg SEA)

> Field similarity might find ones that 2D similarity doesn't

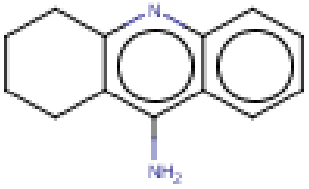
Initial study

- Compare new ligand to ligands from PDB
  - Inverse Blaze
- Raw scores not useful
  - Phe gets a decent score to everything
- Convert to Z-scores

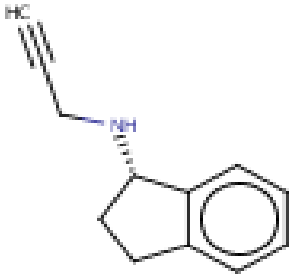
Initial results



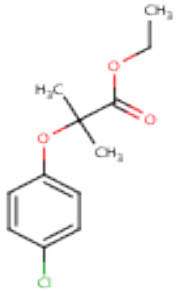
Terbinafine
OTE: MAO & AChE



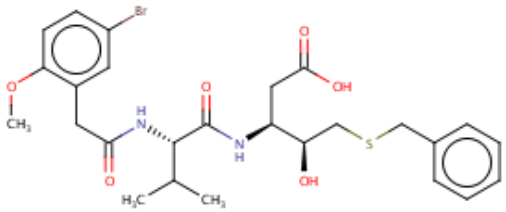
Imipramine
AChE



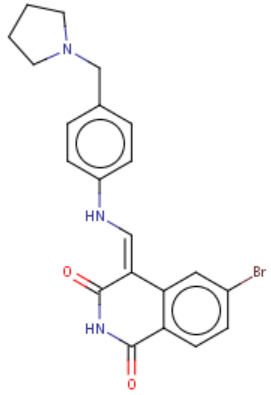
MAO B



Clofibrate
Caspase-1 & tyrosine kinses



Caspases



IGF-1R

OTEs

- Currently extending the method to conformer populations
  - More computationally expensive
  - Harder

What next?

Coming soon…

- More protein fields
  - Protein-protein comparisons
  - Protein-ligand comparisons
  - Using protein info to guide ligand alignments
- Qualitative models
- Extend off-target models
- ADMET models