In silico Compound Screening for Drug Discovery in the "Cloud"

03/09/2010

Gerald J. Wyckoff, Assoc. Prof, UMKC

Working with Vassa Informatics

- Founded (2007) to develop information-theory based technologies that would provide additional tools for scientists studying the functional effects of differences between similar nucleotide and amino acid sequences
- BioVassa was the initial result of this work
  - Collaborations with Indiana University-Northwest, Washington University - St. Louis and the University of Chicago further refined BioVassa
- ChemVassa applies information content analysis to chemical sequences of arbitrary length with an eye towards small molecule screening for drug discovery
  - Development and initial proof of concept work complete

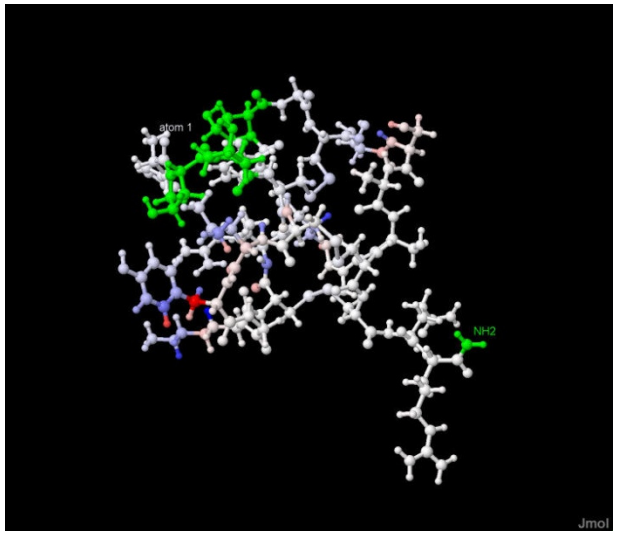
UMKC

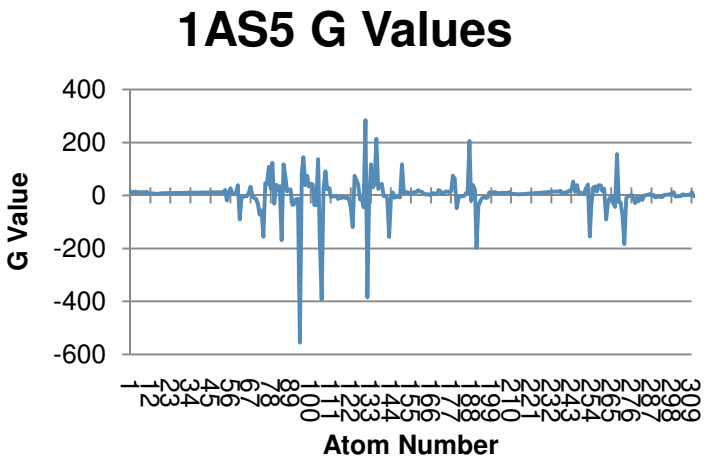
VASSA INFORMATICS

Information Content Overview

Information content is a measurement of a unit's (e.g., a small molecule) compressability versus a theoretical maximum. Units with high information content are not as compressible as those with low information content.

- **We can present this information graphically in several ways; the important point (looking at Conotoxin, PDB: 1AS5 is that we are generally tracking binding or interaction sites.**





UMKC

VASSA INFORMATICS

ChemVassa Overview

- ChemVassa works by calculating the information content of a molecule, utilizing spatial information (taken from PDB, converted SMILES, or other structural information files) to locate an atom within the molecule
- Each atom is converted using a lexicon that accounts for the valence shell content, atomic number, and reactivity of the atom
- The location of each atom is then compared and the reactivity between adjacent atoms is compared
  - The average of the distance multiplied by the reactivity difference is the **"G score"**
- G scores for the backbone of the molecule are calculated as follows:
  - The average for connected non-main-chain molecules is added to the connected main carbon atom and summed across the backbone and averaged; this is the **"M score"**
- A string of G scores may then be searched across a database of compounds

UMKC

VASSA INFORMATICS

ChemVassa Validation

Test Molecule: Lipitor

Question:

Is ChemVassa able to predict novel binding partners for a chemical ligand that cannot be predicted by existing methods *ab initio*?

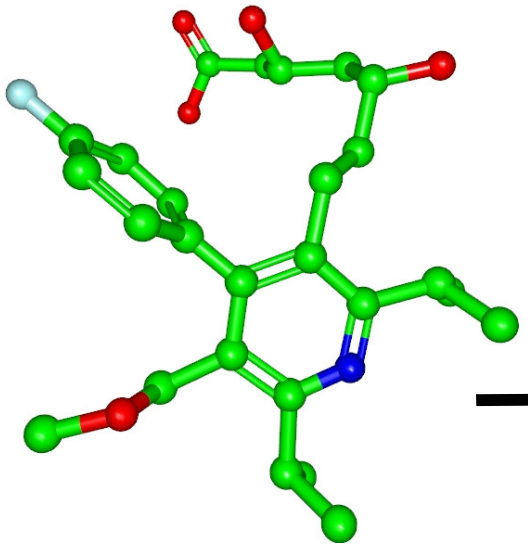
Lipitor

- Lipitor was chosen as:
- Commercially valuable
- Crystal structures of HMG-CoA reductase in complex with six statins are available
- All marketed HMG CoA reductase inhibitors are structurally similar – can we identify novel scaffolds and chemistry?

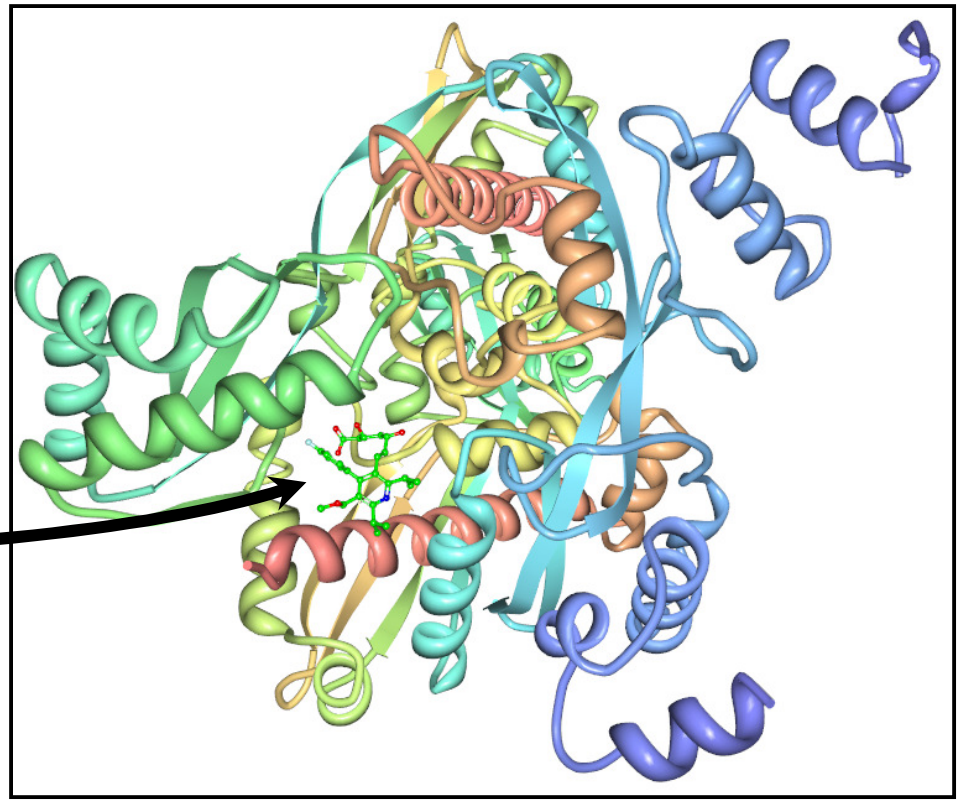
UMKC

VASSA INFORMATICS

Lipitor Structure



Lipitor

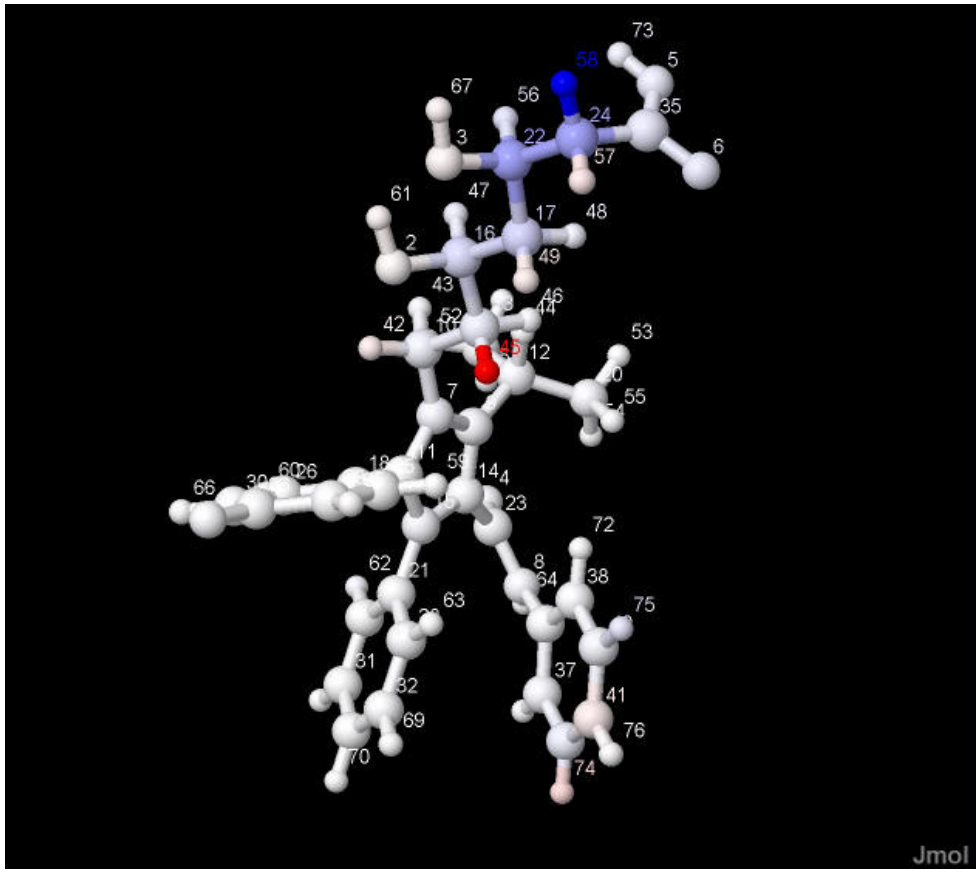


Lipitor works by binding to and inhibiting the liver enzyme HMG-CoA reductase

UMKC

VASSA INFORMATICS

VaSSA Analysis - Lipitor



- Red Regions are high information content
- Blue regions are low
- ChemVassa *correctly identified* the binding region where Lipitor interacts with its target (Hmg-CoA reductase) and predicts most of the important interacting atoms that were determined experimentally

LK1

LK2

UMKC

VASSA INFORMATICS

Slide 9

LK1 is "most of the important interacting atoms..." the best way to say this?
Lisa Kenney, 1/29/2010

LK2 This should go on the Results slide (as written).

I think this third bullet should provide additiona analysis information or explanation
Lisa Kenney, 1/29/2010

Results

- We searched a 600,000 ligand library using the Lipitor information signature
- We categorized the results as: Positive (Validation), Known Binders (Neutral), False Positive, or Novel
- Of these results, about 60% were known binders or novel results; 40% were false positives
- We found 10 Novel, previously unknown results which can be tested for functionality at the bench
  - These novel compounds would not be able to be identified utilizing existing methods

UMKC

VASSA INFORMATICS

Slide 10

LK3 I rewrote this and it still isn't quite right. Theoretically, we should provide the breakdown by category...
Lisa Kenney, 1/29/2010

LK4 Are Known Binders Neural, or are they validation?
Lisa Kenney, 1/29/2010

Positive (Validation) Results

- We utilized the search to see if it would identify other statins
- Creating a statin library, we reliably pulled statins as results if the binding region was used as search input
- This shows that there is a shared set of physical properties that ChemVassa is able to detect within the statins

UMKC

VASSA INFORMATICS

Known Binders (Neutral)

- In some cases, we pulled results that were not statins and NOT structurally similar to lipitor that, however, are known binders of HmG-CoA Reductase.
- An example is Coenzyme A; it was returned as a search result though it is NOT structurally similar to Lipitor.
- However, as CoA binds HmG-CoA reductase, it is NOT a negative result and suggested that the algorithm is tracking a FUNCTIONAL property of HmG-CoA reductase binding, NOT just a physical one.

UMKC

VASSA INFORMATICS

False Positives

- Of course, we also returned some results that do NOT bind HmG-Coenzyme A reductase and are NOT structurally similar to Lipitor.
- An example is Vancomyacin; it was returned as a search result though it is NOT structurally similar to Lipitor and DOES NOT bind HmG-CoA reductase.
- These results fell into two categories; complete non-binders, and cases where a portion of the molecule would likely bind except cannot due to steric hindrance.
- About 40% of the experimental search results fell into this category.

UMKC

VASSA INFORMATICS

Novel Results

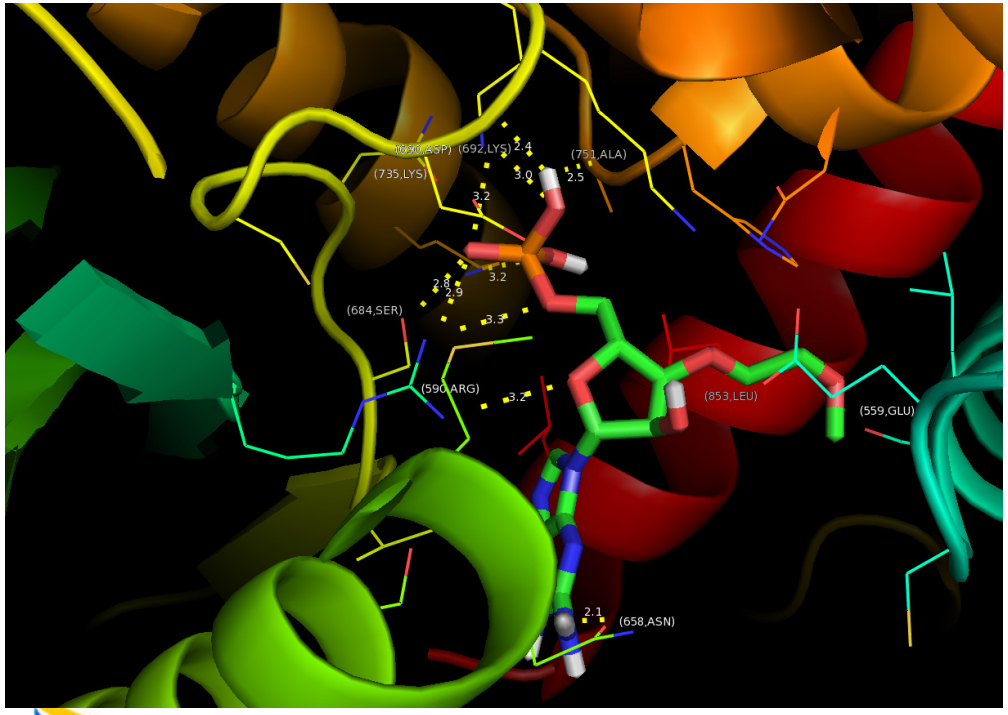
- These results are not structurally similar to Lipitor, but that appear to be capable of binding HmG-CoA reductase in a manner similar to Lipitor
- Modeling allowed us to look at affinity and electrostatic contacts of these results
- About 40% of the experimental search results fell into this category, most with exceptionally good binding.
- These included some hits where little biological information was present, and cases where biological information provided insight into the possible mechanism for the ligand function
- Currently, VaSSA Informatics is utilizing these results for partnership development with several interested parties.

UMKC

VASSA INFORMATICS

Additional Results

G2L (3'-o-methyoxyethyl-guanosine-5'-monophosphate) also interacts with the Hmg-CoA reductase site. Although the compound is not well-studied, it is small and should be bio-accessible.



G2L	AVS
Glu:559	Y
Arg:590	Y*
Leu:658	N
Ser:684	Y*
Asp:690	N
Lys:692	Y*
Lys:735	Y*
Ala:751	Y*
Leu:853	Y*

UMKC

VASSA INFORMATICS

Answer:

YES, our models show that ChemVassa can *ab initio** predict novel binding partners for chemical ligands that cannot be predicted by existing methods.

Cost Model Discussion

How are we moving forward?

- Develop a compound library of about 9 million compounds, including about 1.5 million "sweet-spot" that have good drug-like qualities.
- Developing the infrastructure for this quickly and on a limited budget for deployment, we explored several options:
  - Server purchase
  - Machine rental
  - Cloud services

UMKC

VASSA INFORMATICS

Cloud computing

- We have two tasks:
  - Initial candidate screening, using ChemVaSSA to generate a compound library and screen the library
  - Modeling the results to see if they are compatible with binding.
- Project-based pricing
  - Creating the library: about $300.
  - Screening the library: about $30 per compound.
  - Modeling the results: about $55 per 1000 models.

UMKC

VASSA INFORMATICS

Conclusions

- Cloud computing may work for initial development of computing infrastructure.
  - Not ideal for all cases
- Allows accurate prediction of project times
- Allows quick set-up/tear down of infrastructure
- Costs can be billed back to a source (grant, client, etc)
- Low overhead.

UMKC

VASSA INFORMATICS

Future Directions

- Cloud computing for development of Bioinformatic teaching infrastructure
  - Non-persistant nature of facilities fits well with semester-to-semester changes in enrollment.
  - Cost basis can be readily understood.
  - Limiting student access to ensure effective use of resources
- Development of trial web-based resources for grants and exploratory research.

UMKC

VASSA INFORMATICS