Docking & drug discovery

Máster de bioinformática, Universidad de Valencia
14 - Abril - 2014

fmartinez@pcb.ub.es

Summary

- Introduction
- Small molecules binding site prediction
  - de-novo.
  - comparative.
- Docking.
  - What is docking?
  - Autodock and state-of-the-art methods.
  - An application in drug discovery : ISENTRESS.
- nAnnolyze.

Objective

TO LEARN HOW-TO USE AutoDock Vina FOR DOCKING SMALL MOLECULES IN THE SURFACE OF A PROTEIN

Nomenclature

Ligand: Structure (usually a small molecule) that binds to the binding site.

Receptor: Structure (usually a protein) that contains the active binding site.

Binding site: Set of amino-acids (residues) that physically interact with the ligand (usually within 6 Ångstroms).

From sequence to function...

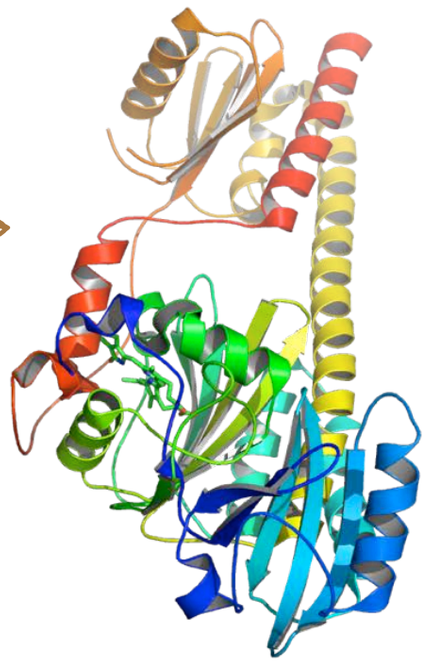
Sequence

Structure

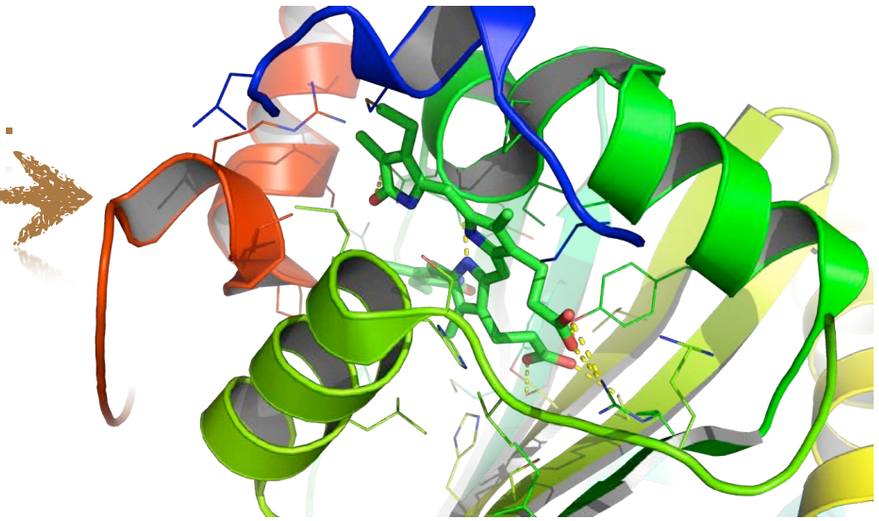
Function

```
MTSITPVTLANCEDEP
IHVPGAIQPHGALVTL
RADGMVLAASENIQAL
LGFVASPGSYLTQEQV
GPEVLRMLEEGLTGNG
P....
```

implies..



implies..

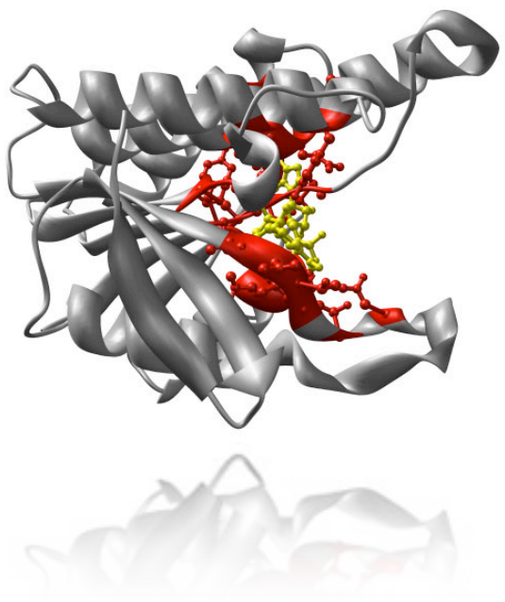


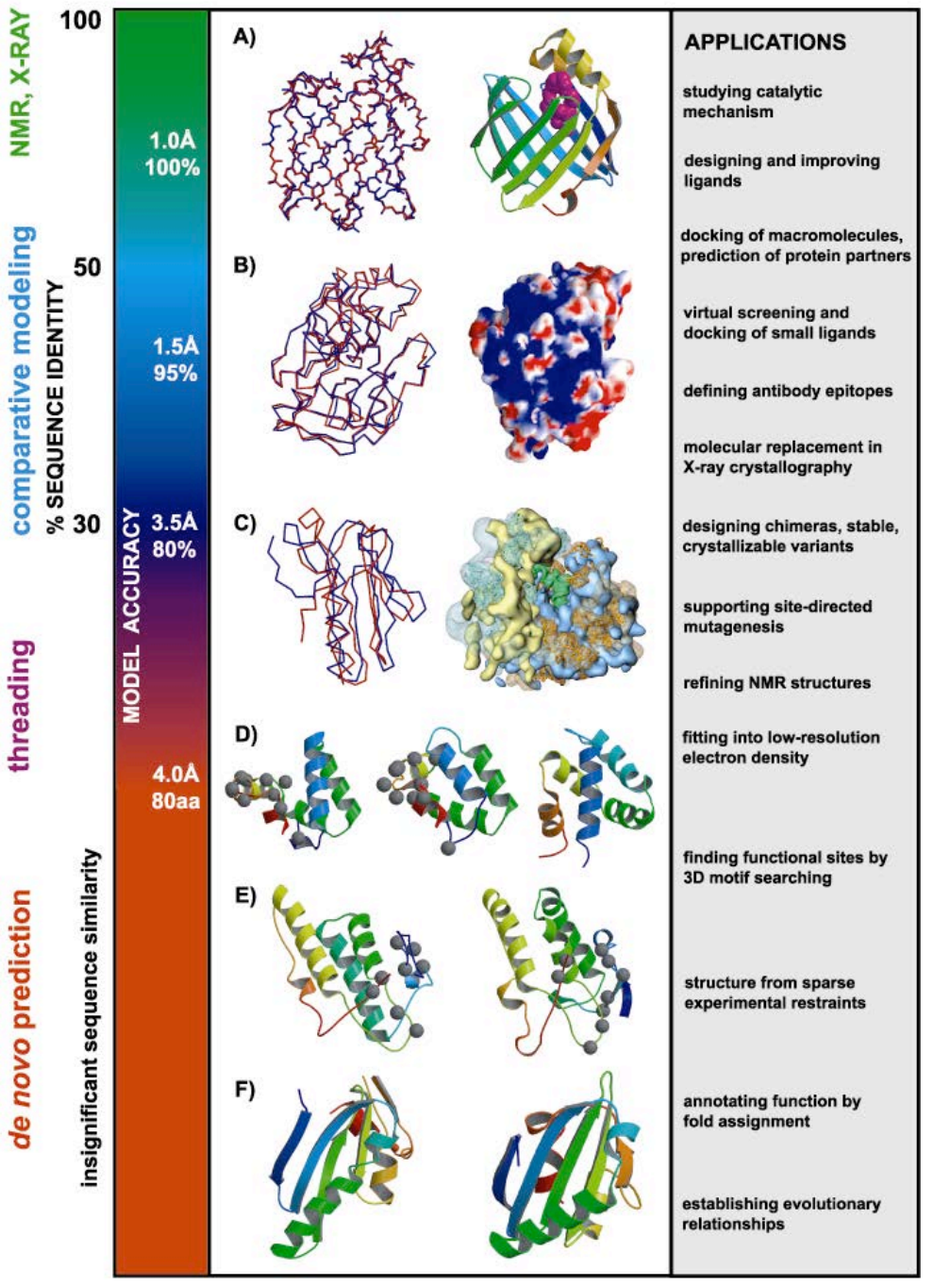
- conserved+

Program

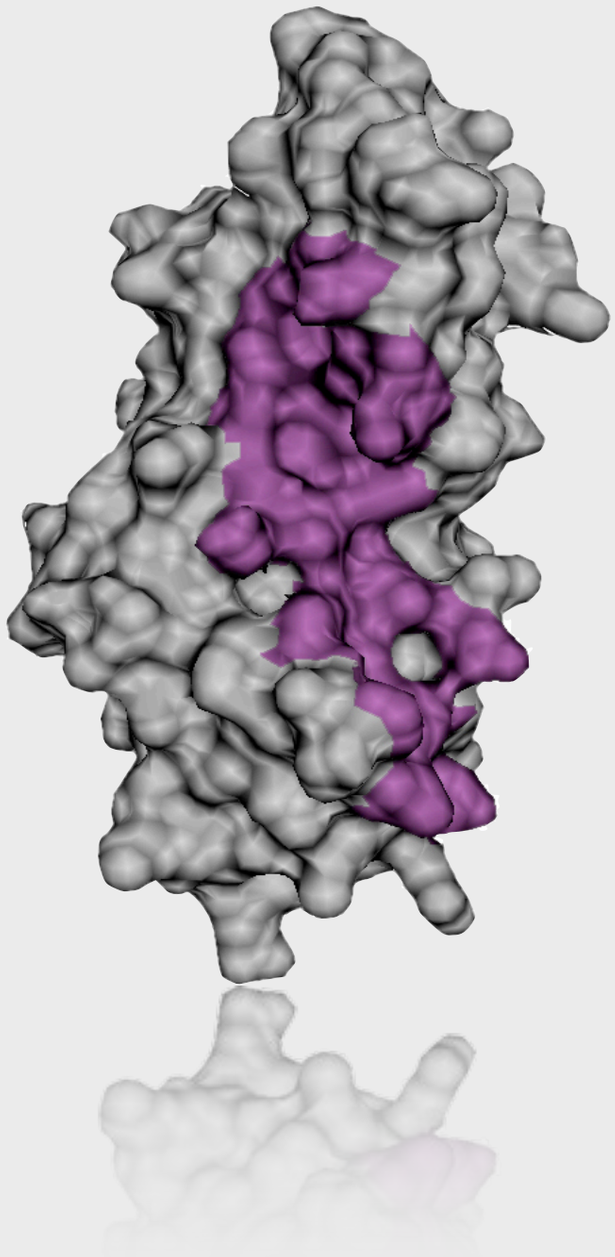
Binding site prediction

AutoDock





binding site prediction



- Sometimes, we know the binding site for a ligand because it has been co-crystalized with the protein.
- Localize the binding site/s for a given molecule.
- There could be several binding sites in a protein surface.
- Two different approaches for binding site prediction : de novo & comparative prediction.

Protein function from structure

ab-initio *localization of binding sites*

Rossi. Localization of binding sites in protein structures by optimization of a composite scoring function. Protein Science (2006) vol. 15 (10) pp. 2366-2380

Localization of binding sites in protein structures by optimization of a composite scoring function

ANDREA ROSSI, MARC A. MARTI-RENOM, AND ANDREJ SALI
Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, California Institute for Quantitative Biomedical Research, University of California, San Francisco, California 94143-2552, USA

(RECEIVED March 28, 2006; FINAL REVISION July 10, 2006; ACCEPTED July 11, 2006)

Abstract

The rise in the number of functionally uncharacterized protein structures is increasing the demand for structure-based methods for functional annotation. Here, we describe a method for predicting the location of a binding site of a given type on a target protein structure. The method begins by constructing a scoring function, followed by a Monte Carlo optimization, to find a good scoring patch on the protein surface. The scoring function is a weighted linear combination of the z-scores of various properties of protein structure and sequence, including amino acid residue conservation, compactness, protrusion, convexity, rigidity, hydrophobicity, and charge density; the weights are calculated from a set of previously identified instances of the binding-site type on known protein structures. The scoring function can easily incorporate different types of information useful in localization, thus increasing the applicability and accuracy of the approach. To test the method, 1008 known protein structures were split into 20 different groups according to the type of the bound ligand. For nonsugar ligands, such as various nucleotides, binding sites were correctly identified in 55%–73% of the cases. The method is completely automated (http://salilab.org/patcher) and can be applied on a large scale in a structural genomics setting.

Keywords: protein function annotation; small ligand binding-site localization

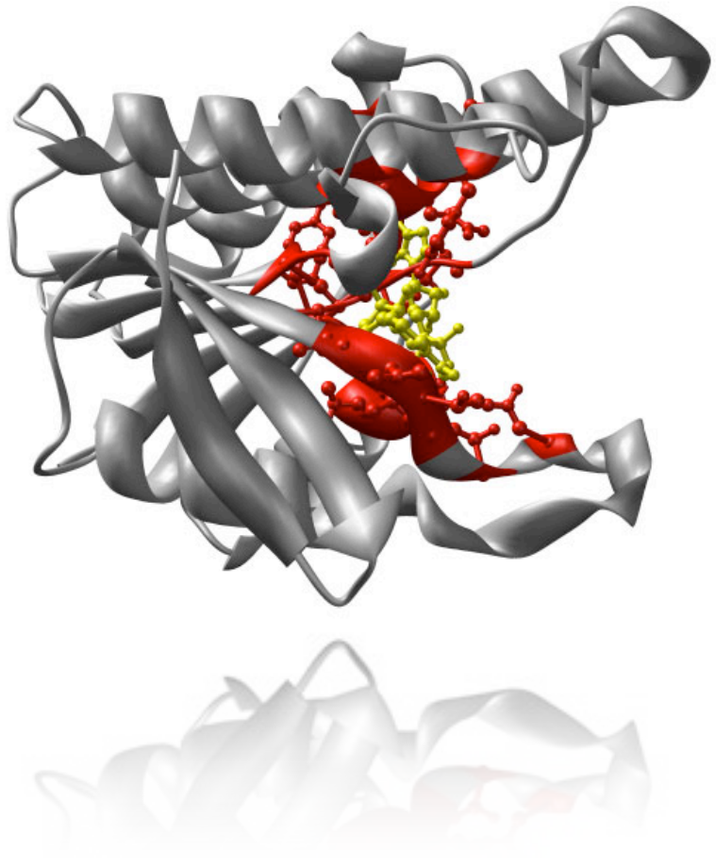
Many protein targets of structural biologists are no longer chosen because of their function, but rather by their location in the protein sequence-structure space (Burley et al. 1999; Brenner 2000, 2001; Sali 2001; Vitkup et al. 2001; Chance et al. 2002; Goldsmith-Fischman and Honig 2003). Therefore, the number of functionally uncharacterized protein structures is growing. Of the 36,606 entries in the Protein Data Bank (PDB) (Kouranov et al. 2006) as of February 23, 2006, 1407 structures were deposited by structural genomics consortia, 985 (70%) of which had an unknown function according to the HEADER record of their PDB files. In contrast, only 174 (0.5%) of the 35,199 protein structures solved outside of structural genomics had no functional annotations in their PDB files.

To classify the functions of thousands of uncharacterized protein structures that will become available over the next few years and millions of comparative models based on the known structures, automated structure-based functional annotation is required (Wallace et al. 1996, 1997; Kleywegt 1999; Thornton et al. 2000; Babbitt 2003; Laskowski et al. 2003). In particular, we need to be able to identify the locations and types of binding sites on a given structure, because the binding sites define the molecular function of a protein.

The most principled computational approach to predicting the molecular function is to dock a large library of potential ligands against the surface of the protein. In

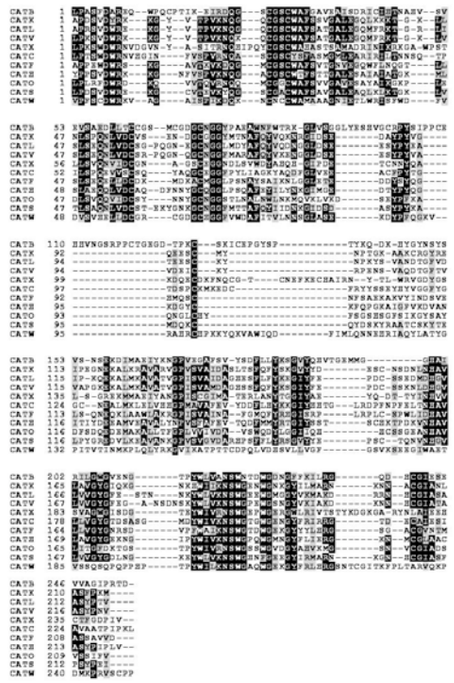
Reprint requests to: Andrea Rossi or Andrej Sali, Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, California Institute for Quantitative Biomedical Research, University of California, San Francisco Byers Hall, Office 503B, 1700 4th Street, San Francisco, CA 94143-2552, USA; e-mail: andrea@salilab.org or sali@salilab.org; fax: (415) 514-4231.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.062247506.

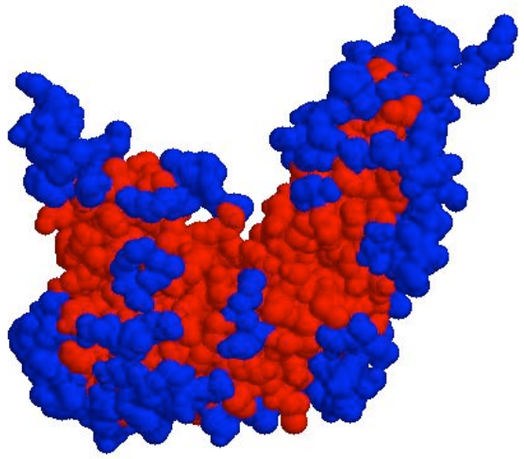


Representation

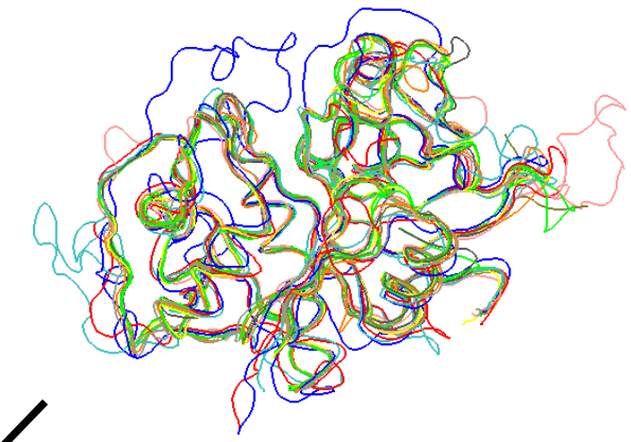
Sequence conservation



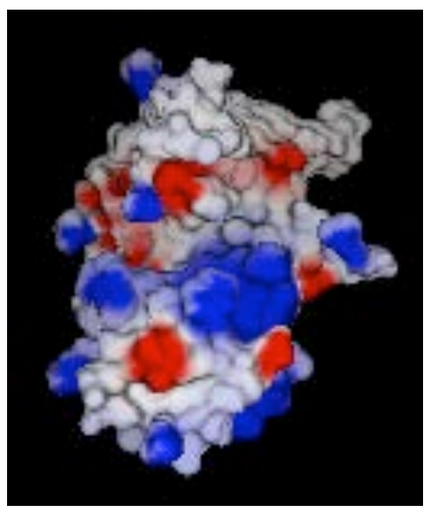
Surface geometry

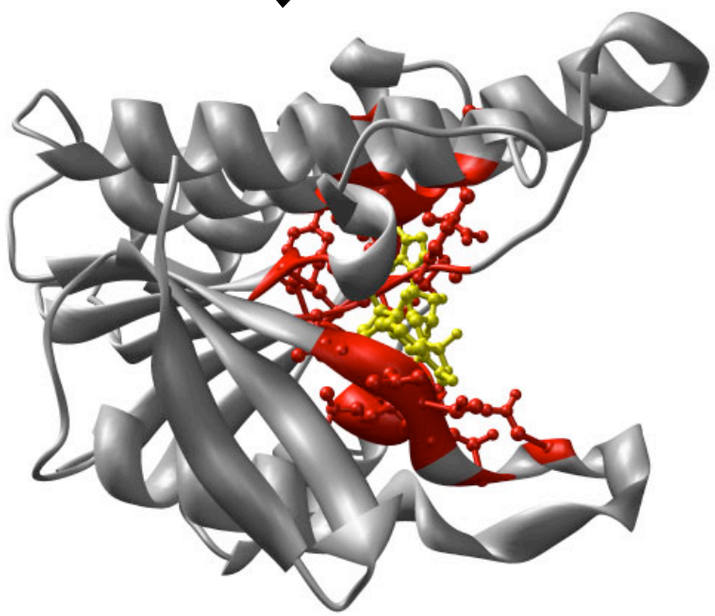


Structure conservation

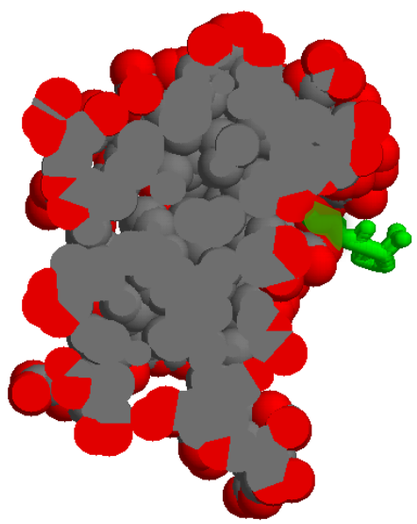


Electrostatics



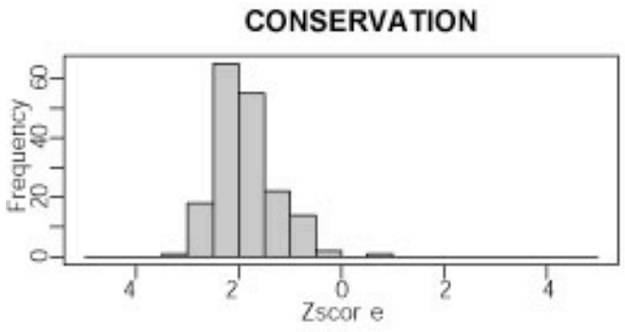


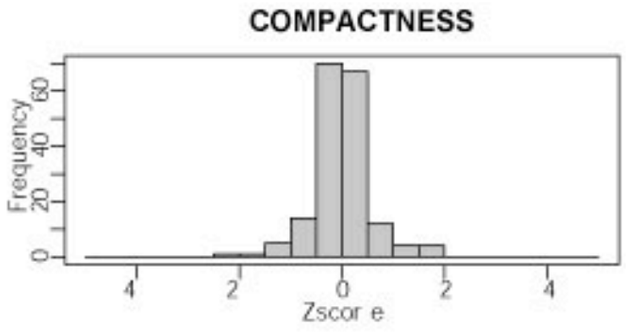
Solvent accessibility

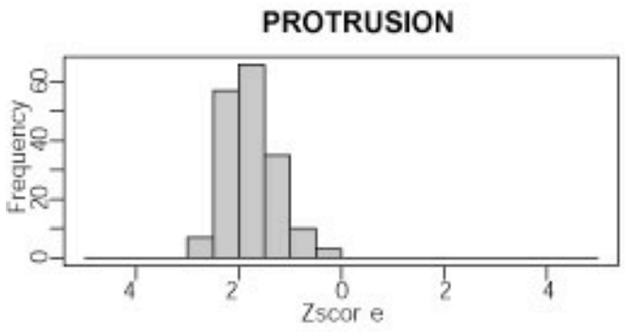


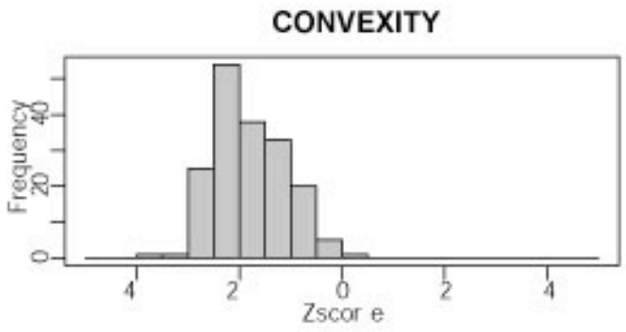
Scoring

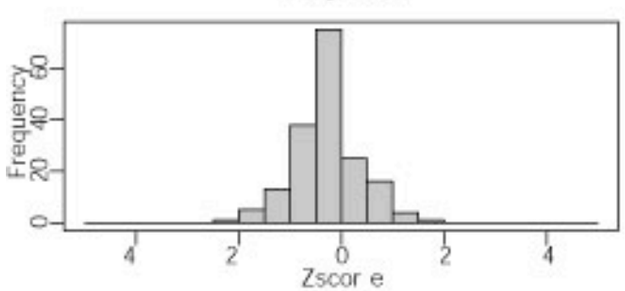
NAD

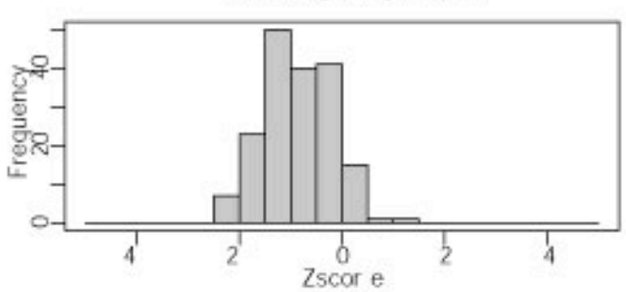


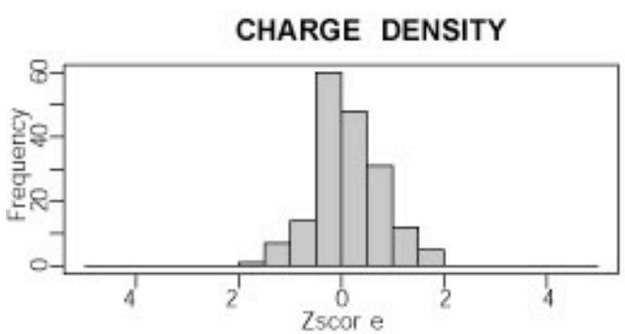












$$\tilde{f} = (f - \langle f \rangle)/\sigma_f.$$

Getting the z-score for each feature.

$$\rightarrow \quad w_k = \frac{1}{M}\sum_{\alpha=1}^{M} \tilde{f}_k^{(\alpha)}$$

M = number of proteins in training set

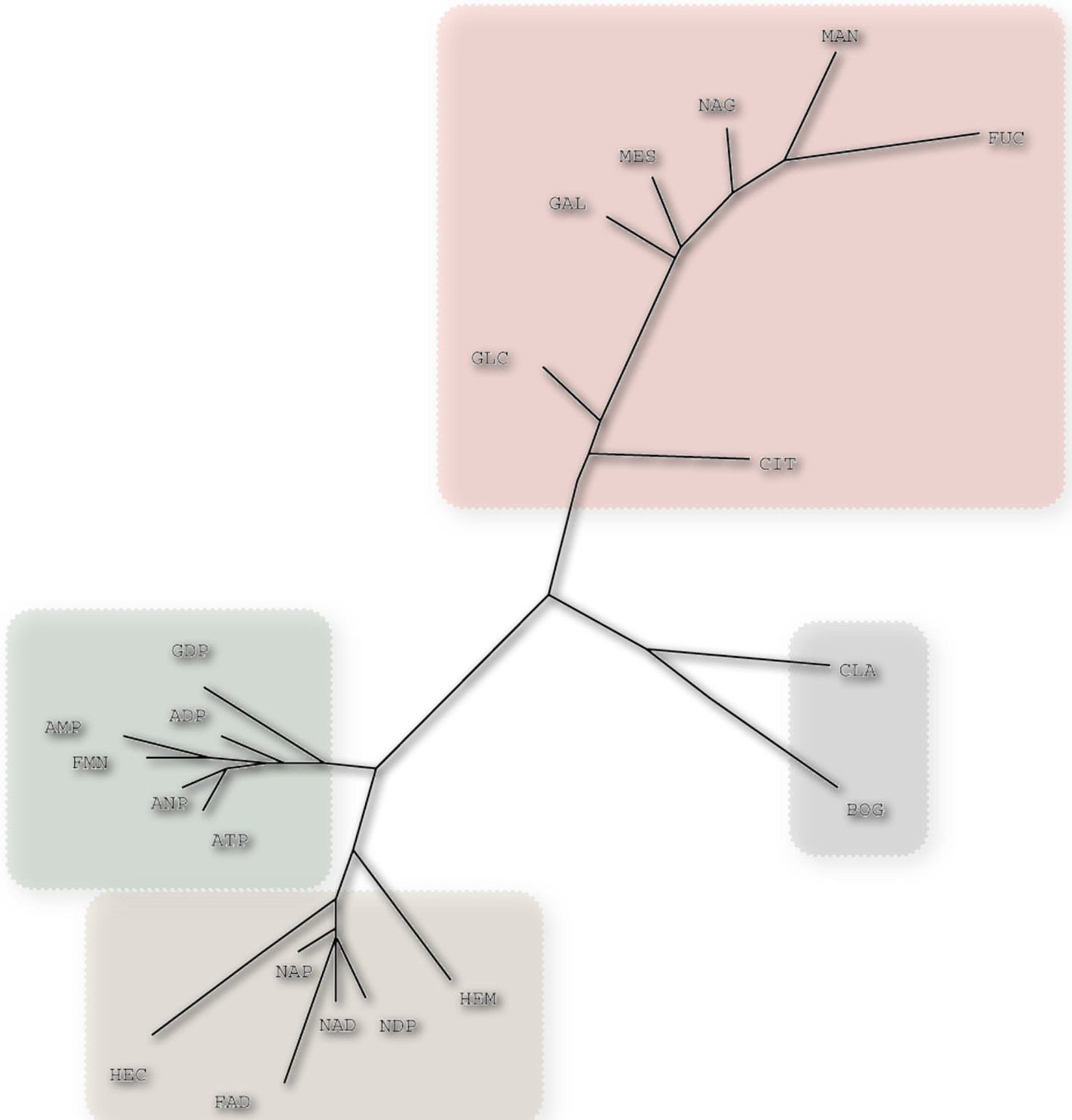
$$F(P) = \sum_{k=1}^{7} w_k \cdot \tilde{f}_k(P),$$

Optimization, maximizing score.

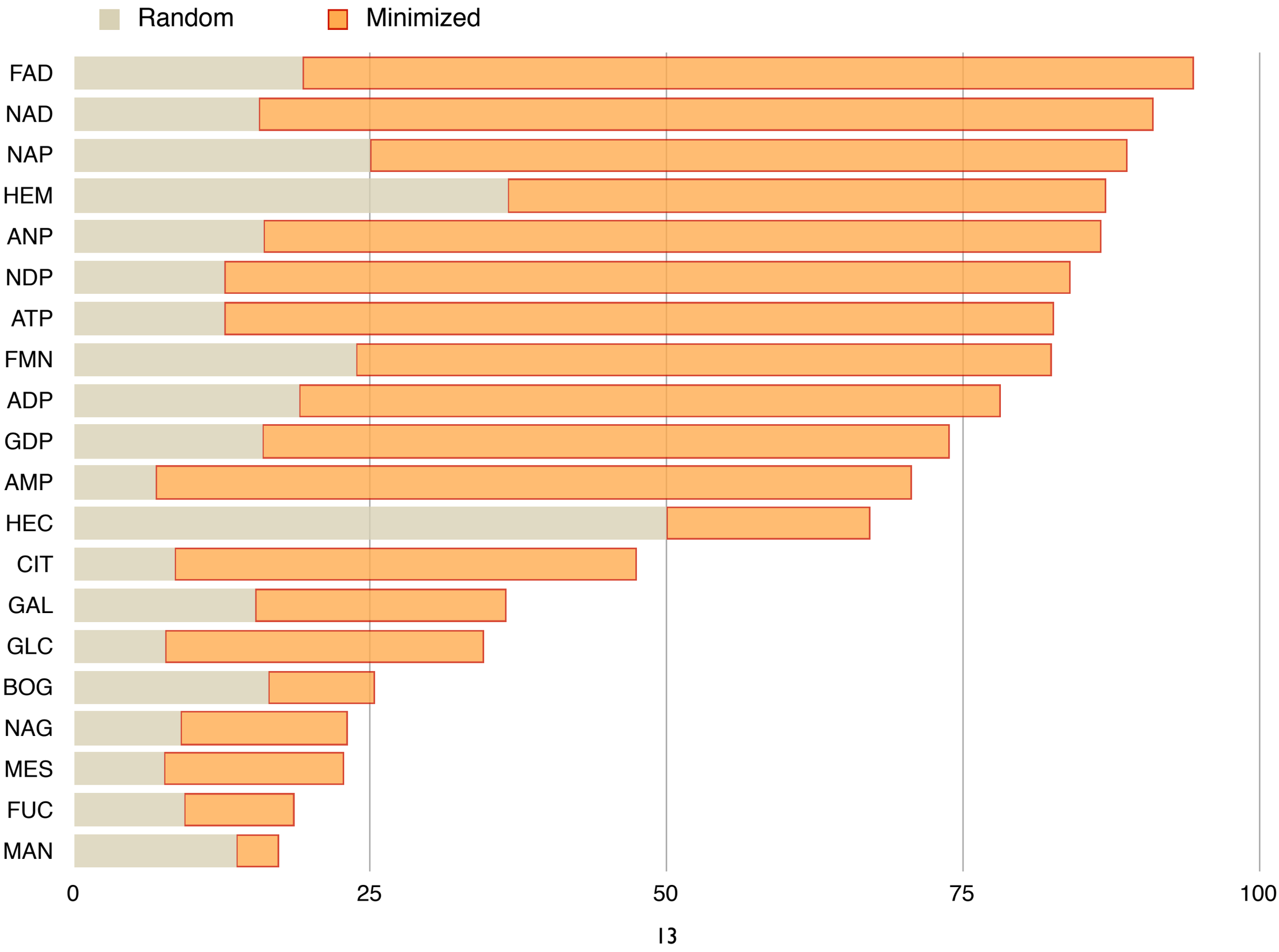
Ligand fingerprints

	Compactness	Conservation	Charge density	B-factor	Protrusion coefficient	Convexity score	Hydrophobicity
ADP	-1.266	-2.009	0.447	-0.414	-1.521	-1.388	-0.118
AMP	-1.62	-1.962	0.341	-0.381	-1.909	-1.944	-0.518
ANP	-1.007	-2.227	0.176	-0.392	-1.706	-1.595	-0.14
ATP	-1.122	-2.156	0.228	-0.274	-1.845	-1.768	0.038
BOG	-2.067	-0.012	0.552	-0.465	-0.356	-0.49	-0.781
CIT	-2.948	-1.58	0.563	-0.527	-0.922	-0.838	-0.113
FAD	0.505	-2.108	0.366	-0.702	-1.735	-1.725	-0.75
FMN	-1.132	-1.98	0.382	-0.387	-1.803	-1.886	-0.695
FUC	-3.43	0.016	-0.295	-0.123	0.002	0.132	0.459
GAL	-3.186	-0.538	-0.234	-0.068	-0.906	-0.987	0.298
GDP	-1.061	-1.471	0.409	-0.81	-1.472	-1.423	0.182
GLC	-2.813	-1.247	-0.207	-0.399	-1.247	-1.337	-0.089
HEC	-0.172	-0.912	0.286	-0.325	-1.153	-1.27	-1.282
HEM	-0.651	-1.571	0.683	-0.51	-1.797	-1.937	-1.47
MAN	-3.72	0.131	0.105	-0.52	-0.605	-0.509	0.405
MES	-3.049	-0.24	-0.338	-0.479	-0.714	-0.926	0.296
NAD	-0.005	-1.852	0.156	-0.232	-1.775	-1.804	-0.858
NAG	-3.419	-0.46	-0.126	-0.154	-0.341	-0.523	-0.078
NAP	-0.009	-1.898	0.612	-0.321	-1.587	-1.656	-0.336
NDP	0.217	-1.741	0.535	-0.312	-1.463	-1.562	-0.498

Ligand fingerprints



Prediction accuracy



Protein function from structure

Comparative annotation. AnnoLite and AnnoLyze.

Marti-Renom et al. The AnnoLite and AnnoLyze programs for comparative annotation of protein structures. BMC Bioinformatics (2007) vol. 8 (Suppl 4) pp. S4

BMC Bioinformatics

BioMed Central

Proceedings

Open Access

The AnnoLite and AnnoLyze programs for comparative annotation of protein structures

Marc A Marti-Renom*[1], Andrea Rossi[2], Fátima Al-Shahrour[3], Fred P Davis[2], Ursula Pieper[2], Joaquín Dopazo[3] and Andrej Sali[2]

Address: [1]Structural Genomics Unit, Bioinformatics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain, [2]Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, University of California at San Francisco, San Francisco, CA 94143, USA and [3]Functional Genomics Unit, Bioinformatics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain

Email: Marc A Marti-Renom* - mmarti@cipf.es; Andrea Rossi - andrea@salilab.org; Fátima Al-Shahrour - falshahrour@cipf.es; Fred P Davis - fred@salilab.org; Ursula Pieper - Ursula@salilab.org; Joaquín Dopazo - jdopazo@cipf.es; Andrej Sali - sali@salilab.org

* Corresponding author

from The Second Automated Function Prediction Meeting
La Jolla, CA, USA. 30 August – 1 September 2006

Published: 22 May 2007

BMC Bioinformatics 2007, **8**(Suppl 4):S4 doi:10.1186/1471-2105-8-S4-S4

This article is available from: http://www.biomedcentral.com/1471-2105/8/S4/S4

Abstract

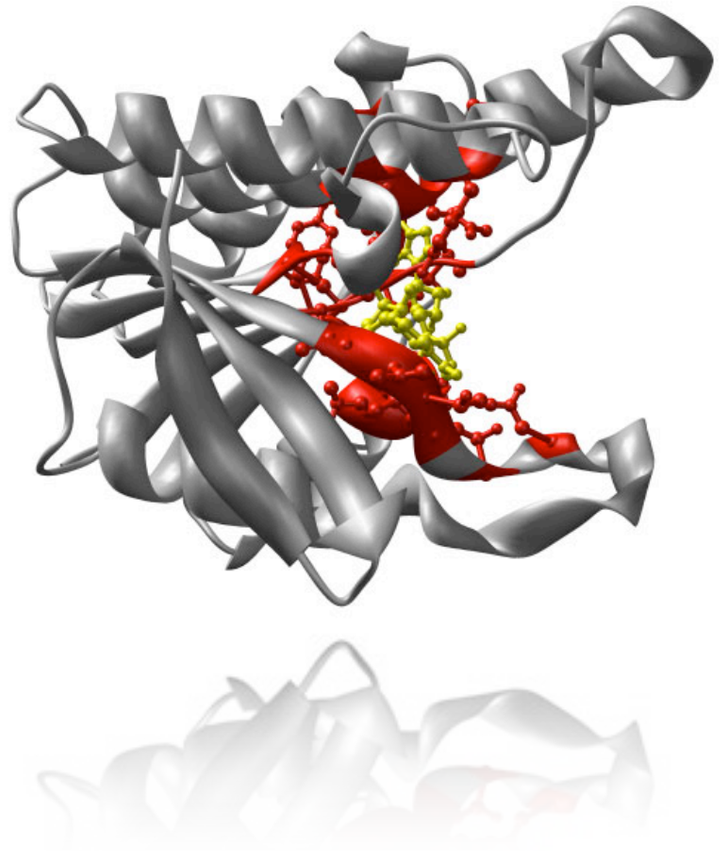
Background: Advances in structural biology, including structural genomics, have resulted in a rapid increase in the number of experimentally determined protein structures. However, about half of the structures deposited by the structural genomics consortia have little or no information about their biological function. Therefore, there is a need for tools for automatically and comprehensively annotating the function of protein structures. We aim to provide such tools by applying comparative protein structure annotation that relies on detectable relationships between protein structures to transfer functional annotations. Here we introduce two programs, AnnoLite and AnnoLyze, which use the structural alignments deposited in the DBAli database.

Description: AnnoLite predicts the SCOP, CATH, EC, InterPro, PfamA, and GO terms with an average sensitivity of ~90% and average precision of ~80%. AnnoLyze predicts ligand binding site and domain interaction patches with an average sensitivity of ~70% and average precision of ~30%, correctly localizing binding sites for small molecules in ~95% of its predictions.

Conclusion: The AnnoLite and AnnoLyze programs for comparative annotation of protein structures can reliably and automatically annotate new protein structures. The programs are fully accessible via the Internet as part of the DBAli suite of tools at http://salilab.org/DBAli/.

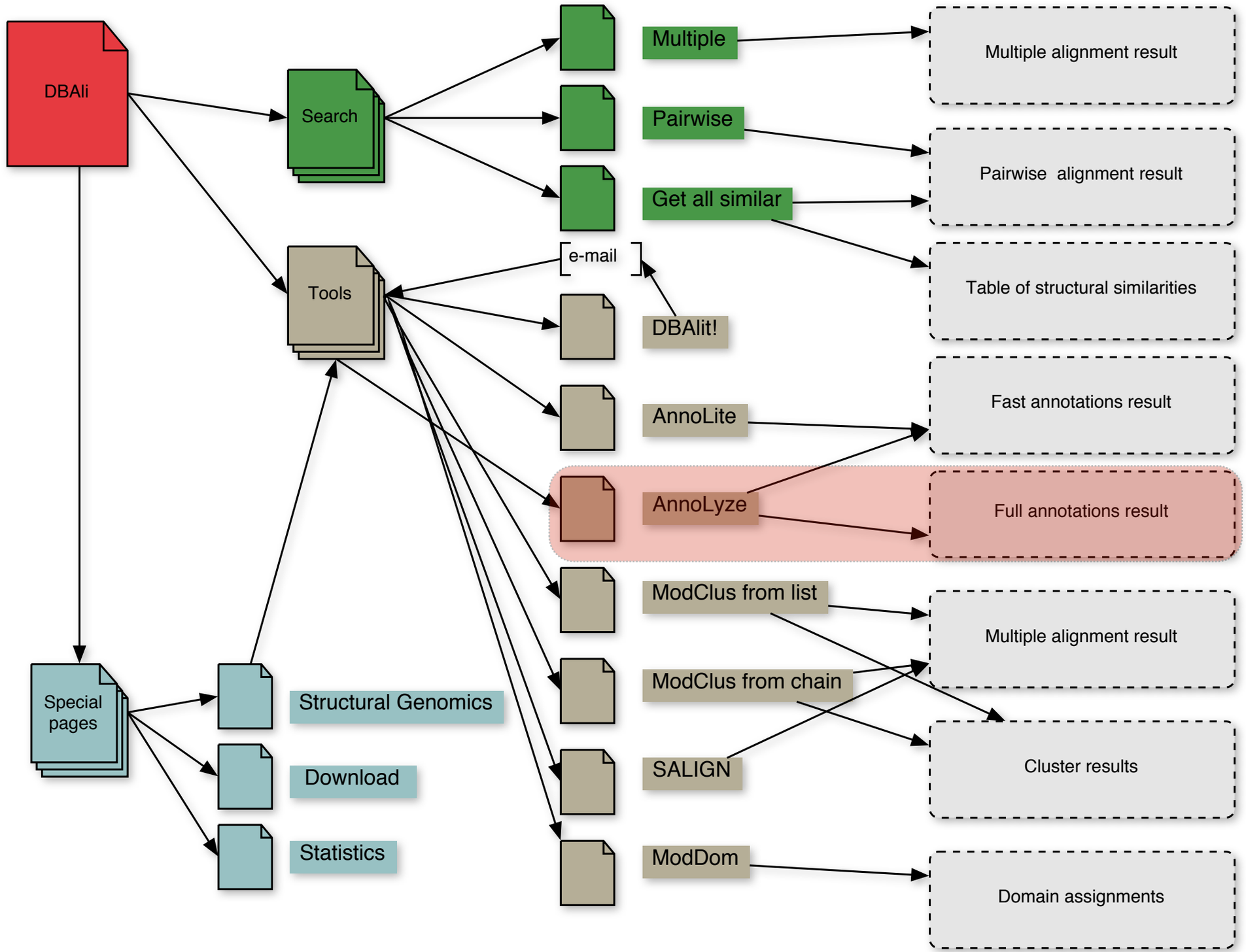
Background

Genomic efforts are providing us with complete genetic blueprints for hundreds of organisms, including humans. We are now faced with assigning, understanding, and modifying the functions of proteins encoded by these genomes. This task is generally facilitated by protein 3D



DBAli$_{v2.0}$ database

http://www.dbali.org



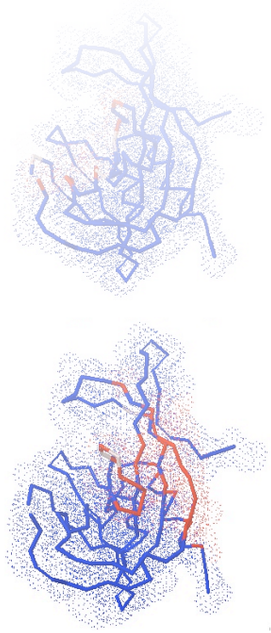
AnnoLyze

Inherited ligands: 4

Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
MO2	59.03	0.185	48 49 52 62 63 66 67 113 116
CRY	20.00	0.111	23 29 31 37 44 48 49 83 85 94 96 103 121
8OG	20.00	0.111	19 20 21 48 49 51 96 98 136
ACY	15.87	0.163	23 29 31 37 44 45 81 83 85 94 96 98 103 121 135

Inherited partners:1

Partner	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
d.113.1.1	23.68	0.948	19 20 50 51 52 53 54 55 56 57 58 77 78 79 80 81 82 83 84 85 93 95 97 99 134 135 138 142 145



Benchmark

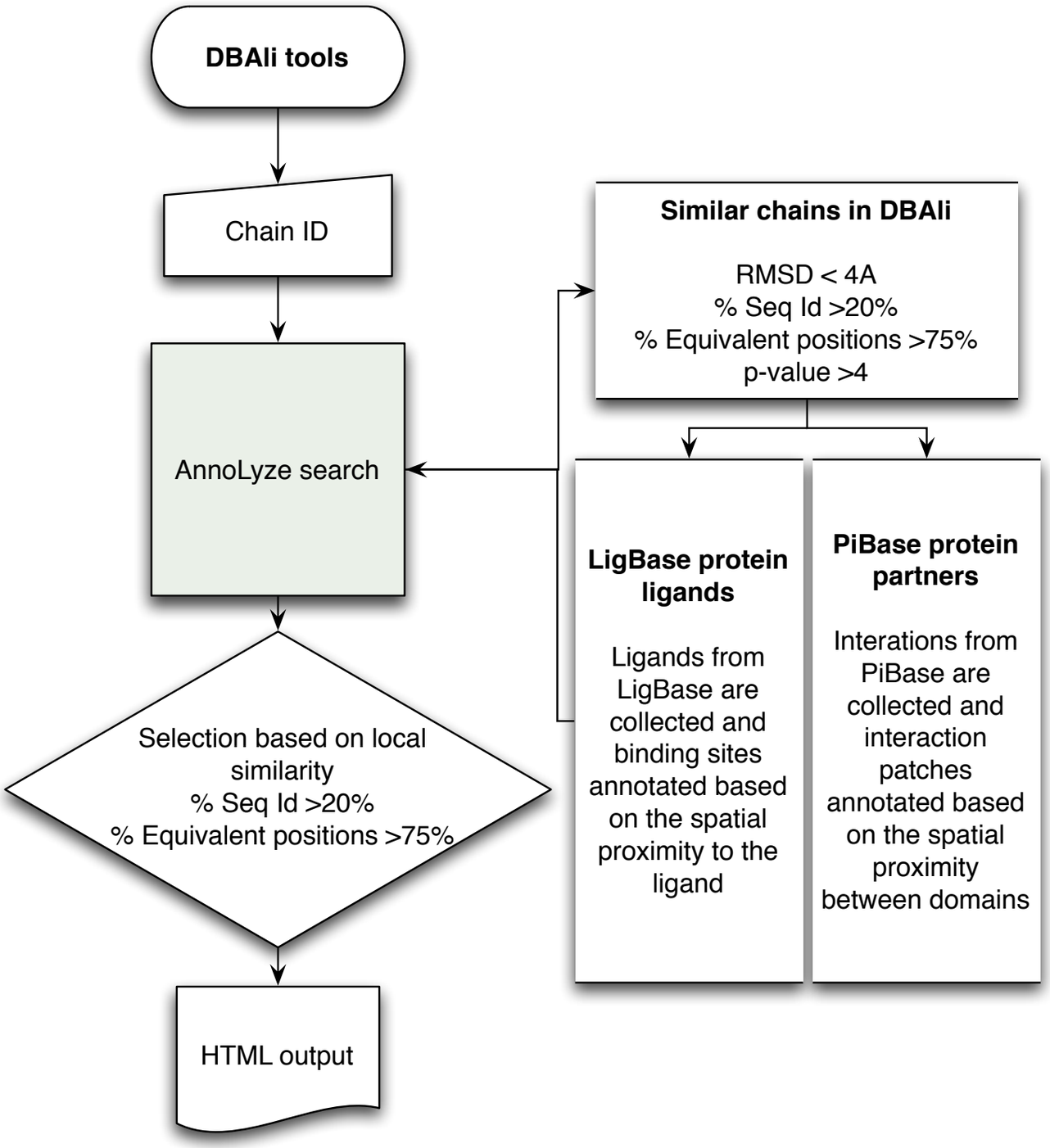
	Number of chains
Initial set*	78,167
LigBase**	30,126
Non-redundant set***	4,948 (8,846 ligands)

*all PDB chains larger than 30 aminoacids in length (8th of August, 2006)
**annotated with at least one ligand in the LigBase database
***not two chains can be structurally aligned within 3A, superimposing more than 75% of their Cα atoms, result in a sequence alignment with more than 30% identity, and have a length difference inferior to 50aa

	Number of chains
Initial set*	78,167
πBase**	30,425
Non-redundant set***	4,613 (11,641 partnerships)

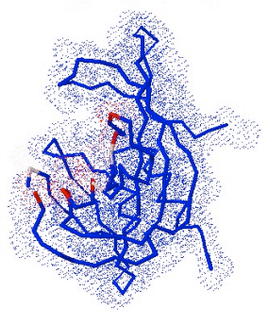
*all PDB chains larger than 30 aminoacids in length (8th of August, 2006)
**annotated with at least one partner in the πBase database
***not two chains can be structurally aligned within 3A, superimposing more than 75% of their Cα atoms, result in a sequence alignment with more than 30% identity, and have a length difference inferior to 50aa

Method



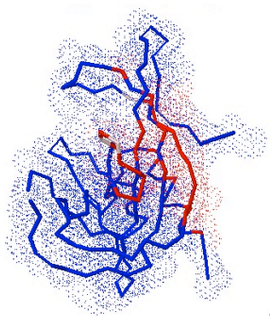
Inherited ligands: 4

Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
MO2	59.03	0.185	48 49 52 62 63 66 67 113 116
CRY	20.00	0.111	23 29 31 37 44 48 49 83 85 94 96 103 121
8OG	20.00	0.111	19 20 21 48 49 51 96 98 136
ACY	15.87	0.163	23 29 31 37 44 45 81 83 85 94 96 98 103 121 135



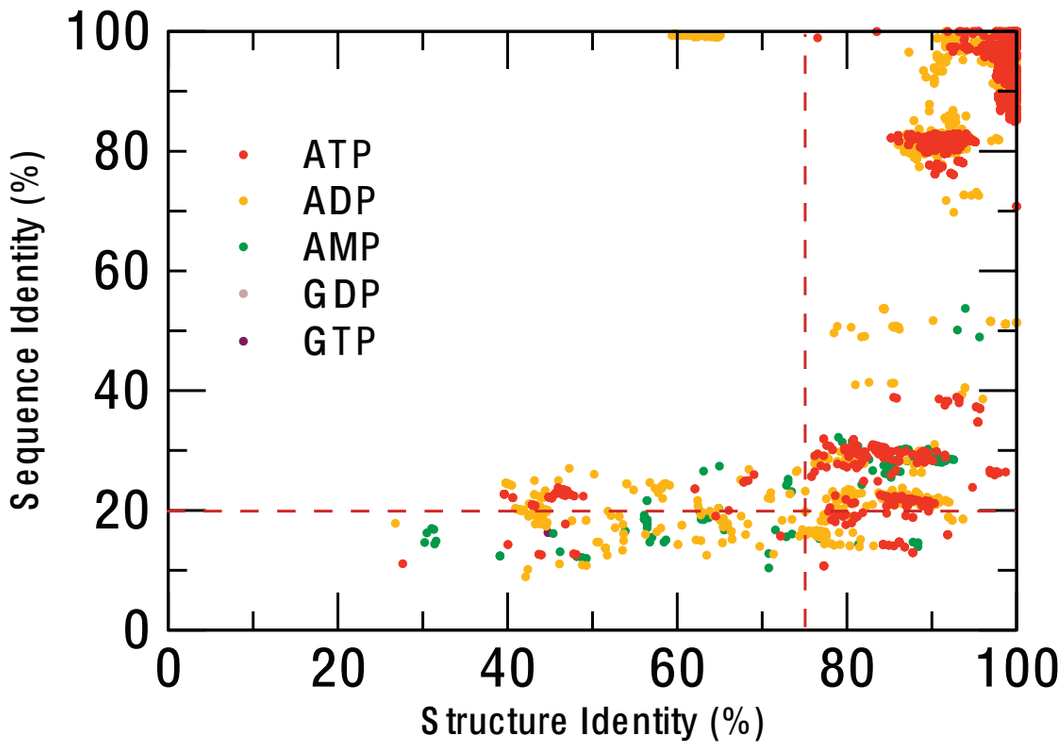
Inherited partners:1

Partner	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
d.113.1.1	23.68	0.948	19 20 50 51 52 53 54 55 56 57 58 77 78 79 80 81 82 83 84 85 93 95 97 99 134 135 138 142 145

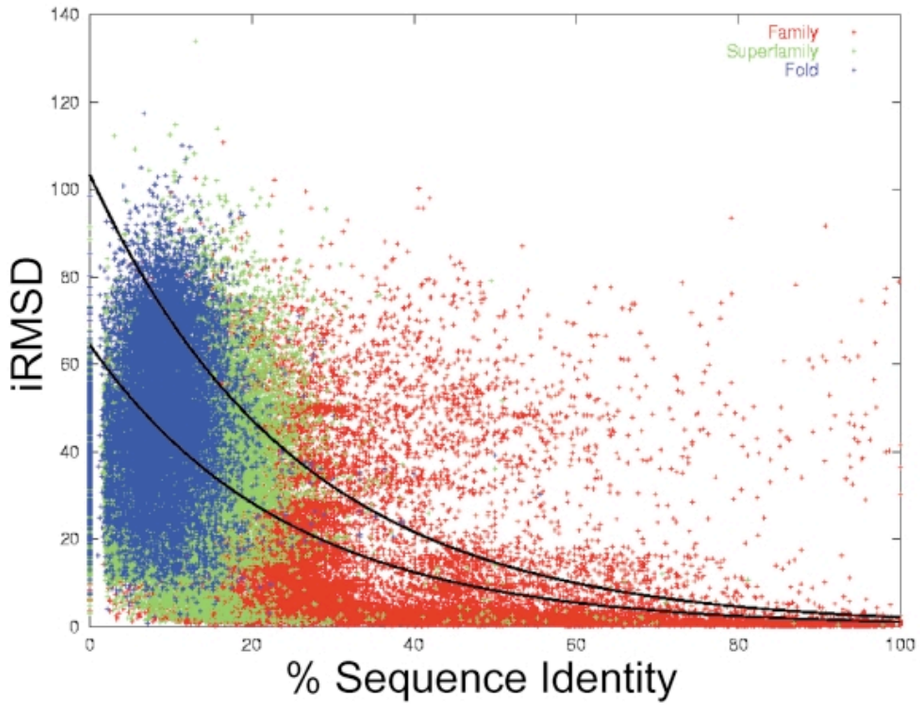


Scoring function

Ligands



Partners



Aloy *et al.* (2003) J.Mol.Biol. 332(5):989-98.

Sensitivity .vs. Precision

	Optimal cut-off	Sensitivity (%) Recall or TPR	Precision (%)
Ligands	30% BS COV	71.9	13.7
Partners	40% PS COV	72.9	55.7

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP}$$

However, 90-95% of aa correctly predicted

Other binding-site prediction web methods

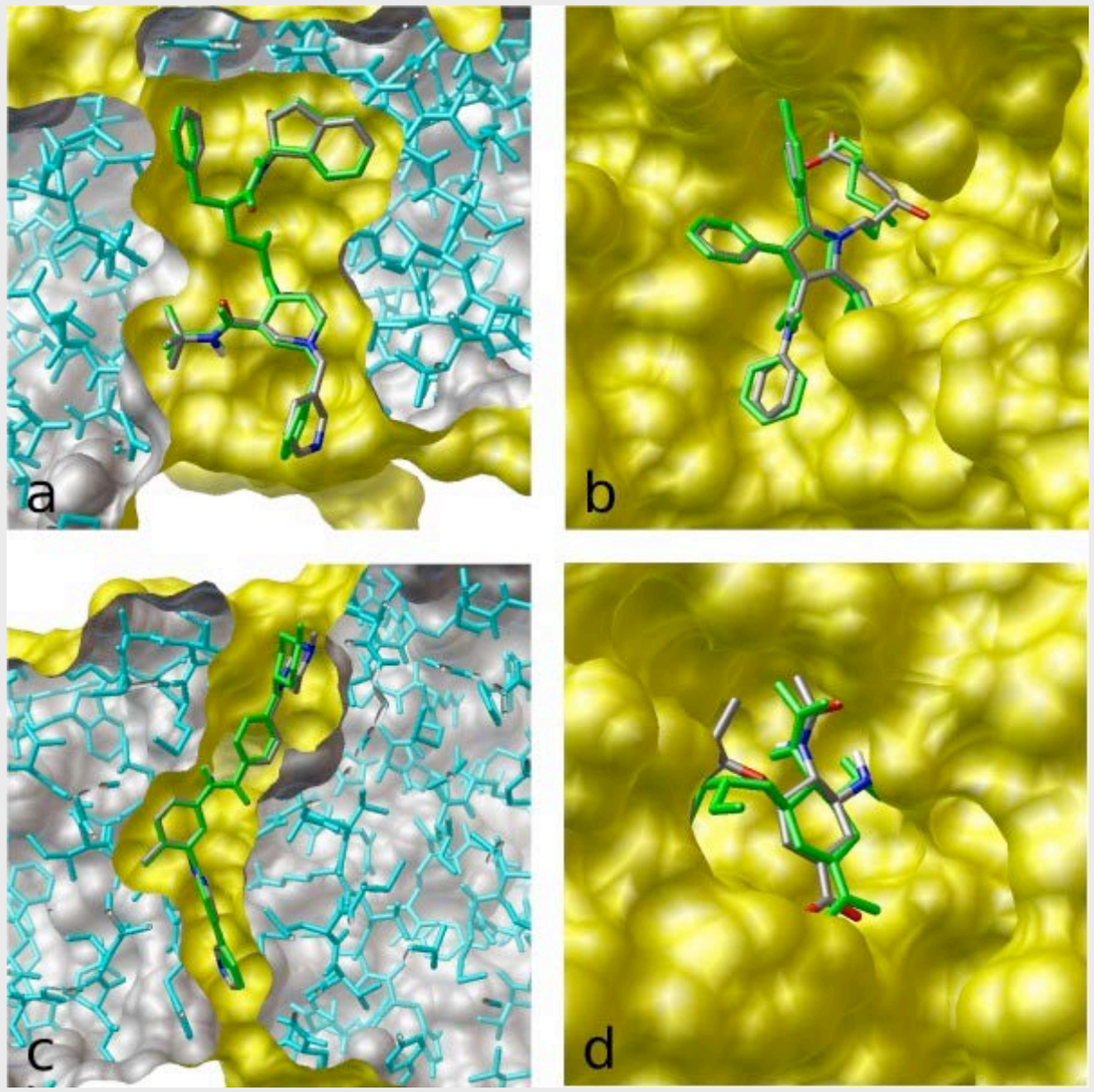
- **Metapocket 2.0** (http://projects.biotec.tu-dresden.de/metapocket/).

  ★ Metapredictor : LIGSITE,PASS, Q-SiteFinder, SURNET, Fpocket,GECOM, ConCavity, POCASA.

- **LISE** (http://lise.ibms.sinica.edu.tw).

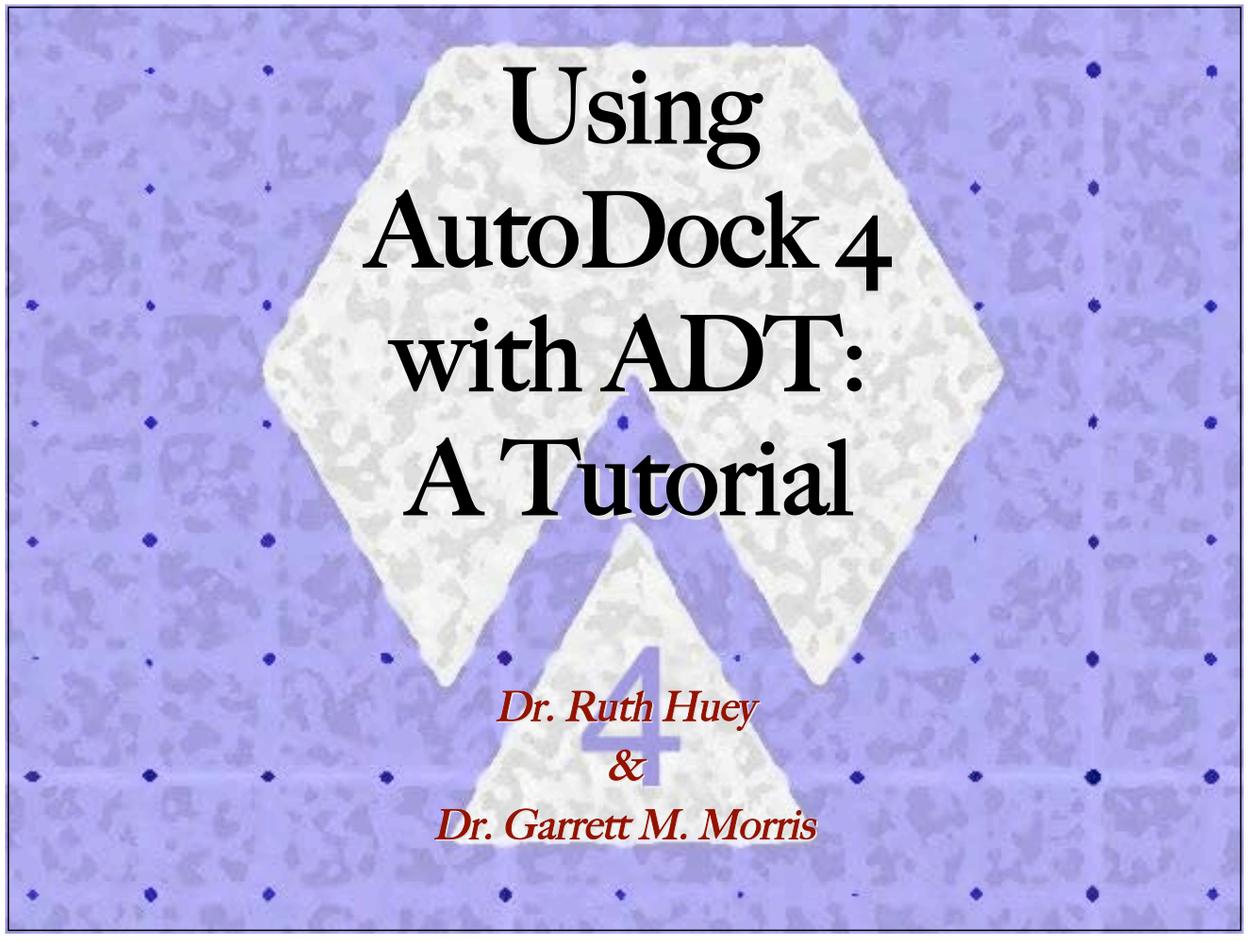
  ★ Binding Site-Enriched Protein Triangles method. Published in April 2012.

Docking of small molecules. Autodock Vina



DISCLAIMER!

Credit should go to Dr. Oleg Trott, Dr. Ruth Huey and Dr. Garret M. Morris



Software News and Update
AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading

OLEG TROTT, ARTHUR J. OLSON
Department of Molecular Biology, The Scripps Research Institute, La Jolla, California

Received 3 March 2009; Accepted 21 April 2009
DOI 10.1002/jcc.21334
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: AutoDock Vina, a new program for molecular docking and virtual screening, is presented. AutoDock Vina achieves an approximately two orders of magnitude speed-up compared with the molecular docking software previously developed in our lab (AutoDock 4), while also significantly improving the accuracy of the binding mode predictions, judging by our tests on the training set used in AutoDock 4 development. Further speed-up is achieved from parallelism, by using multithreading on multicore machines. AutoDock Vina automatically calculates the grid maps and clusters the results in a way transparent to the user.

© 2009 Wiley Periodicals, Inc. J Comput Chem 00: 000–000, 2009

Key words: AutoDock; molecular docking; virtual screening; computer-aided drug design; multithreading; scoring function

Introduction

Molecular docking is a computational procedure that attempts to predict noncovalent binding of macromolecules or, more frequently, of a macromolecule (receptor) and a small molecule (ligand) efficiently, starting with their unbound structures, structures obtained from MD simulations, or homology modeling, etc. The goal is to predict the bound conformations and the binding affinity.

The prediction of binding of small molecules to proteins is of particular practical importance because it is used to screen virtual libraries of drug-like molecules to obtain leads for further drug development. Docking can also be used to try to predict the bound conformation of known binders, when the experimental holo structures are unavailable.[1]

One is interested in maximizing the accuracy of these predictions while minimizing the computer time they take, because the computational resources spent on docking are considerable. For example, hundreds of thousands of computers are used for running docking in `FightAIDS@Home` and similar projects.[2]

Theory

In the spectrum of computational approaches to modeling receptor-ligand binding,

a. molecular dynamics with explicit solvent,
b. molecular dynamics and molecular mechanics with implicit solvent, and
c. molecular docking

can be seen as making an increasing trade-off of the representational detail for computational speed.[3]

Among the assumptions made by these approaches is the commitment to a particular protonation state of and charge distribution in the molecules that do not change between, for example, their bound and unbound states. Additionally, docking generally assumes much or all of the receptor rigid, the covalent lengths, and angles constant, while considering a chosen set of covalent bonds freely rotatable (referred to as active rotatable bonds here).

Importantly, although molecular dynamics directly deals with energies (referred to as force fields in chemistry), docking is ultimately interested in reproducing chemical potentials, which determine the bound conformation preference and the free energy of binding. It is a qualitatively different concept governed not only by the minima in the energy profile but also by the shape of the profile and the temperature.[4,5]

Docking programs generally use a scoring function, which can be seen as an attempt to approximate the standard chemical potentials of the system. When the superficially physics-based terms like the 6–12 van der Waals interactions and Coulomb energies are used in the scoring function, they need to be significantly empirically weighted, in part, to account for this difference between energies and free energies.[4,5]

Correspondence to: A.J. Olson; e-mail: olson@scripps.edu
Contract/grant sponsor: NIH; contract/grant number: 2R01GM069832

O. Trott, A. J. Olson, Journal of Computational Chemistry (2009)

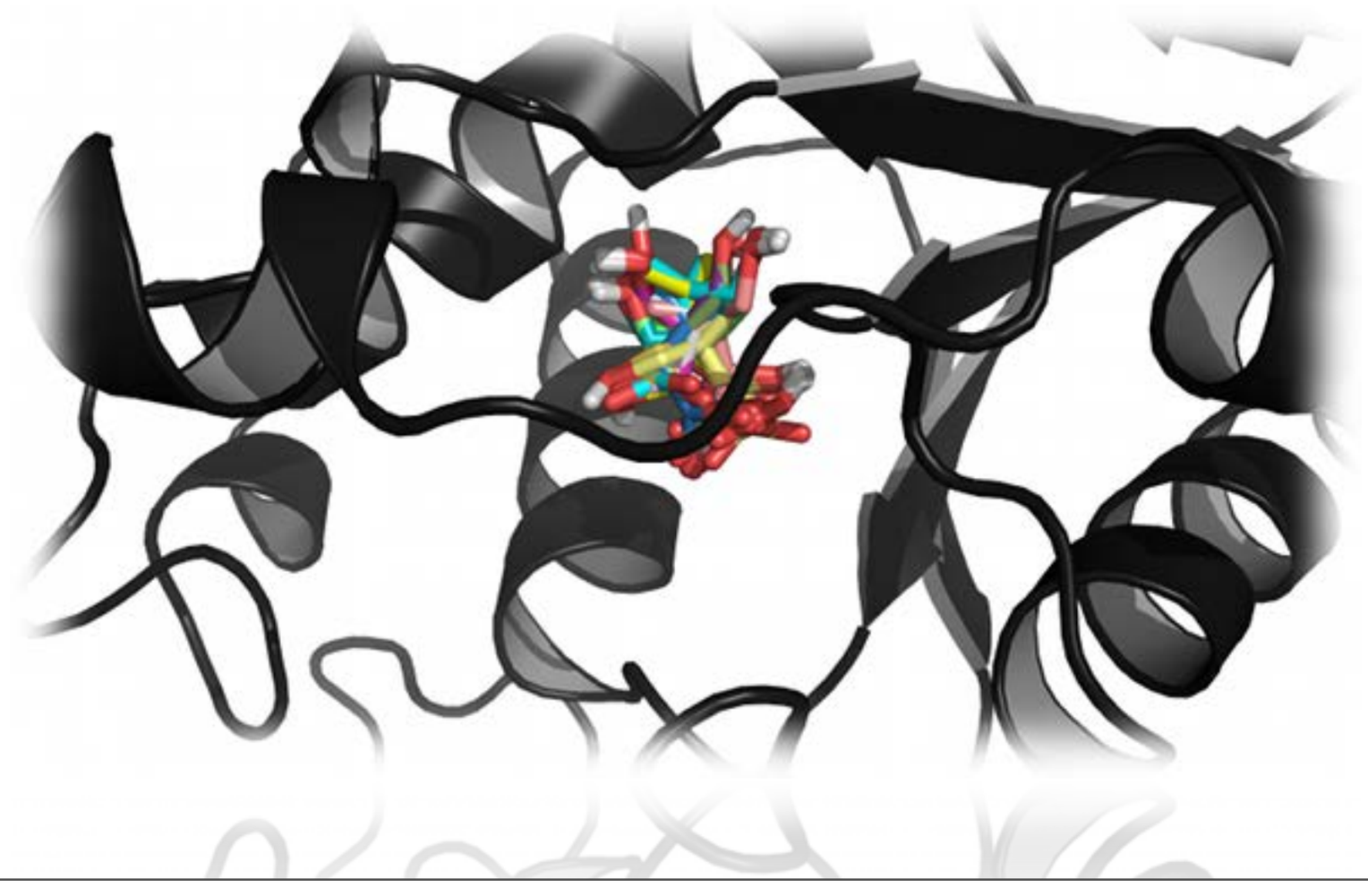
http://autodock.scripps.edu

http://vina.scripps.edu

What is docking?

Predicting the best ways two molecules interact.

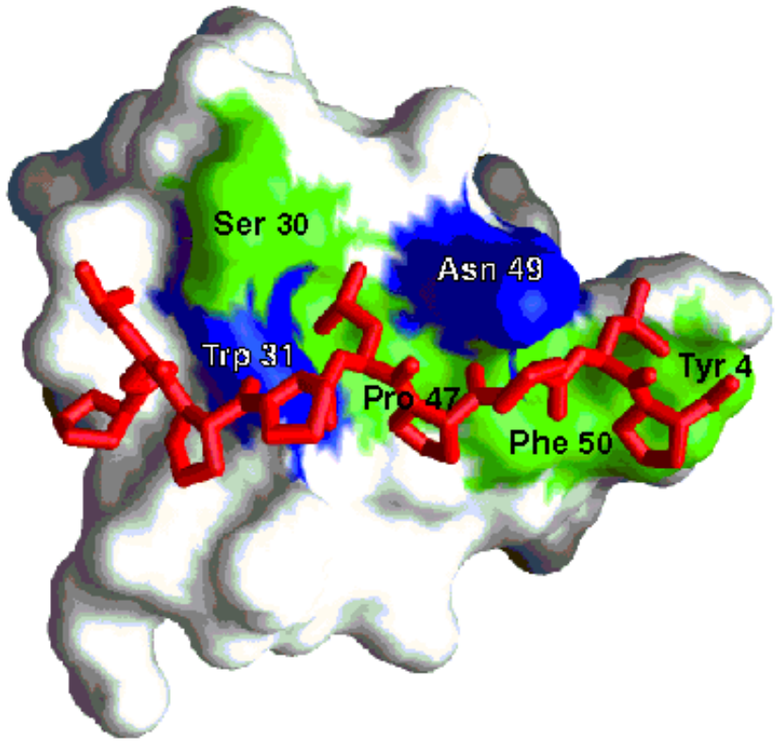
- Obtain the 3D structures of the two molecules.
- Locate the best binding site (**Remember AnnoLyze, Metapocket...**)
- Here, small molecule docking in protein.
- **Determine the best binding mode.** (POSE) .



What is docking?

Predicting the **best** ways two molecules interact.

- We need to quantify or rank solutions
- We need a good scoring function for such ranking
- Can we determine the best solution?

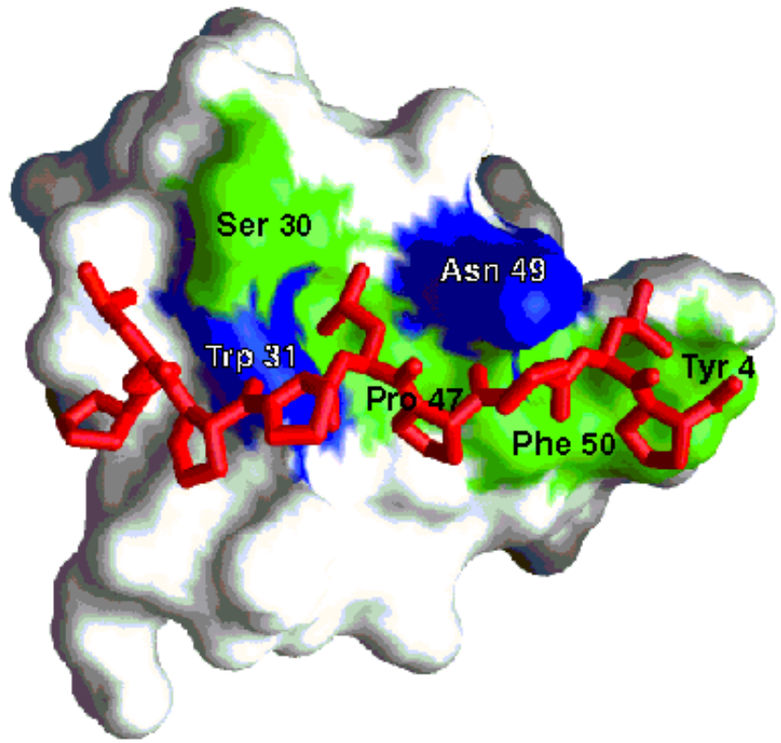


Scoring!

What is docking?

Predicting the best **ways** two molecules interact.

- X-ray and NMR structures are just ONE of the possible solutions
- There is a need for a search solution.
- Can we get all possible solutions?

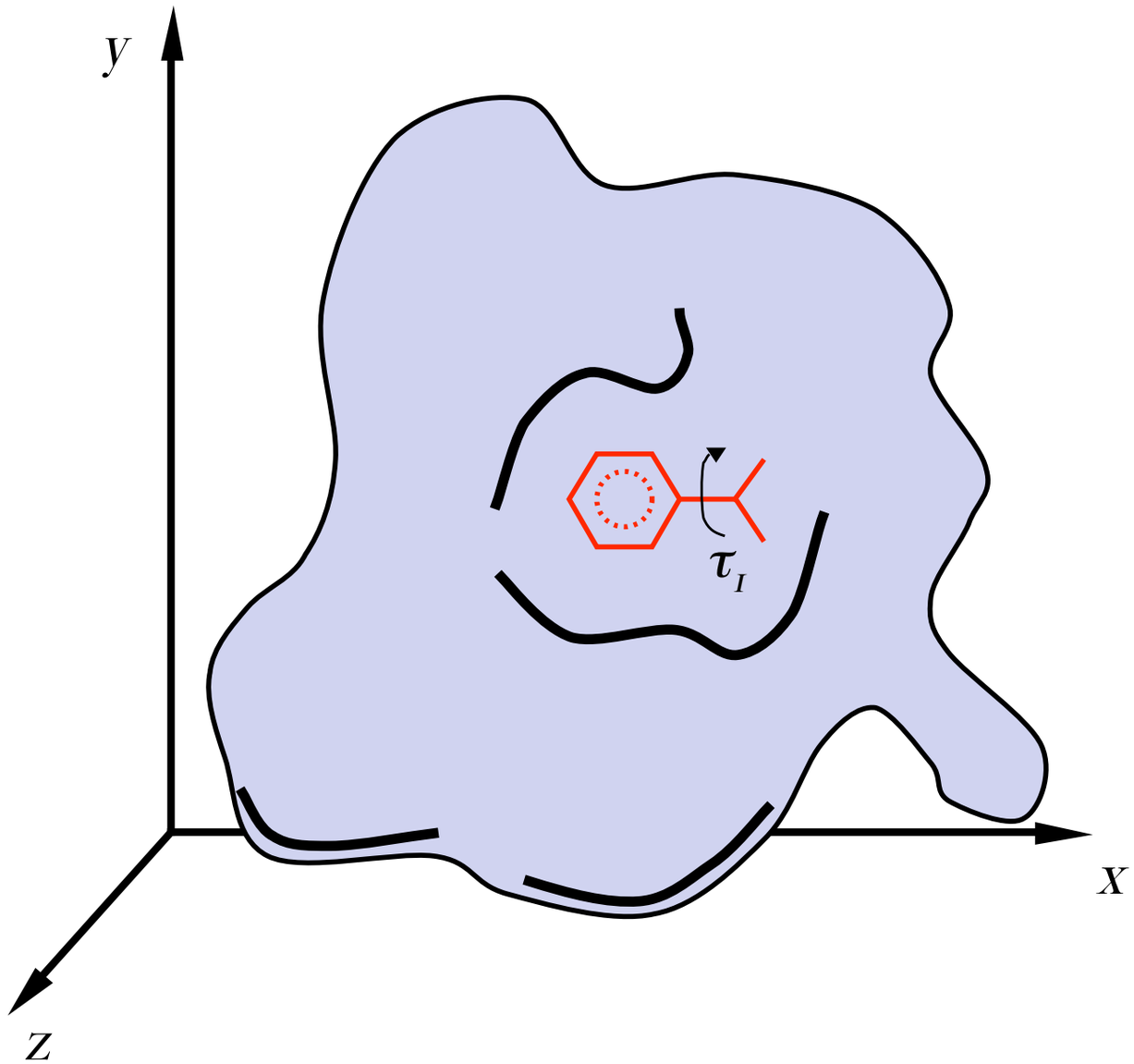


Sampling!

As everything in BIOINFORMATICS...

REPRESENTATION
SCORING
SAMPLING

REPRESENTATION



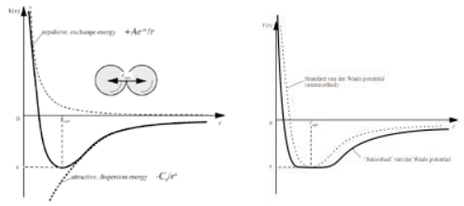
SCORING

AutoDock Vina

$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$$

- ΔG_{vdW}

  12-6 Lennard-Jones potential



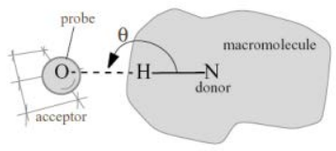
- ΔG_{elec}

  Coulombic with Solmajer-dielectric

$$\varepsilon(r) = A + \frac{B}{1 + ke^{-\lambda Br}}$$

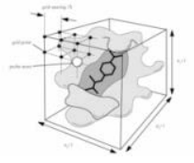
- ΔG_{hbond}

  12-10 Potential with Goodford Directionality



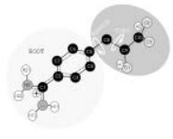
- ΔG_{desolv}

  Stouten Pairwise Atomic Solvation Parameters



- ΔG_{tors}

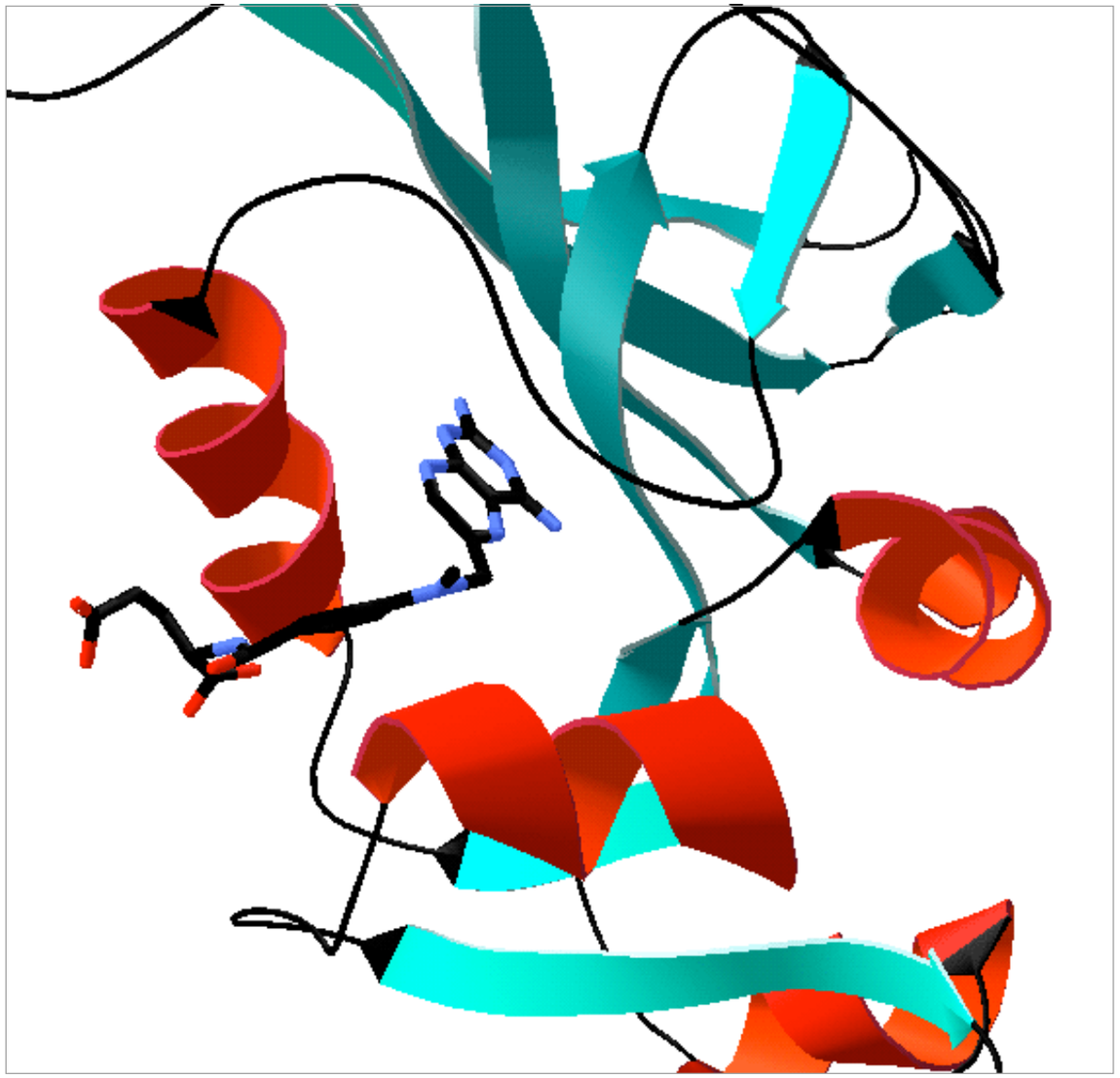
  Number of rotatable bonds



http://autodock.scripps.edu/resources/science/equations

PROBLEM!

Unaffordable CPU time...



Dihidrofolate reductase with a metotrexate (4dfr.pdb)

$$N=T^{360/i}$$

N: number of conformations

T: number of rotable bonds

I: incremental degrees

Metotrexato

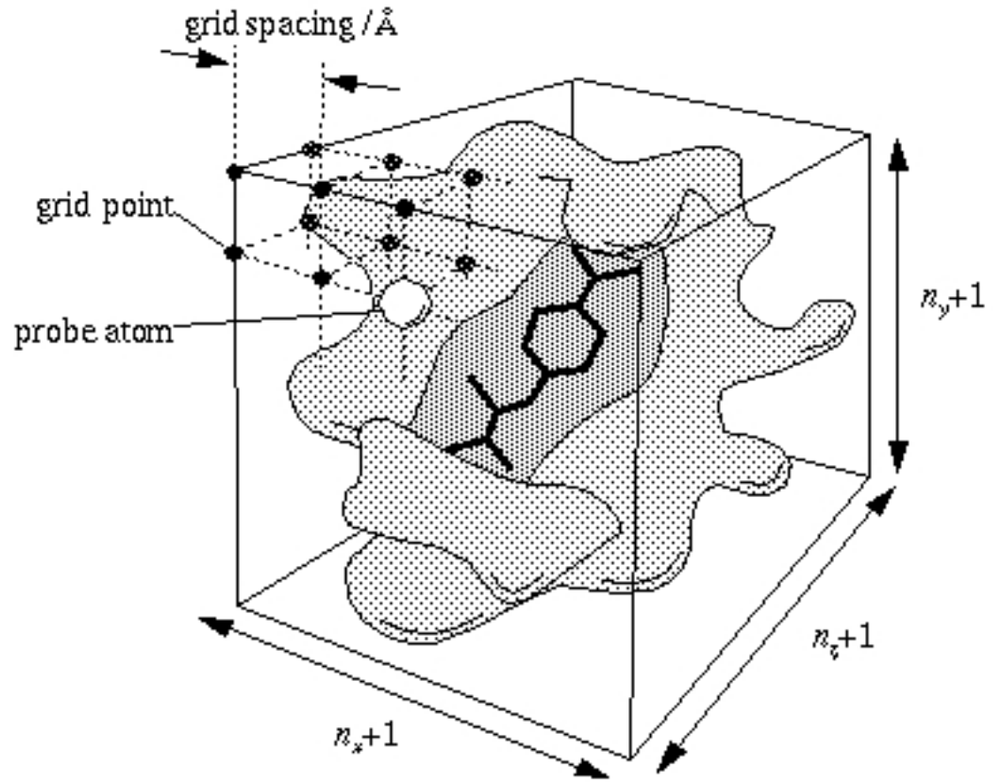
10 rotable bonds

30° increments (discrete)

10^{12} plausible conformations!

SOLUTION

Use of grid maps!

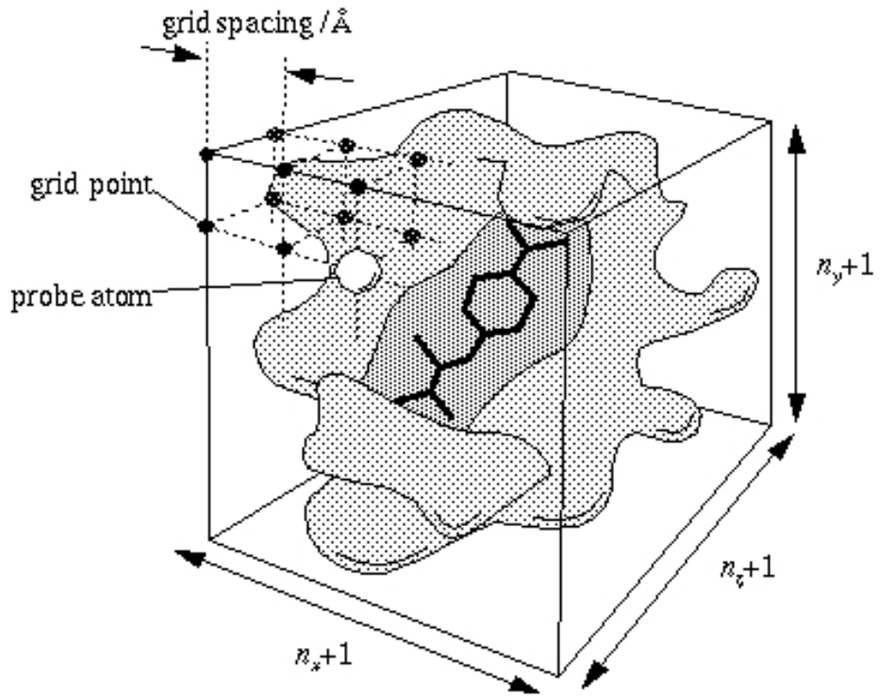


- Saves lots of time (compared to classical MM/MD).
- Need to map each atom to a grid point.
- Limits the search space!. From continue to discrete space.

AutoGrid Vina + ADT Tools

Use of grid maps!

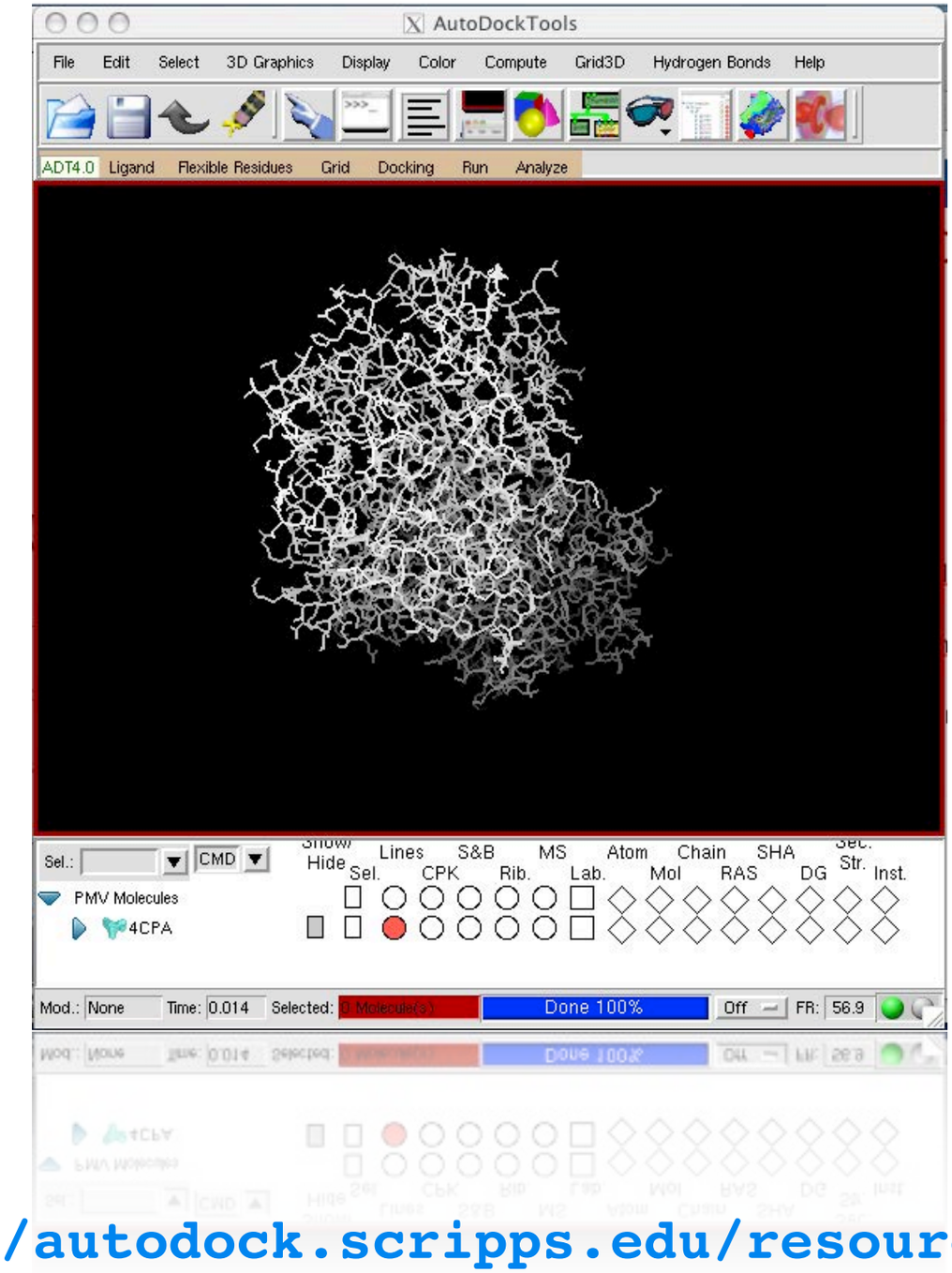
- Center of grid *
  - center of a bind ligand.
  - a selected atom or coordinate.
  - **Binding Site Center of Mass (CoM) .**
- Box dimension *
  - At least, **two** times the size of the ligand.
  - 3-Dimensions X,Y, Z.
- Grid resolution (spacing)
  - default **0.375 Angstroms**.
- Number of grid points (dimension)
  - use ONLY even numbers



With VINA + ADT Tools much simplified (*)

Vina + AutoDock Tools

Good that we have AutoDock Tools (ATD)

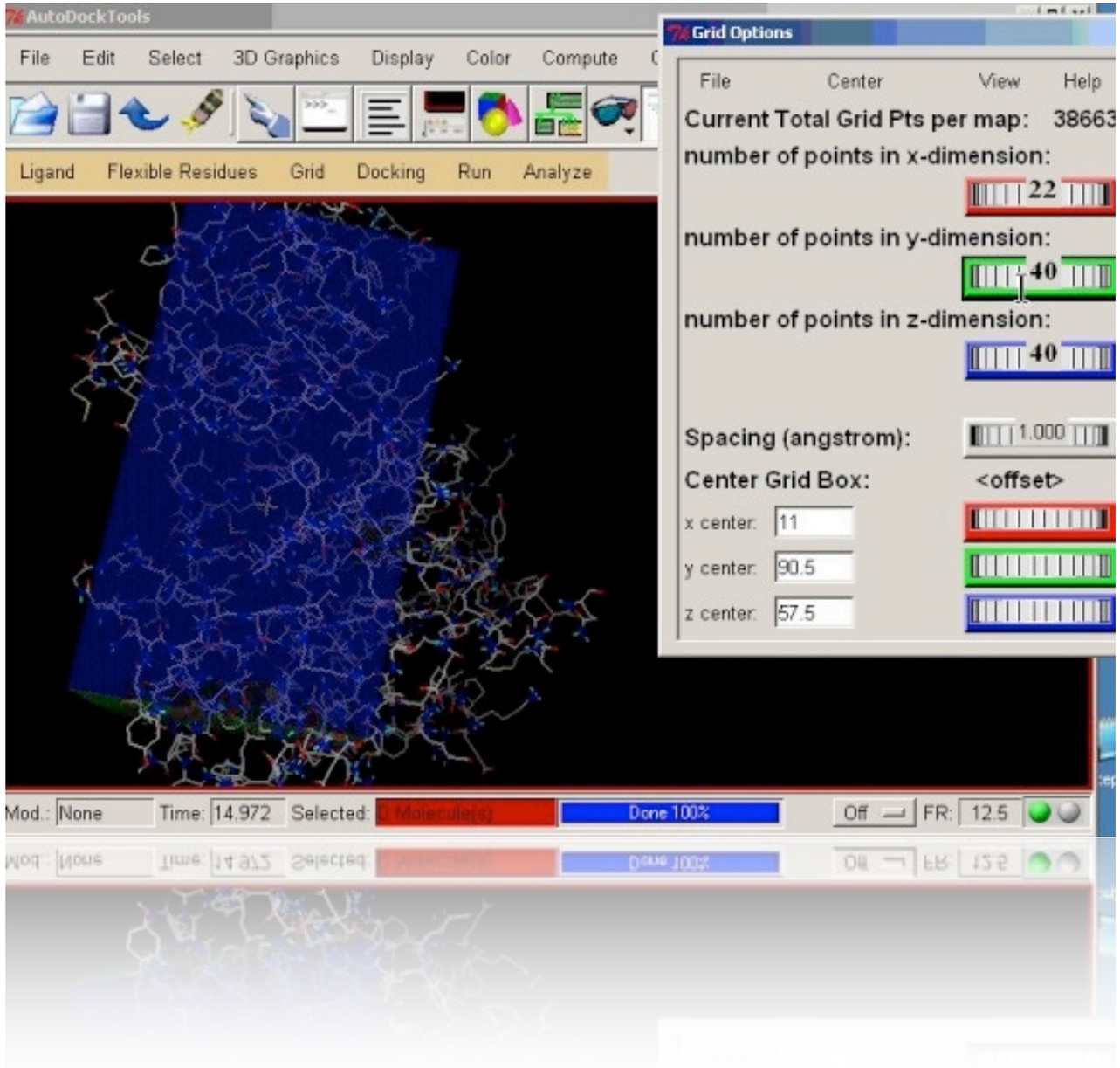


http://autodock.scripps.edu/resources/adt

AutoDock Tools

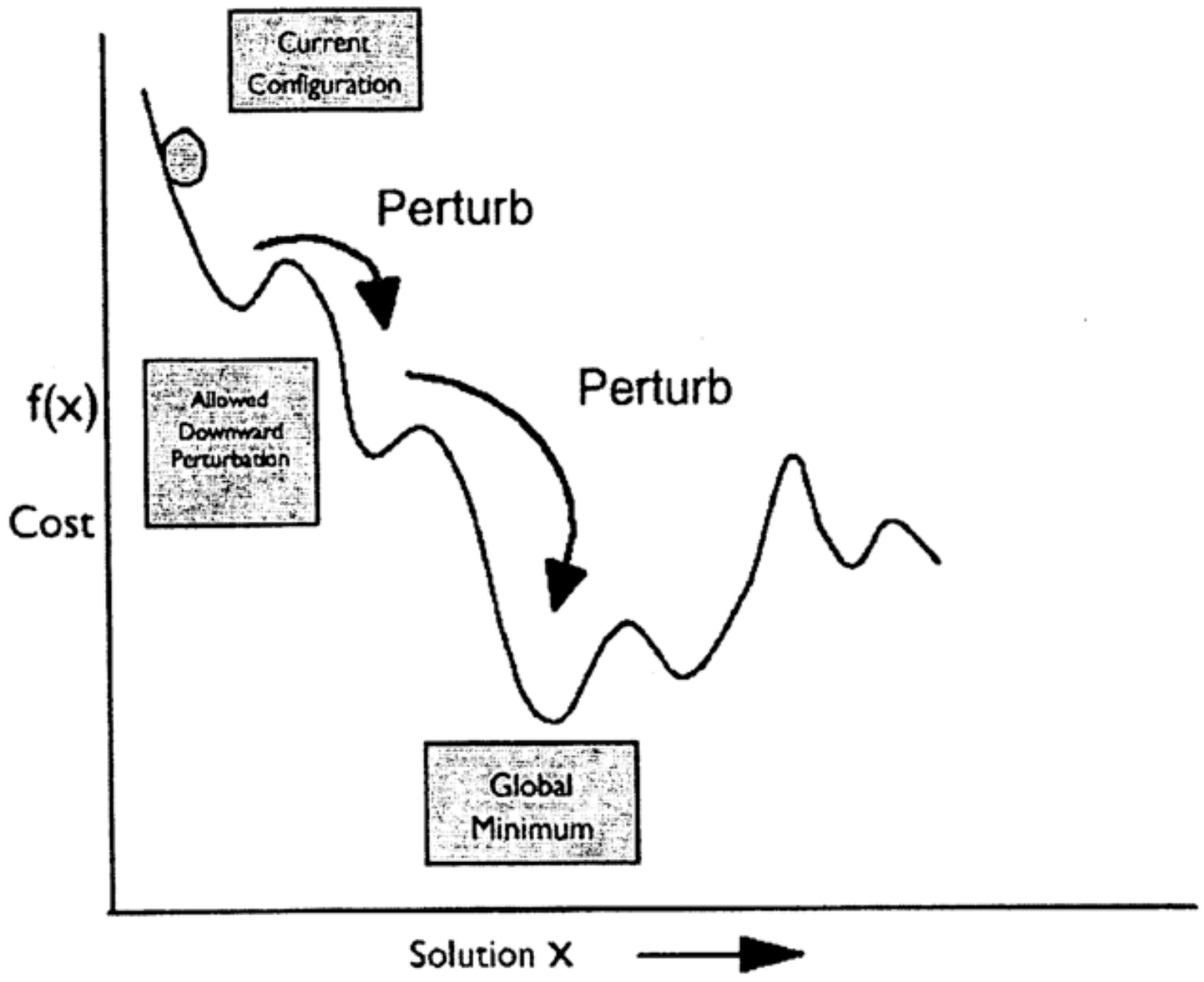
Very useful and easy to use!

- **Is the free GUI for AutoDock.**
- **We can use it for setting up grid size and grid position.**
- **We can also prepare the input molecules :**
  - **Adding all hydrogens or only polar hydrogens.**
  - **Assigning polar charges to the ligand and the receptor.**
  - **Set up rotatable bonds in the ligand using a graphical version of AutoTors.**
  - **Select the flexible side chains in flexible docking.**
- **Useful for analyzing the results, after vina docking.**



Search algorithms

Simulated Annealing

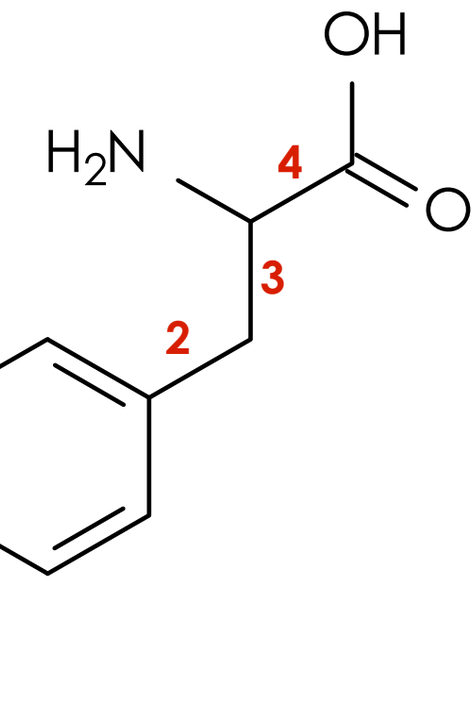


Search algorithms

Genetic Algorithm

Use of a Genetic Algorithm as a sampling method

- Each conformation is described as a set of rotational angles.
- 64 possible angles are allowed to each of the bond in the ligand.
- Each plausible dihedral angle is codified in a set of binary bits (2^6=64)
- Each conformation is codified by a so called chromosome with 4 × 6 bits (0 or 1)



111010.010110.001011.010010

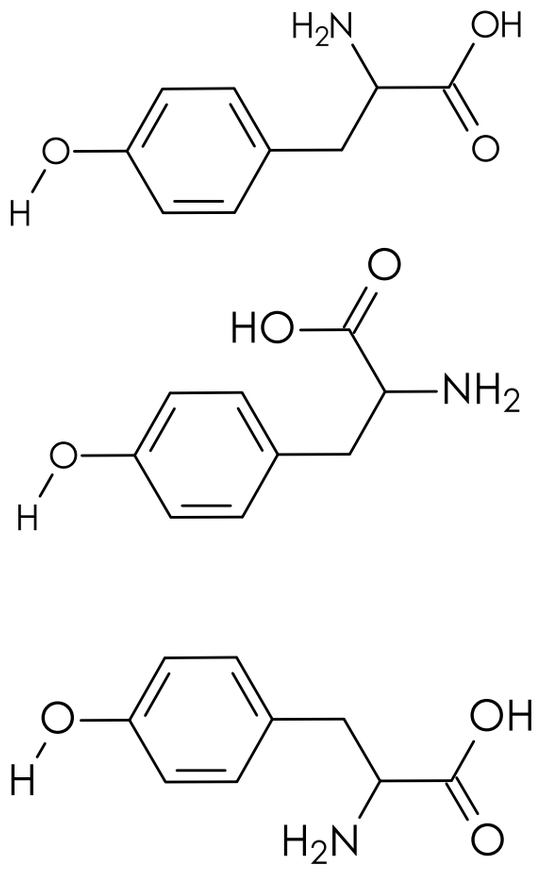
Φ_1 Φ_2 …

$$\Phi_1 = 1\times2^5 + 1\times2^4 + 1\times2^3 + 0\times2^2 + 1\times2^1 + 0\times2^0 = 58°$$

Search algorithms

Genetic Algorithm

Population (*ie*, set of chromosomes or configurations)



011010.010110.011010.010111 ← Chromosome

111010.010110.001011.010010

001010.010101.000101.010001

101001.101110.101010.001000

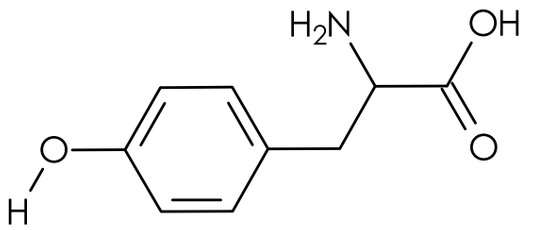
001010.101000.011101.001011

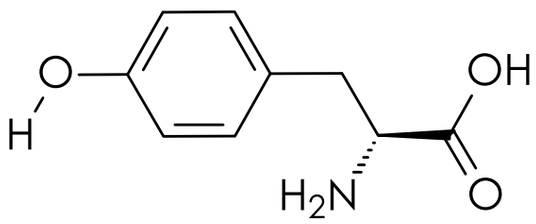
Gene

Search algorithms

Genetic Algorithm

Genetic operators...





011010.010110.011010.010111

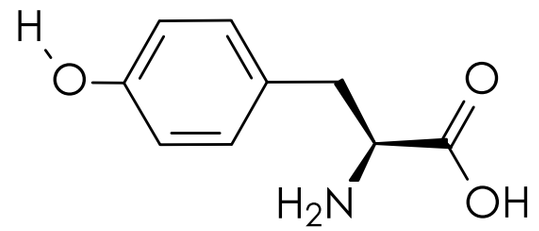
Single mutation

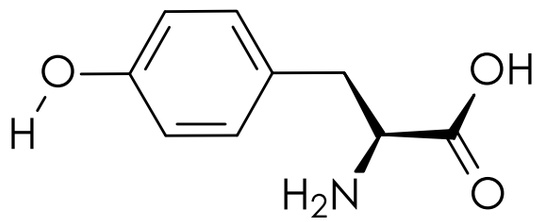
011010.011110.011110.010111

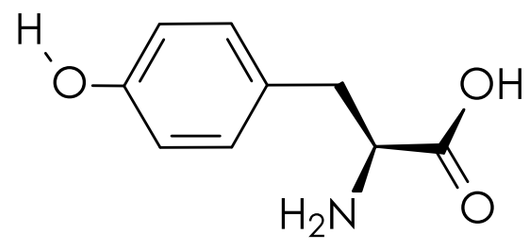
Search algorithms

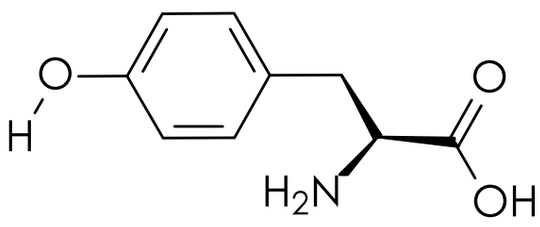
Genetic Algorithm

Genetic operators...









001010.010101.000101.010001

011010.010110.011010.010111

Recombination

001010.010101.011010.010111

011010.010110. 000101.010001

Search algorithms

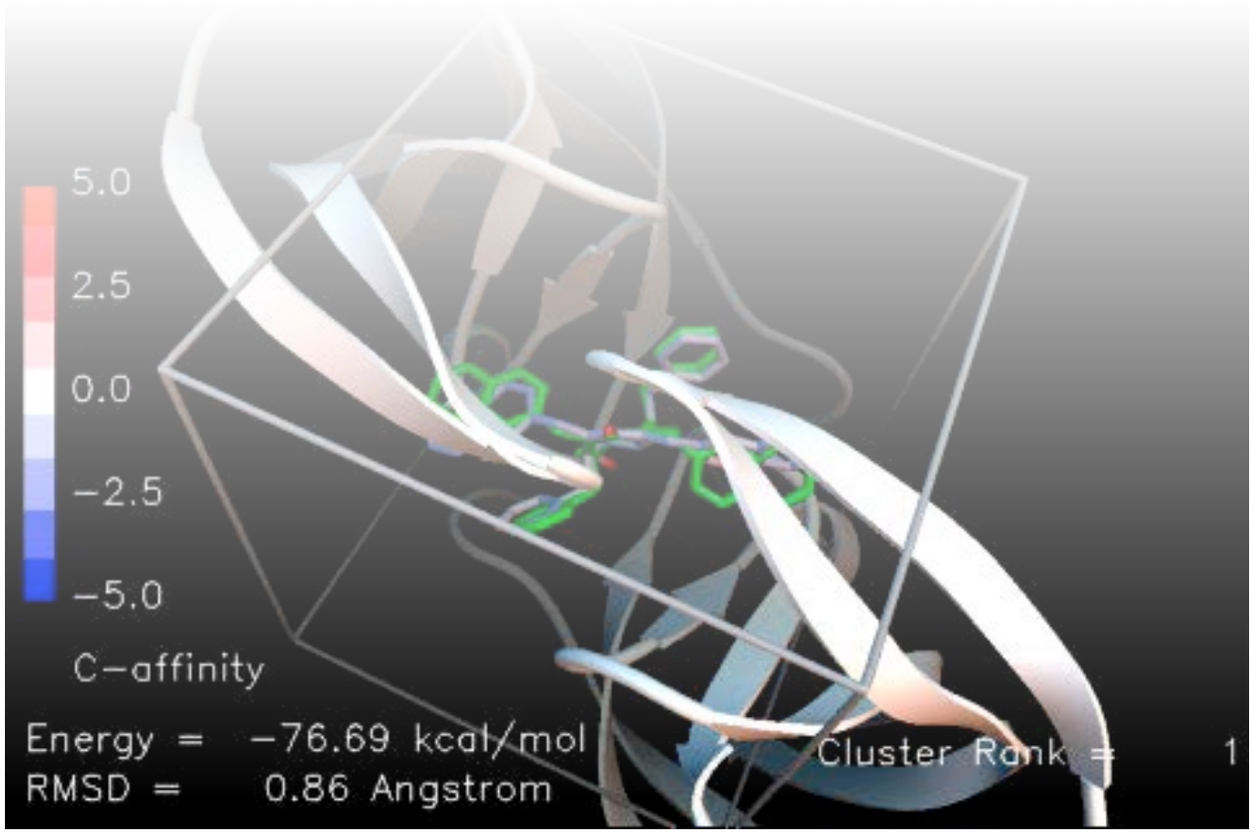
Genetic Algorithm

Genetic operators...

Vina docking results

Goodsell, D. S. and Olson, A. J. (1990), Automated Docking of Substrates to Proteins by Simulated Annealing Proteins:Structure, Function and Genetics., 8: 195-202.
Morris, G. M., et al. (1996), Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4 J. Computer-Aided Molecular Design, 10: 293-304.
Morris, G. M., et al. (1998), Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function J. Computational Chemistry, 19: 1639-1662.
Huey, R., et al. (2007), A Semiempirical Free Energy Force Field with Charge-Based Desolvation J. Computational Chemistry, 28: 1145-1152.

Vina docking results



Goodsell, D. S. and Olson, A. J. (1990), Automated Docking of Substrates to Proteins by Simulated Annealing Proteins:Structure, Function and Genetics., 8: 195-202.
Morris, G. M., et al. (1996), Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4 J. Computer-Aided Molecular Design, 10: 293-304.
Morris, G. M., et al. (1998), Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function J. Computational Chemistry, 19: 1639-1662.
Huey, R., et al. (2007), A Semiempirical Free Energy Force Field with Charge-Based Desolvation J. Computational Chemistry, 28: 1145-1152.

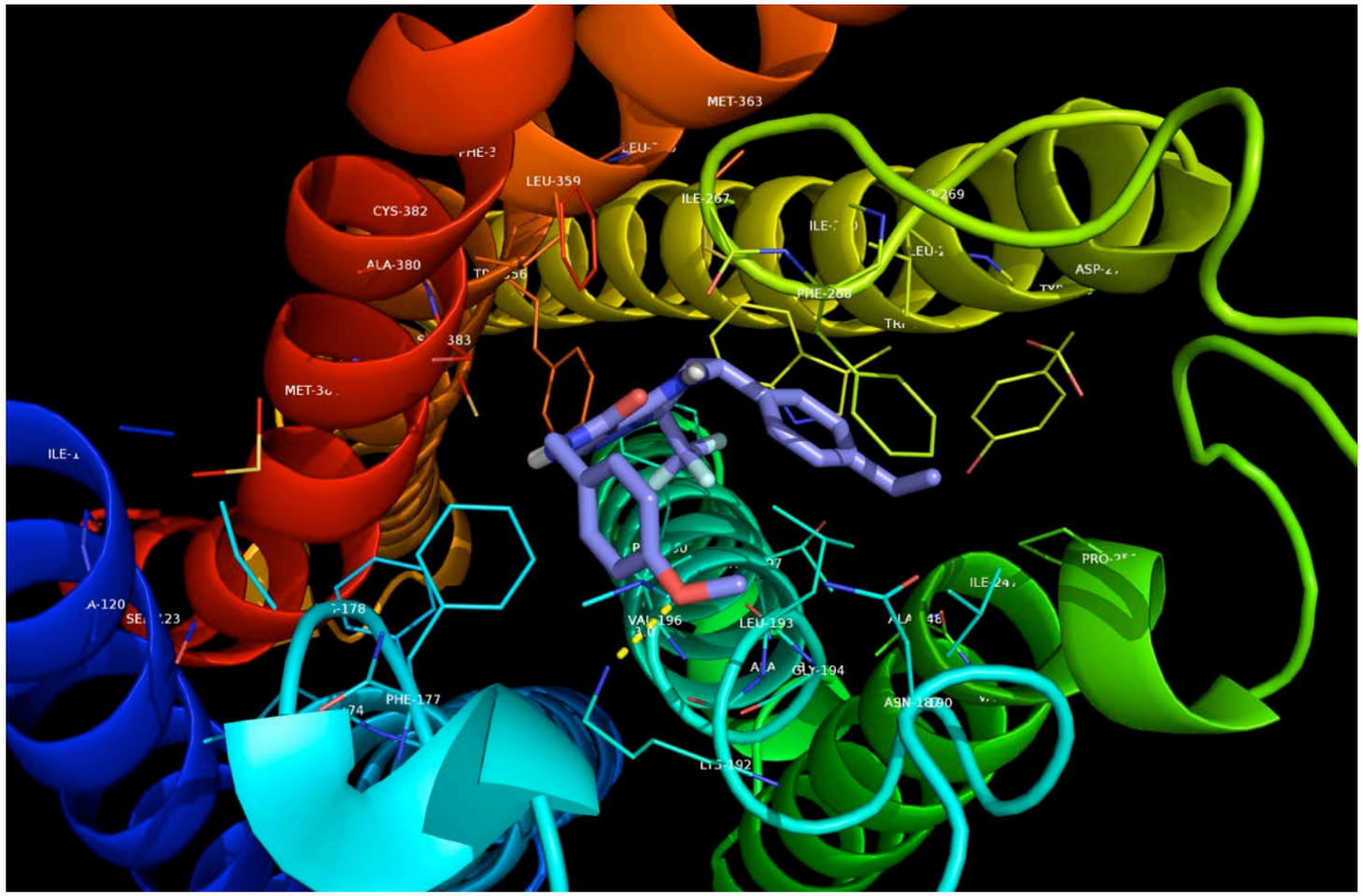
Vina docking results

One practical case...

Vina output log

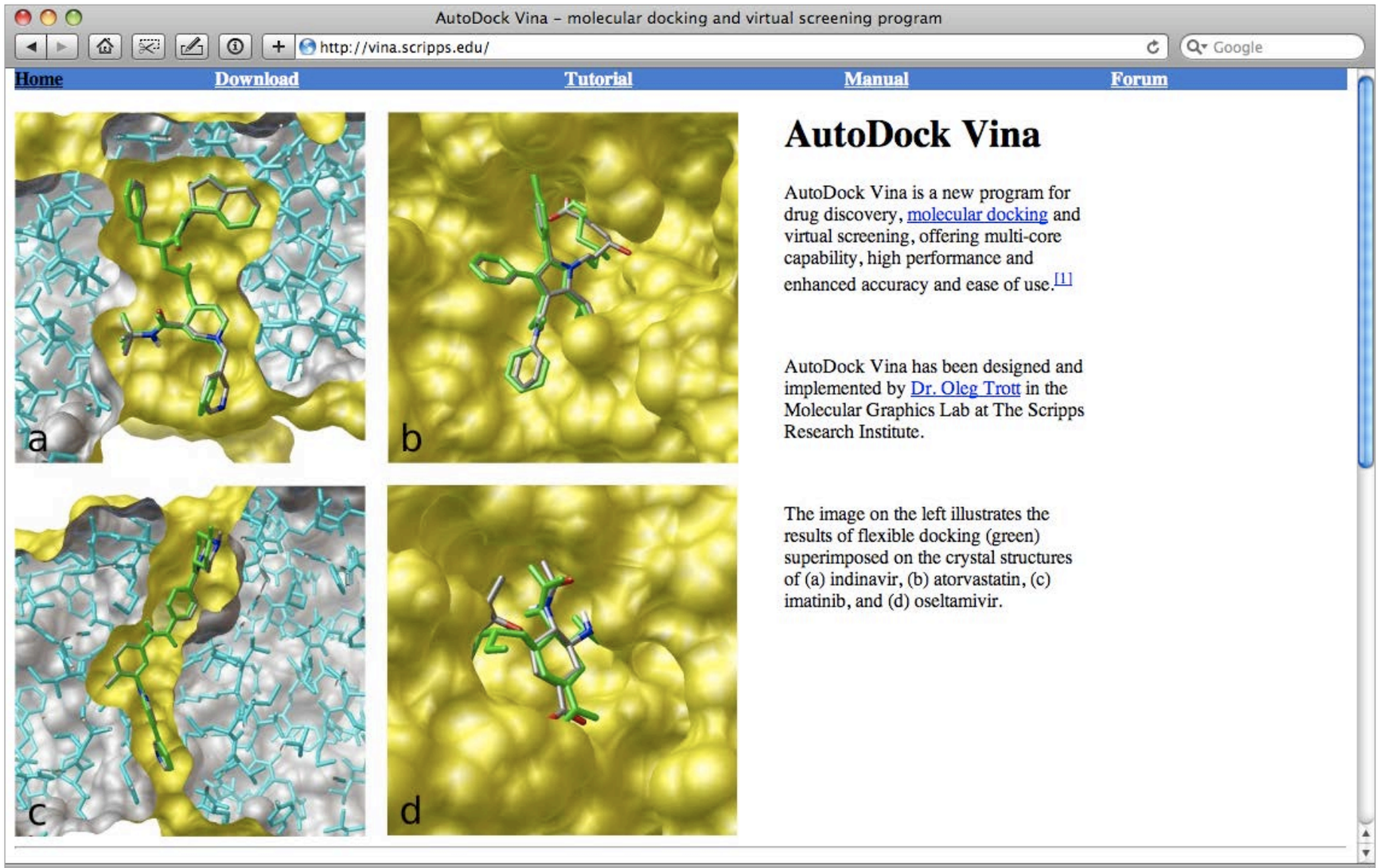
```
# If you used AutoDock Vina in your work, please cite:          #
#                                                               #
# O. Trott, A. J. Olson,                                        #
# AutoDock Vina: improving the speed and accuracy of docking    #
# with a new scoring function, efficient optimization and       #
# multithreading, Journal of Computational Chemistry 31 (2010)  #
# 455-461                                                       #
#                                                               #
# DOI 10.1002/jcc.21334                                         #
#                                                               #
# Please see http://vina.scripps.edu for more information.      #
#################################################################

WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Detected 4 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 682849564
Performing search ... done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -11.3      0.000      0.000
   2        -10.4      1.077      2.294
   3        -10.2      1.327      2.006
   4        -10.0      2.334      4.484
   5         -9.9     14.488     16.499
   6         -9.9      1.542      3.005
   7         -9.8     36.046     37.733
   8         -9.8     36.084     37.975
   9         -9.8     32.479     34.497
Writing output ... done.
fran@davide-desktop:~/Documents/TestProject/autodock_vina_1_1_2_linu
```

HCBR + Rimonabant



AutoDock Vina

Where to get help...



http://vina.scripps.edu

Vina 1.1.1

Alternatives

Progressive building

- FLEXX
- DOCK
- GROW
- GroupBUILD
- LUDI
- LEGEND
- SPROUT
- BUILDER
- GENSTAR

Conformational search

- MIMUMBA
- COBRA
- WIZRAD

Binding site description

- GRID

Genetic algorithms

- GOLD
- Others

Virtual screening

- AutoDOCK

Molecular dynamics

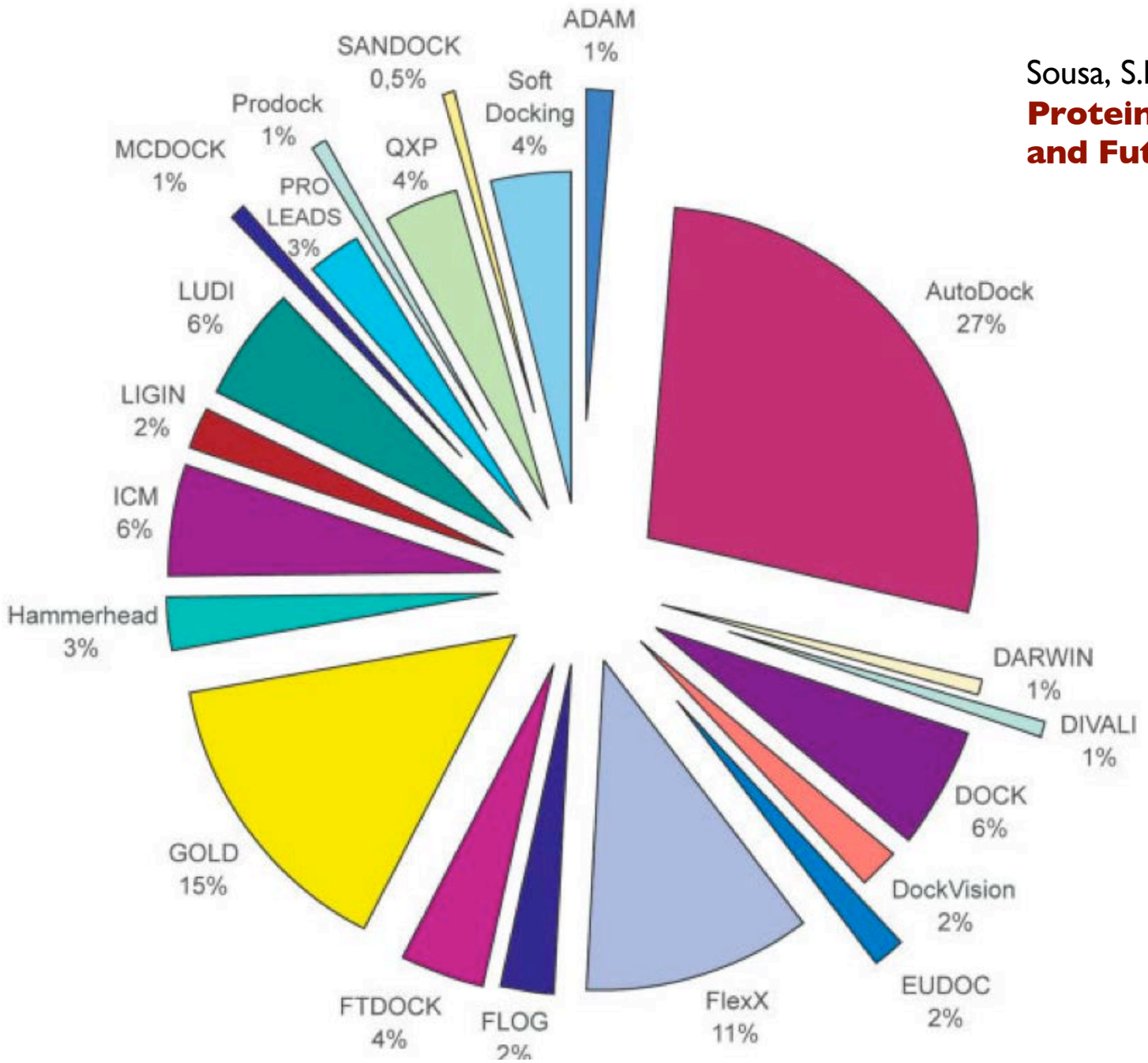
- MCSS
- CONCEPTS

Databases

- CAVEAT
- FOUNDATION
- CLIX
- NEWLEAD
- LEAPFROG

AutoDock 4.0

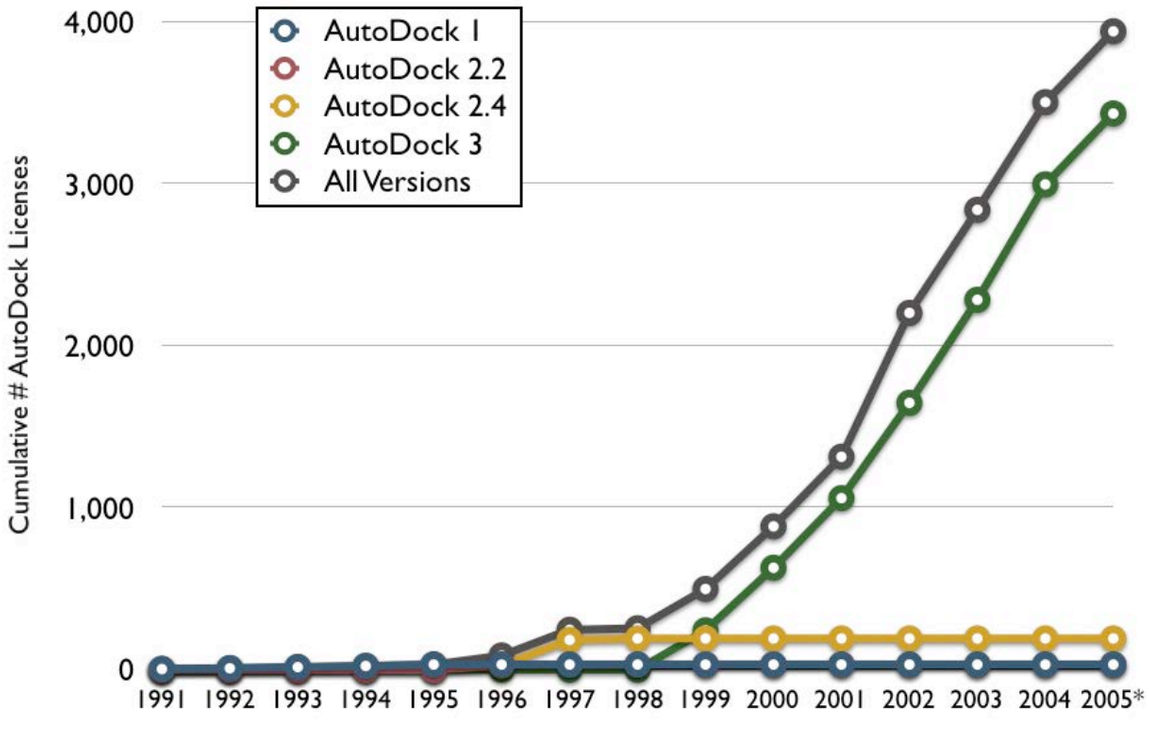
Why AutoDock over others

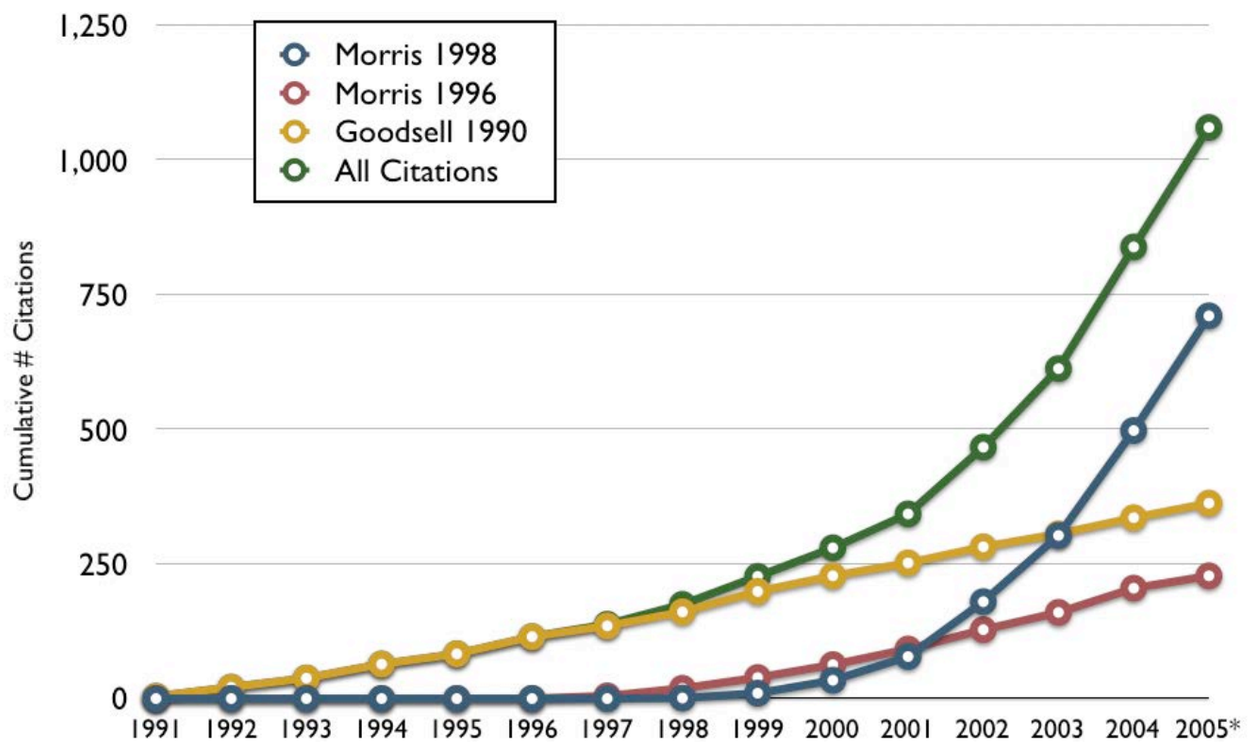


Sousa, S.F., Fernandes, P.A. & Ramos, M.J. (2006)
Protein-Ligand Docking: Current Status and Future Challenges *Proteins*, **65**:15-26

AutoDock 4.0

Why AutoDock over others

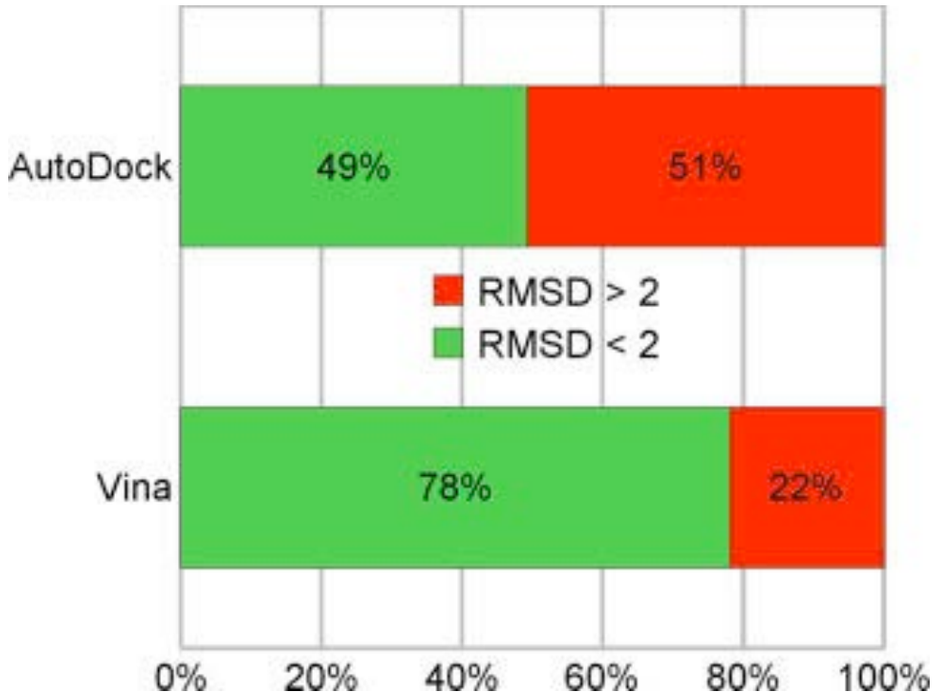




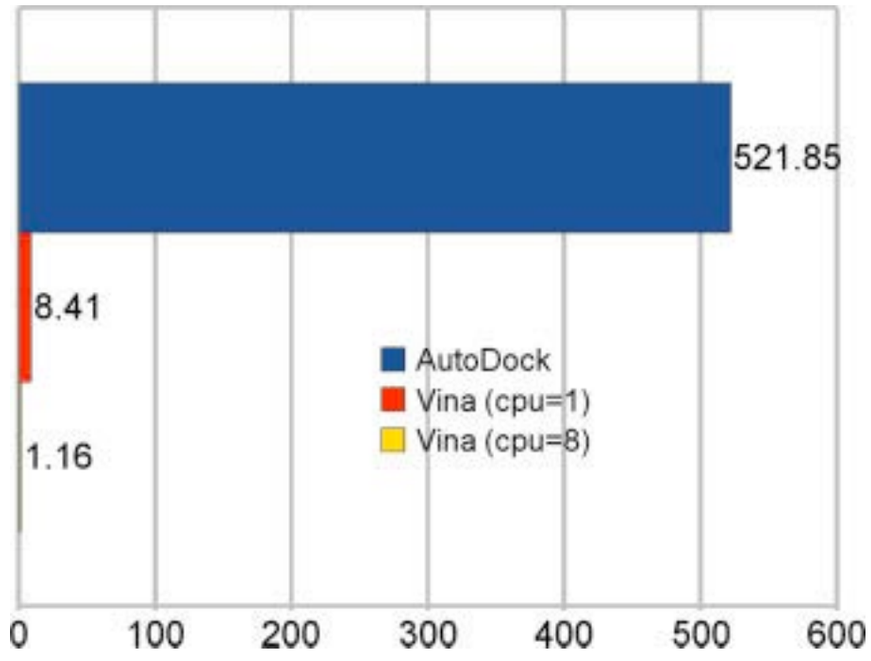
Vina vs. Autodock 4

Important improvements...

More accurate



4-fold faster



O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461

Improvements of Vina

Recently published...

Handoko, S. D., Xuchang Ouyang, Chinh Tran To Su, Chee Keong Kwoh & Yew Soon Ong. *IEEE/ACM Trans. Comput. Biol. and Bioinf.* **9,** 1266–1272

QuickVina: Accelerating AutoDock Vina Using Gradient-Based Heuristics for Global Optimization

Stephanus Daniel Handoko, Xuchang Ouyang, Chinh Tran To Su, Chee Keong Kwoh, and Yew Soon Ong

Abstract—Predicting binding between macromolecule and small molecule is a crucial phase in the field of rational drug design. AutoDock Vina, one of the most widely used docking software released in 2009, uses an empirical scoring function to evaluate the binding affinity between the molecules and employs the iterated local search global optimizer for global optimization, achieving a significantly improved speed and better accuracy of the binding mode prediction compared its predecessor, AutoDock 4. In this paper, we propose further improvement in the local search algorithm of Vina by heuristically preventing some intermediate points from undergoing local search. Our improved version of Vina—dubbed QVina—achieved a maximum acceleration of about 25 times with the average speed-up of 8.34 times compared to the original Vina when tested on a set of 231 protein-ligand complexes while maintaining the optimal scores mostly identical. Using our heuristics, larger number of different ligands can be quickly screened against a given receptor within the same time frame.

Index Terms—Artificial intelligence, bioinformatics, global optimization, gradient methods.

1 Background

Molecular docking is a computational process trying to find the binding between a macromolecule (the receptor) and a small molecule (the ligand). Since it can be used in predicting binding conformations and affinities between drug molecules and their target proteins, leading to the understanding of the biological mechanism behind those bindings, molecular docking is with great value to drug design [1].

Generally, docking is an optimization problem that attempts to find the binding conformation with global lowest energy, the landscape of which is approximated by a scoring function. The introduction of flexibility in the ligand, or further in the receptor as well, will make the problem more sophisticated [1], [2]. The major issue of the difficulty comes from the large number of degrees of freedom in modeling the molecular system. Since 1980s, various programs and software have been developed in order to perform molecular binding, such as DOCK [1], AutoDock [3], GOLD [4], ICM [5], and FlexX [6] and different scoring functions have been proposed. However, after decades of development, docking is still a time-consuming task even with the most powerful computing resources to-date. In 2009, AutoDock Vina [7] (referred to as Vina afterward) was released by the same group who invented the earlier versions of AutoDock, which is one of the most popular docking software. Vina uses an empirical scoring function to evaluate the binding affinity between the molecules, and the iterated local search global optimizer for global optimization. This combination is reported to be successful to achieve approximately two orders of magnitude improvement in speed, and simultaneously, a significantly better accuracy of the binding mode prediction compared to AutoDock 4 [7].

In this paper, we proposed an improvement in the local search procedure of Vina. By heuristically preventing some of the intermediate points from performing local search, our improved version of Vina, named QuickVina (QVina), achieved a maximum speed-up of about 25 times with an average speed-up of 8.34 over a testing data set of 231 protein-ligand complexes from the PDBBind [8] and a tendency to have a higher speed-up with the larger number of degrees of freedom, without compromising the quality of docking result.

- *S.D. Handoko is with the Centre for Computational Intelligence, School of Computer Engineering, Nanyang Technological University, Blk N4, #B1a-02, Nanyang Avenue, Singapore 639798. E-mail: sdhandoko@ntu.edu.sg.*
- *X. Ouyang and C.T.T. Su are with the BioInformatics Research Centre, School of Computer Engineering, Nanyang Technological University, Blk NS4, #04-33, Nanyang Avenue, Singapore 639798. E-mail: {xouyang1, sutr0003}@e.ntu.edu.sg.*
- *C.K. Kwoh and Y.S. Ong are with the School of Computer Engineering, Nanyang Technological University, Blk N4, #02a-26, Nanyang Avenue, Singapore 639798. E-mail: {asckkwoh, asysong}@ntu.edu.sg.*

Manuscript received 8 Nov. 2011; revised 11 Mar. 2012; accepted 20 Apr. 2012; published online 23 May 2012.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBBSI-2011-11-0290.
Digital Object Identifier no. 10.1109/TCBB.2012.82.

Published by the IEEE CS, CI, and EMB Societies & the ACM

2.1 Analyzing the Global Optimization Algorithm in Vina

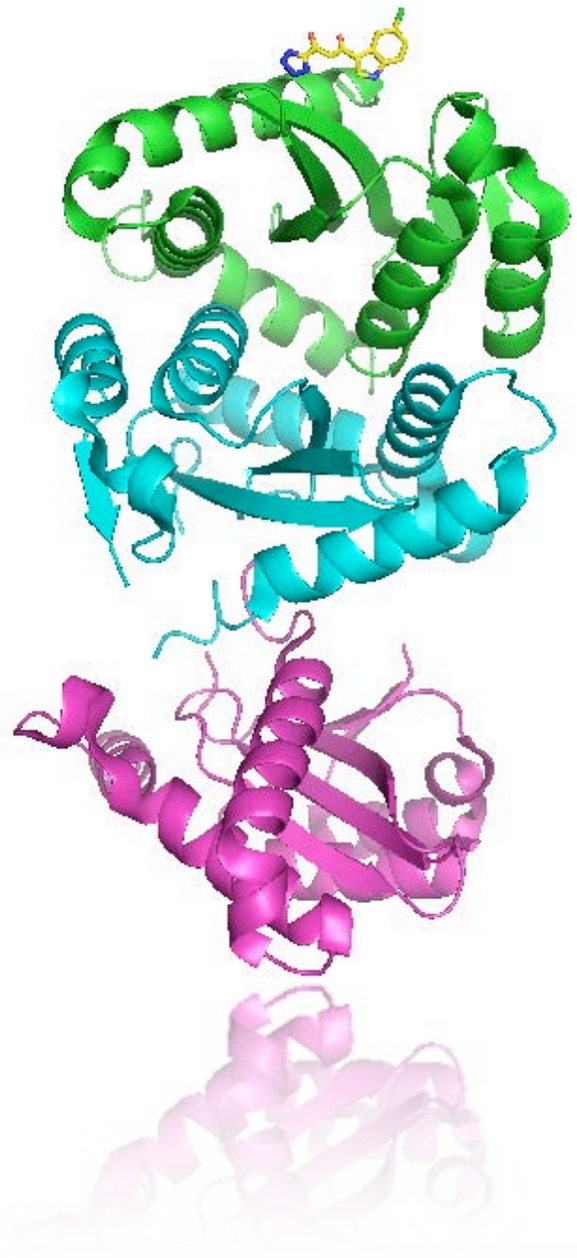
At the time this paper is drafted, the source code of the AutoDock Vina is available free of charge at its website: http://vina.scripps.edu/. With the lack of detailed explanation on how exactly the search algorithm works in Vina, we performed a thorough analysis of the source code. In Fig. 1, we present the pseudocode of the global optimization approach employed by Vina. Fundamentally, it is a form of

AutoDock Example

Discovery of a novel binding trench in HIV Integrase

Schames, J.R., R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, and J.A. McCammon, Discovery of a novel binding trench in HIV integrase. J Med Chem, 2004. 47(8): 1879-81

ISENTRESS example

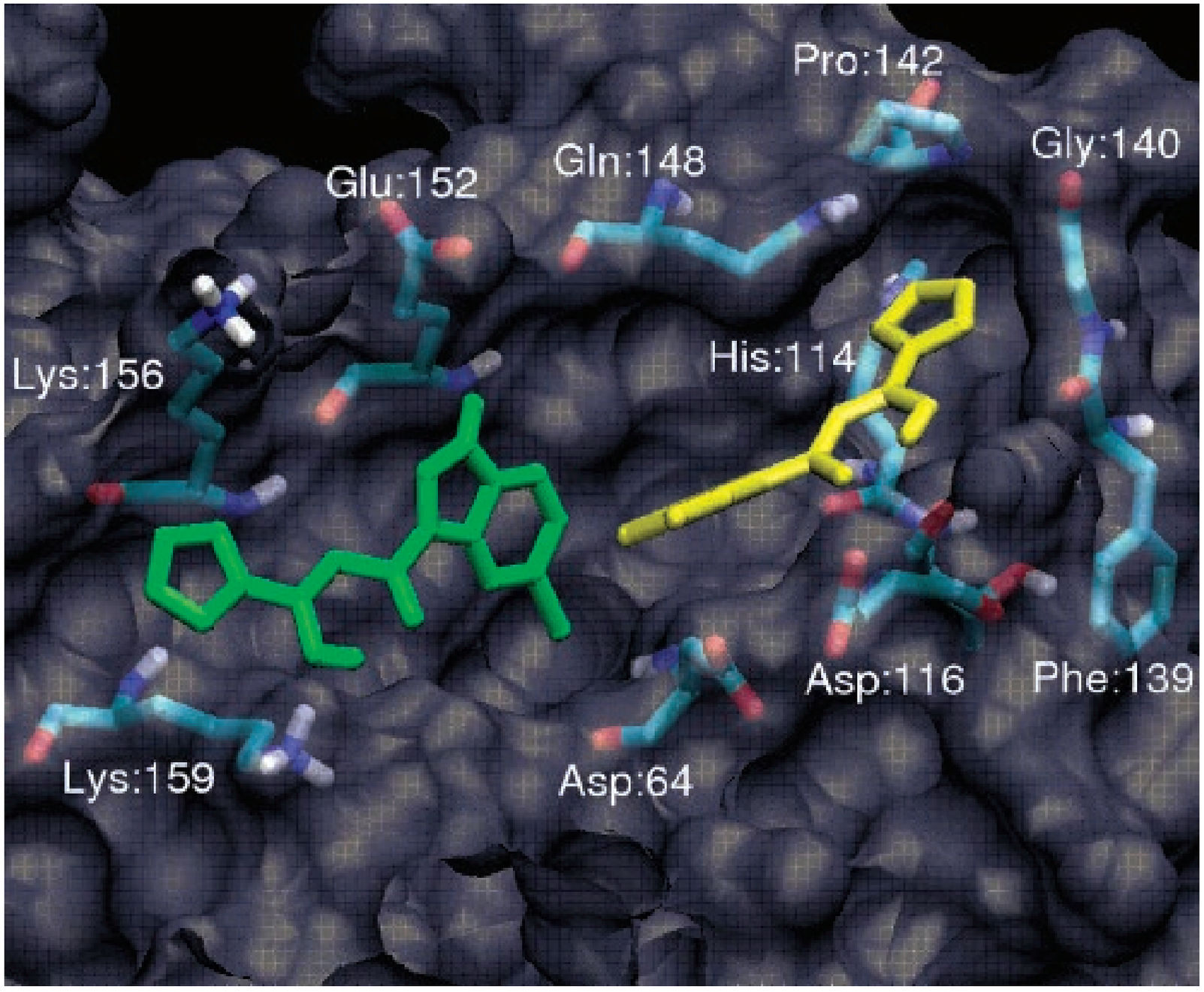


- One structure known with 5CITEP
  - Not clear (low resolution)
  - Binding site near to DNA interacting site
  - Loop near the binding

- Docking + Molecular Dynamics
  - AMBER snapshots
  - AutoDock flexible torsion thetetrazolering and indole ring.

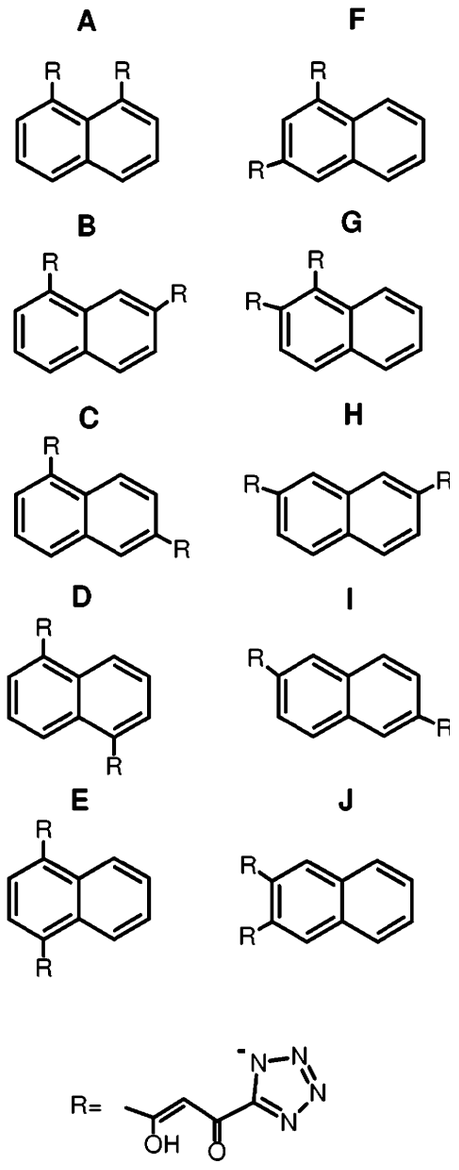
Schames, J.R., R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, and J.A. McCammon, Discovery of a novel binding trench in HIV integrase. J Med Chem, 2004. 47(8): 1879-81

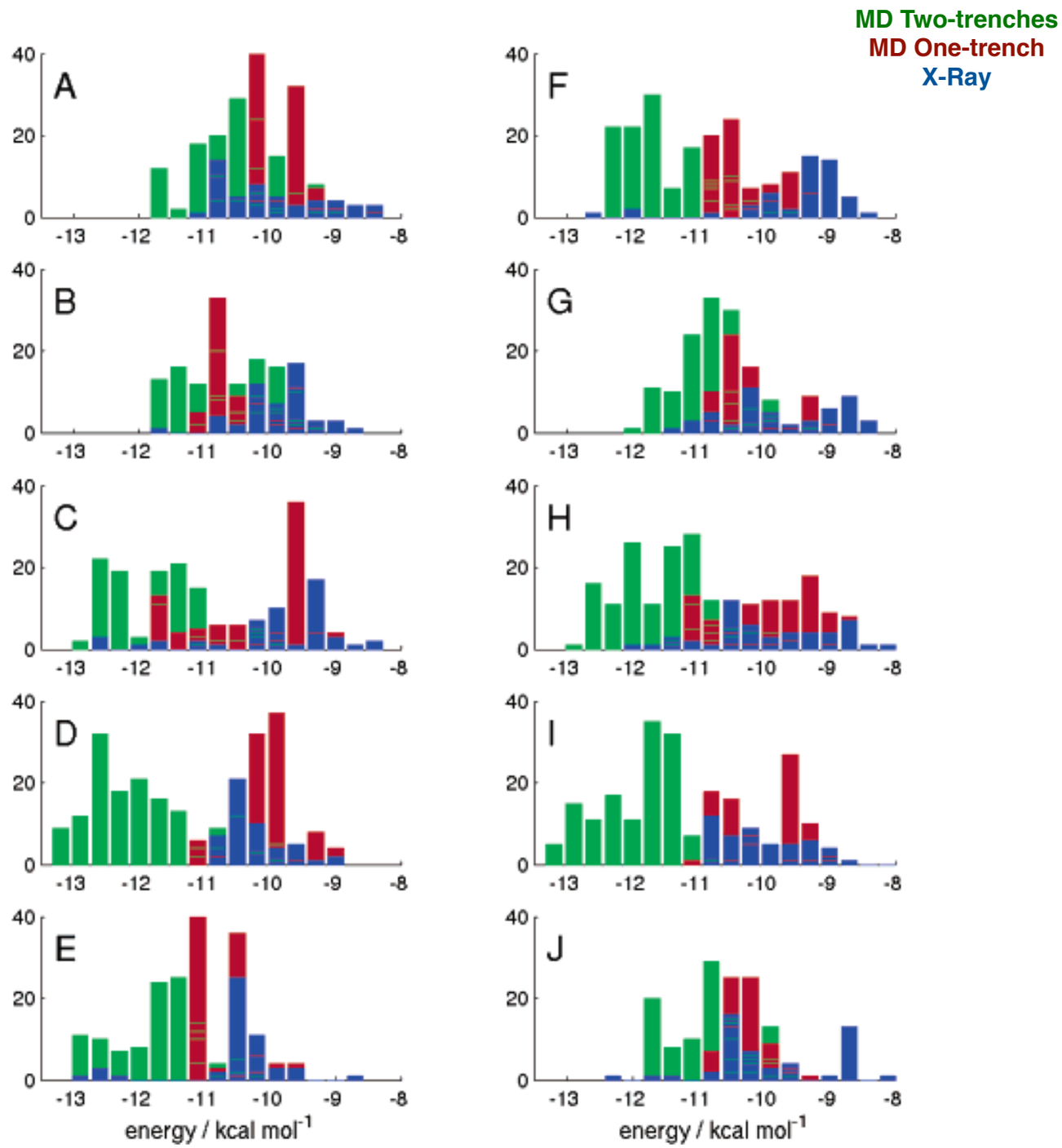
ISENTRESS example



Schames, J.R., R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, and J.A. McCammon, Discovery of a novel binding trench in HIV integrase. J Med Chem, 2004. 47(8): 1879-81

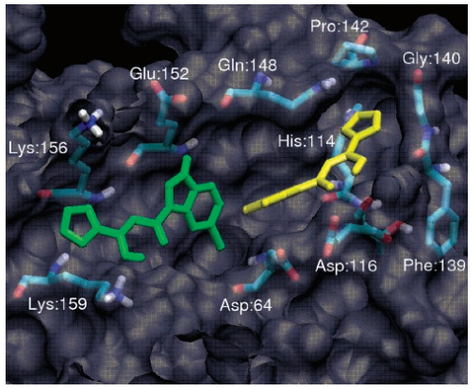
ISENTRESS example

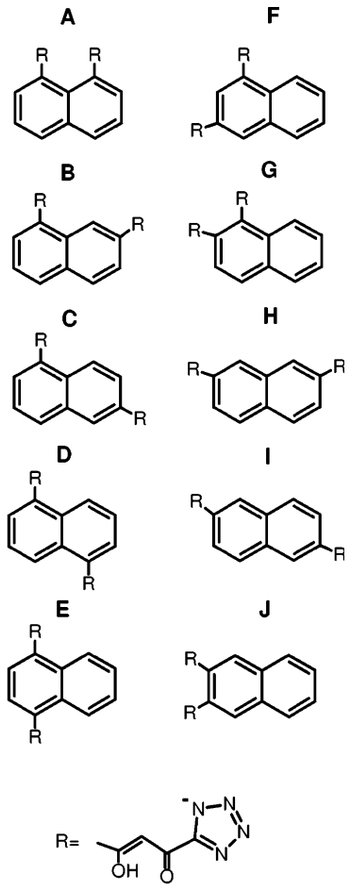


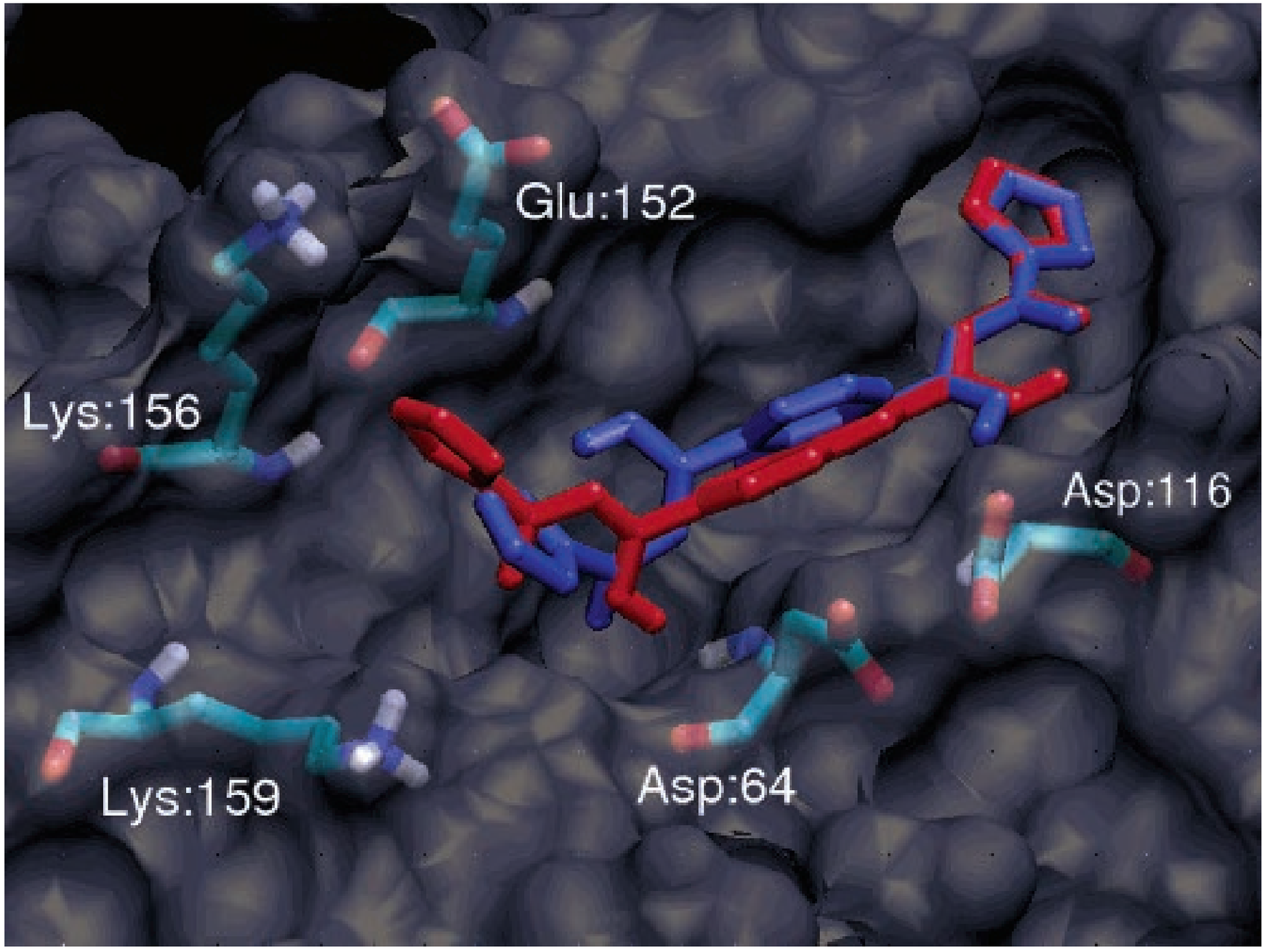


Schames, J.R., R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, and J.A. McCammon, Discovery of a novel binding trench in HIV integrase. J Med Chem, 2004. 47(8): 1879-81

ISENTRESS example

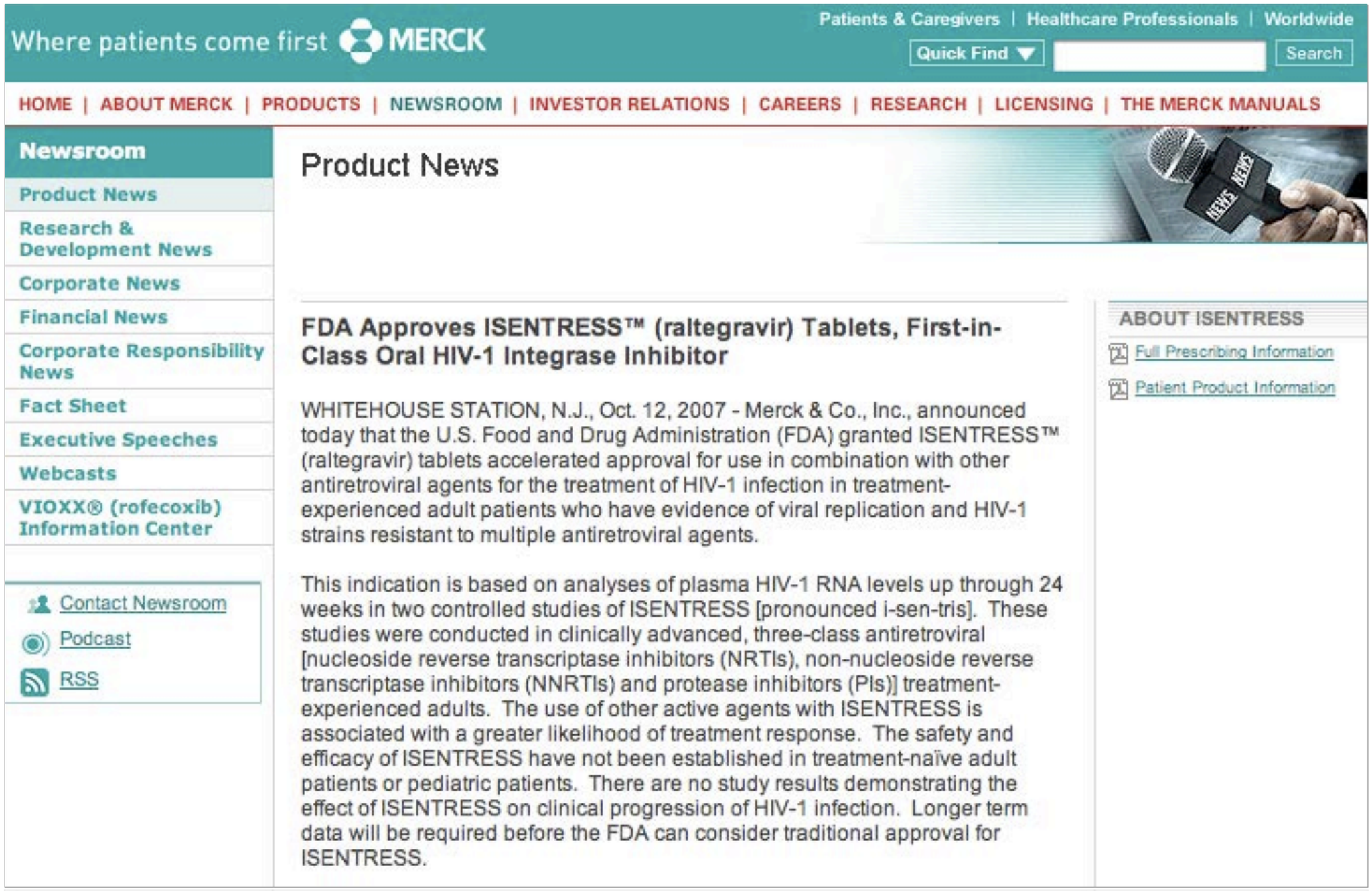






Schames, J.R., R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, and J.A. McCammon, Discovery of a novel binding trench in HIV integrase. J Med Chem, 2004. 47(8): 1879-81

ISENTRESS example

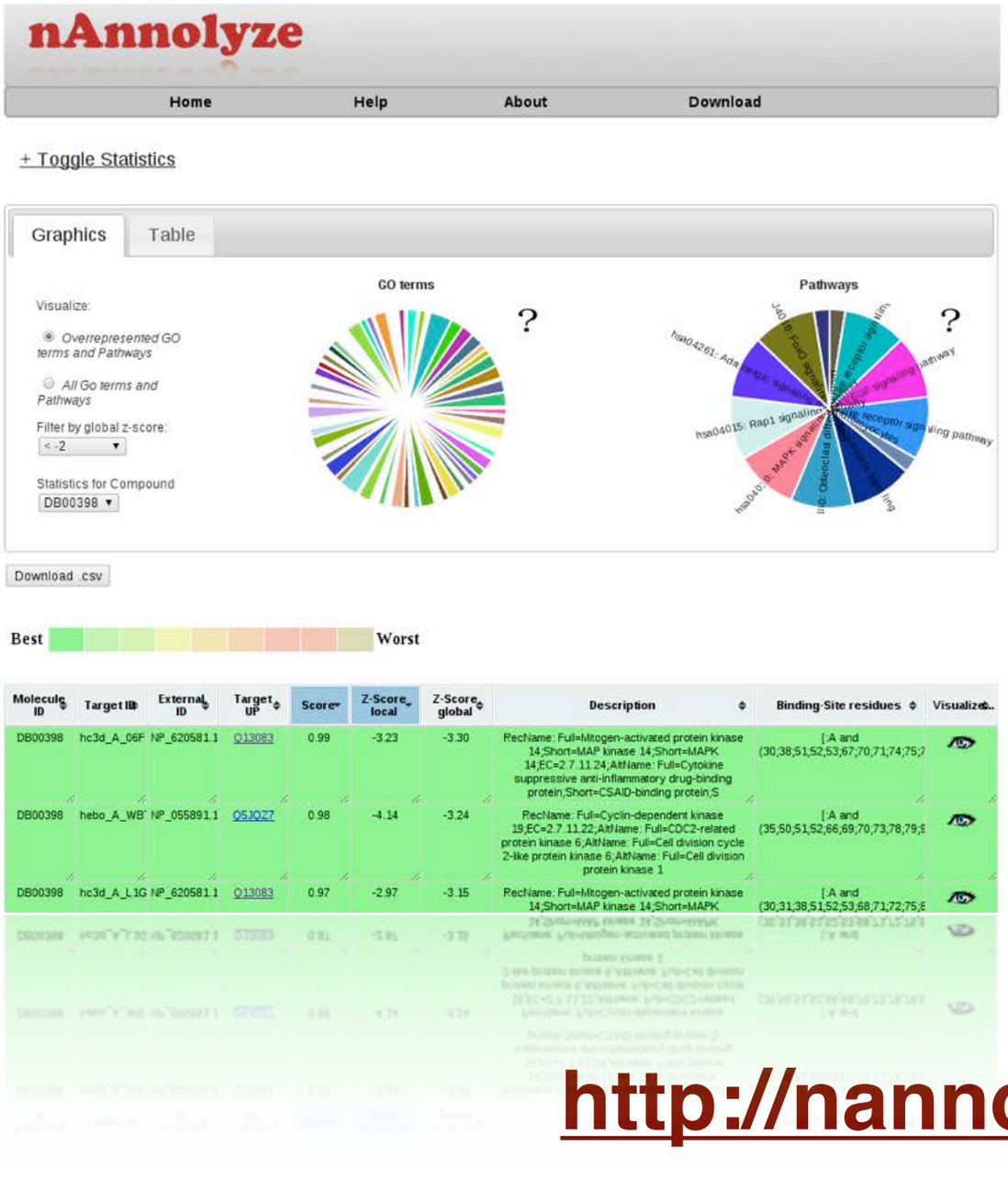


AutoDock / Vina

Practical considerations

- What problem does AutoDock solve?
  - *Flexible* ligands (4.0 *flexible* protein).
- What range of problems is feasible?
  - Depends on the search method:
    - LGA > GA >> SA >> LS
    - SA : can output trajectories, D < about 8 torsions.
    - LGA : D < about 8-32 torsions.
- When is AutoDock not suitable?
  - No 3D-structures are available;
  - Modelled structure of poor quality;
  - Too many (32 torsions, 2048 atoms, 22 atom types);
  - Target protein too flexible.

nAnnolyze binding-site + drug interaction prediction



http://nannolyze.cnag.cat

nAnnolyze in open source drug discovery

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Target Prediction for an Open Access Set of Compounds Active against *Mycobacterium tuberculosis*

Francisco Martínez-Jiménez[1,2], George Papadatos[3], Lun Yang[4], Iain M. Wallace[3], Vinod Kumar[4], Ursula Pieper[5], Andrej Sali[5], James R. Brown[4]*, John P. Overington[3]*, Marc A. Marti-Renom[1,2]*

1 Genome Biology Group, Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Spain, **2** Gene Regulation Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona, Spain, **3** European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, **4** Computational Biology, Quantitative Sciences, GlaxoSmithKline, Collegeville, Pennsylvania, United States of America, **5** Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, United States of America

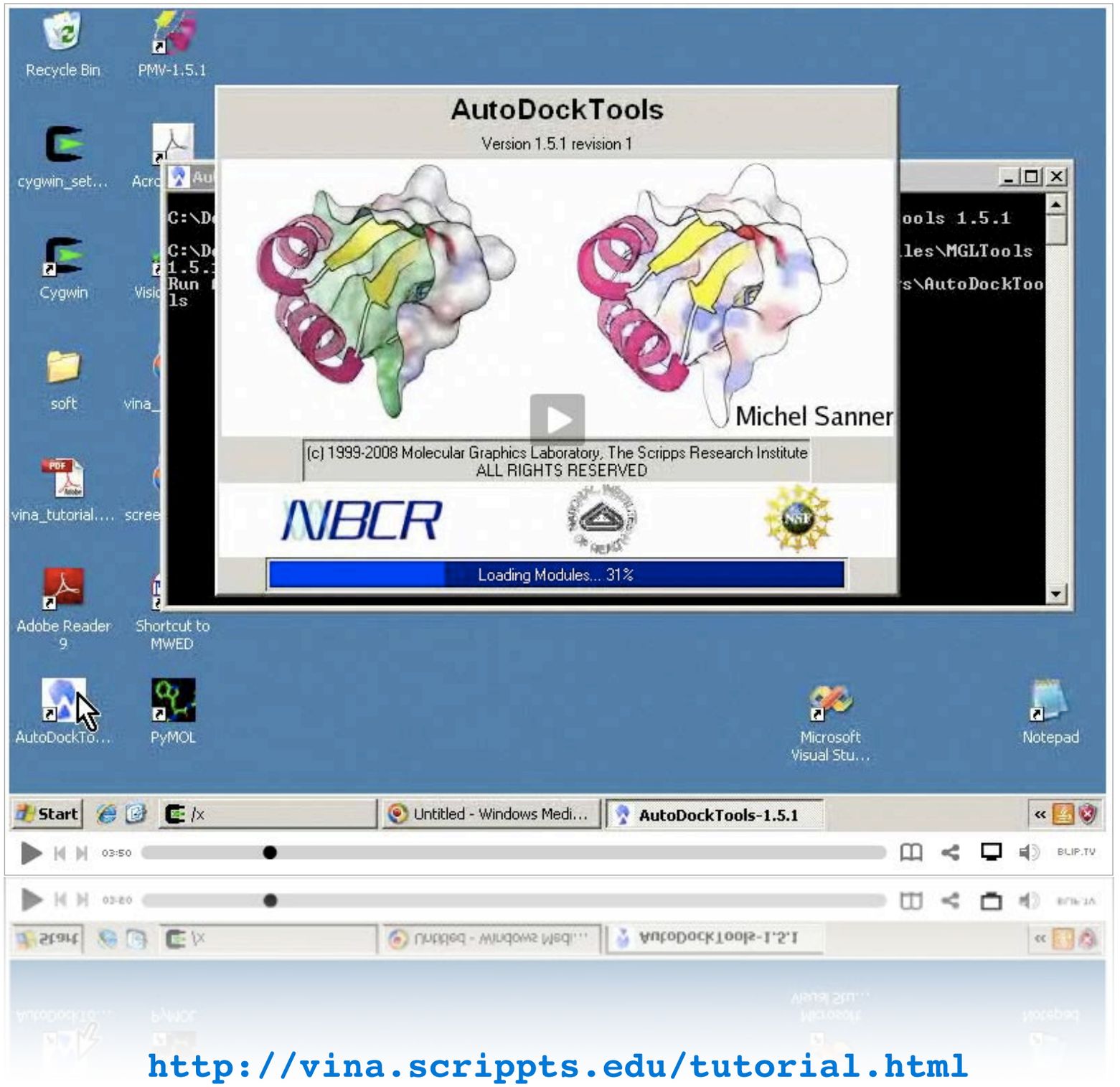
Hands on !

- Monday 4th May : AutoDock Vina tutorial.
- We will use (in LINUX!) :
  - AutoDock Tools ADT . http://mgltools.scripps.edu/downloads
  - AutoDock Vina. http://vina.scripps.edu/download.html
  - Pymol . http://www.pymol.org/

fmartinez@pcb.ub.es

Vina

There is a nice tutorial, let´s try it :)



http://vina.scrippts.edu/tutorial.html

Acknowledgements

This presentation was based on:

“Using AutoDock 4 with ADT. A tutorial”
by Dr. Ruth Huey and Dr. Garret M. Morris

Vina Tutorial
by Dr. Oleg Trott