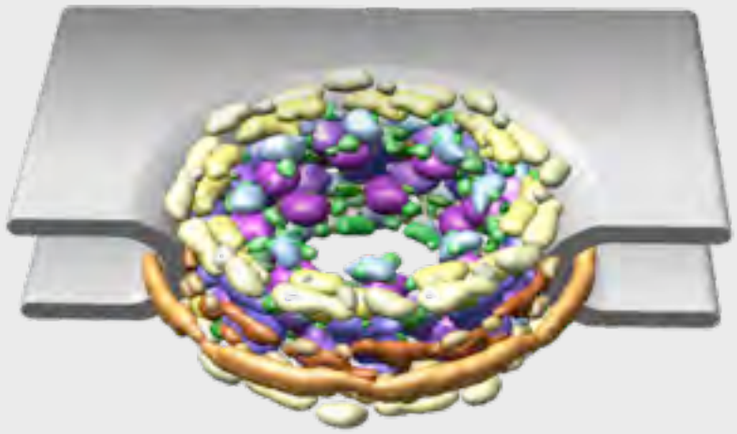
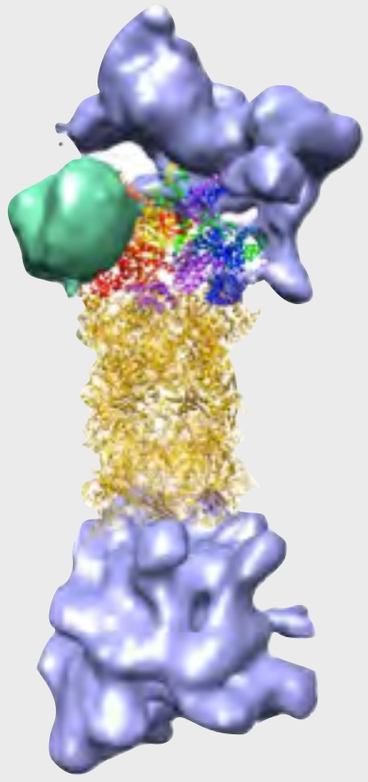
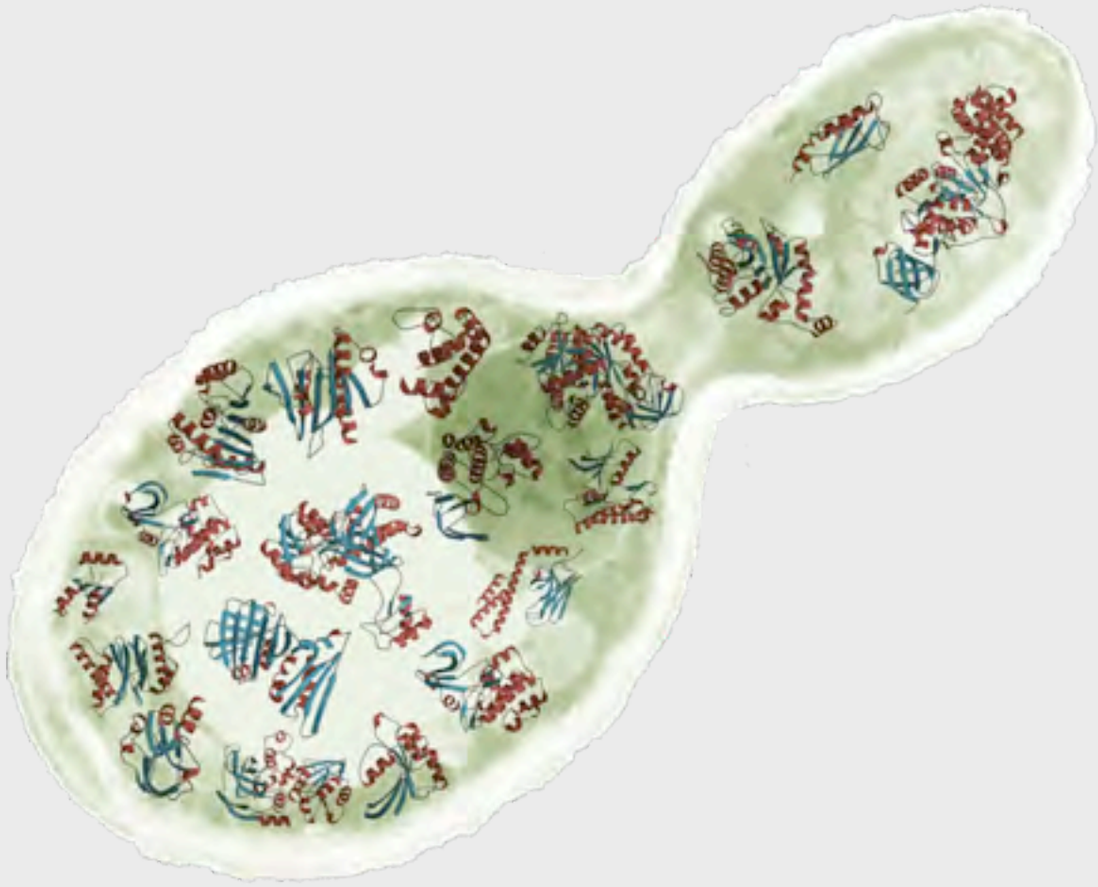
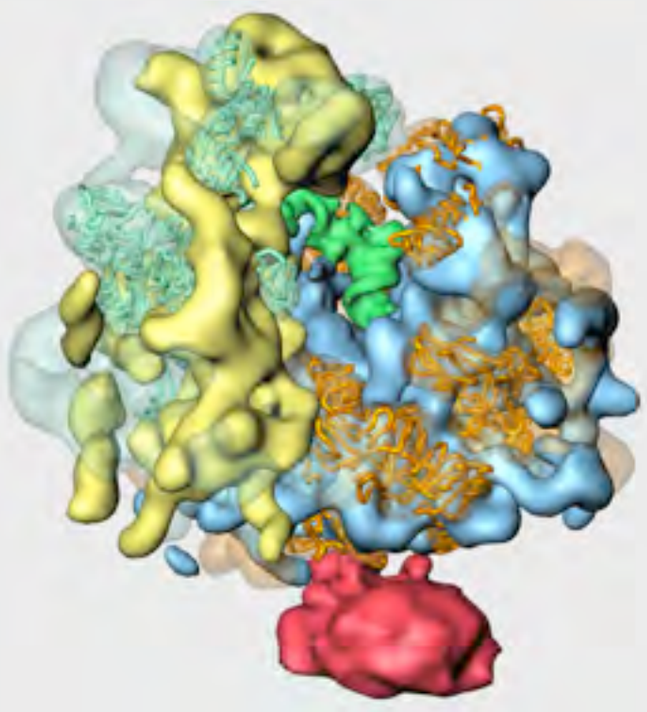
Virtual screening against comparative protein models (bioinformatician's toying with small molecules)









Andrej Sali

http://salilab.org/

UCSF

qb3 ucb ucsc ucsf

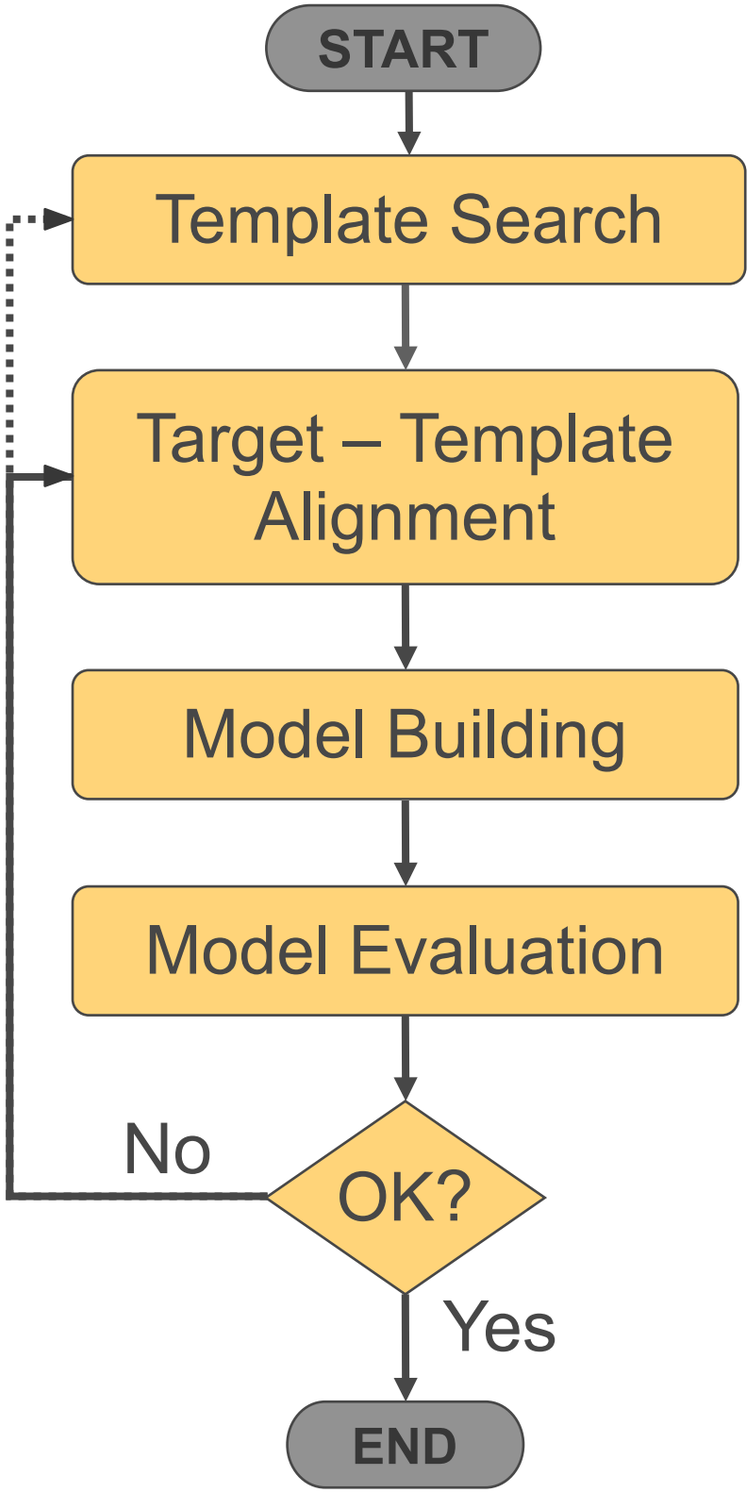
Department of Bioengineering and Therapeutic Sciences

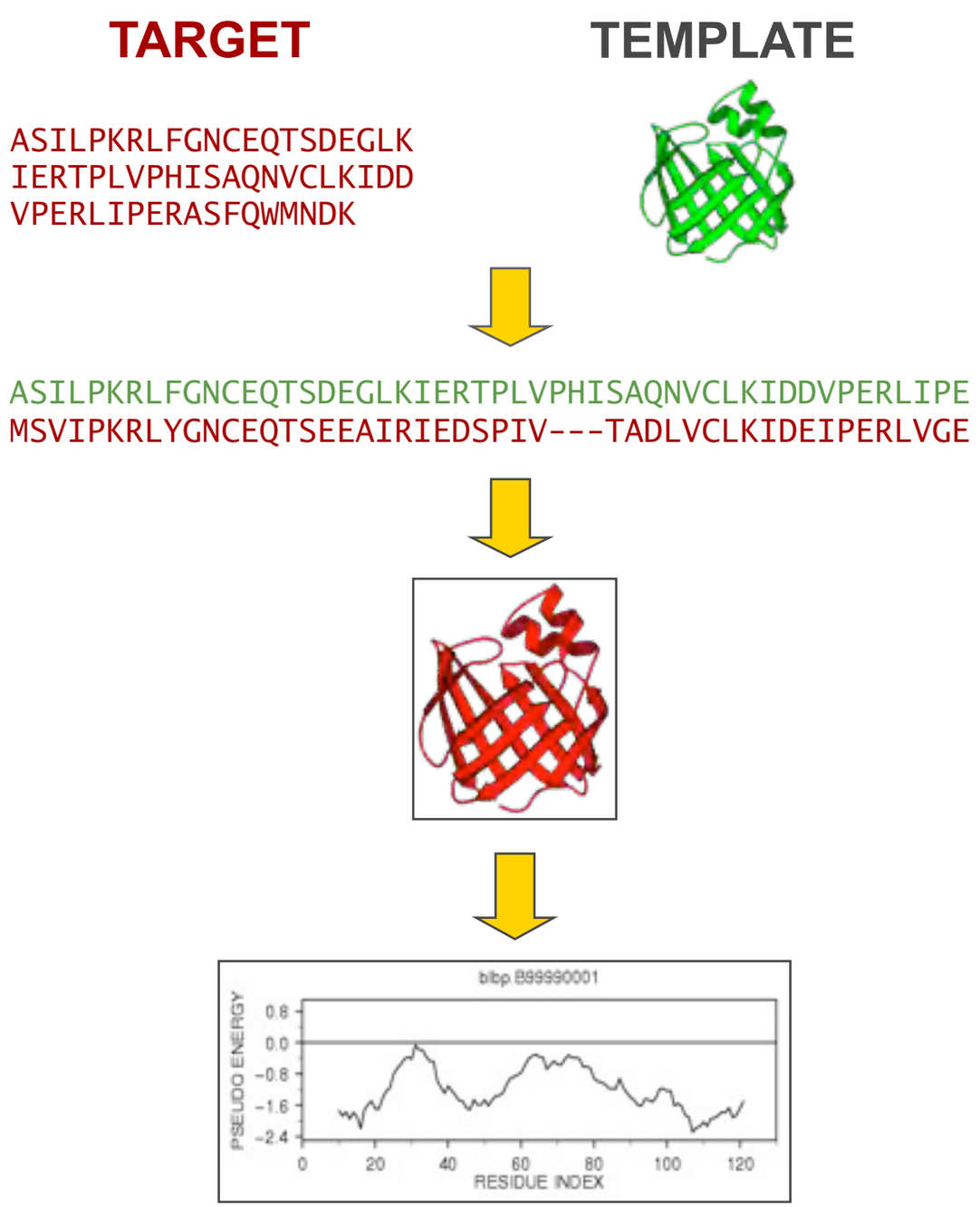
Department of Pharmaceutical Chemistry

California Institute for Quantitative Biosciences

University of California, San Francisco

Steps in Comparative Protein Structure Modeling





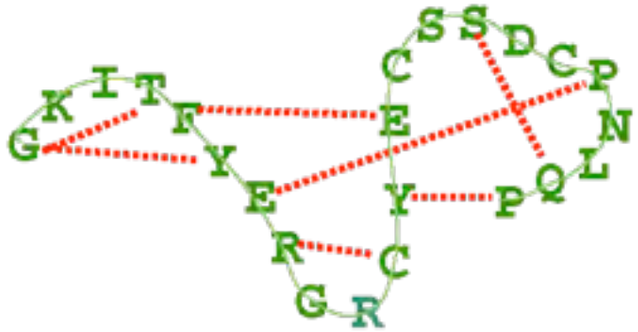
M. Marti-Renom *et al*. *Ann. Rev. Biophys. Biomolec. Struct*. **29**, 291, 2000.
http://salilab.org/

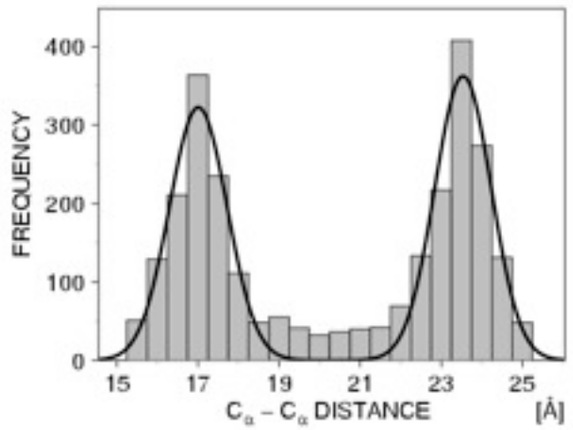
Comparative modeling by satisfaction of spatial restraints
MODELLER

3D GKITFYERGFQGHCYESDC-NLQP…

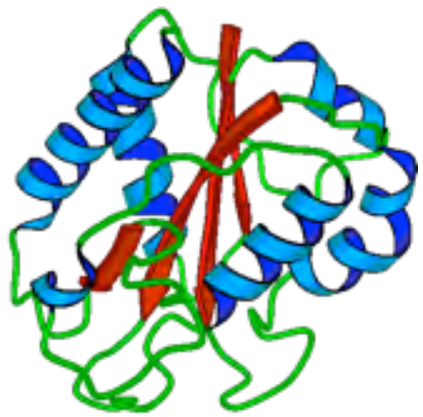
SEQ GKITFYERG---RCYESDCPNLQP…

1. Extract spatial restraints





2. Satisfy spatial restraints



$$P(\mathbf{R} / \mathbf{I}) = \prod_i p_i(\mathbf{r}_i / \mathbf{I}_i)$$

A. Šali & T. Blundell. *J. Mol. Biol.* **234**, 779, 1993.
J.P. Overington & A. Šali. *Prot. Sci*. **3**, 1582, 1994.
A. Fiser, R. Do & A. Šali, *Prot. Sci*., **9**, 1753, 2000.

http://salilab.org/

05/27/2006

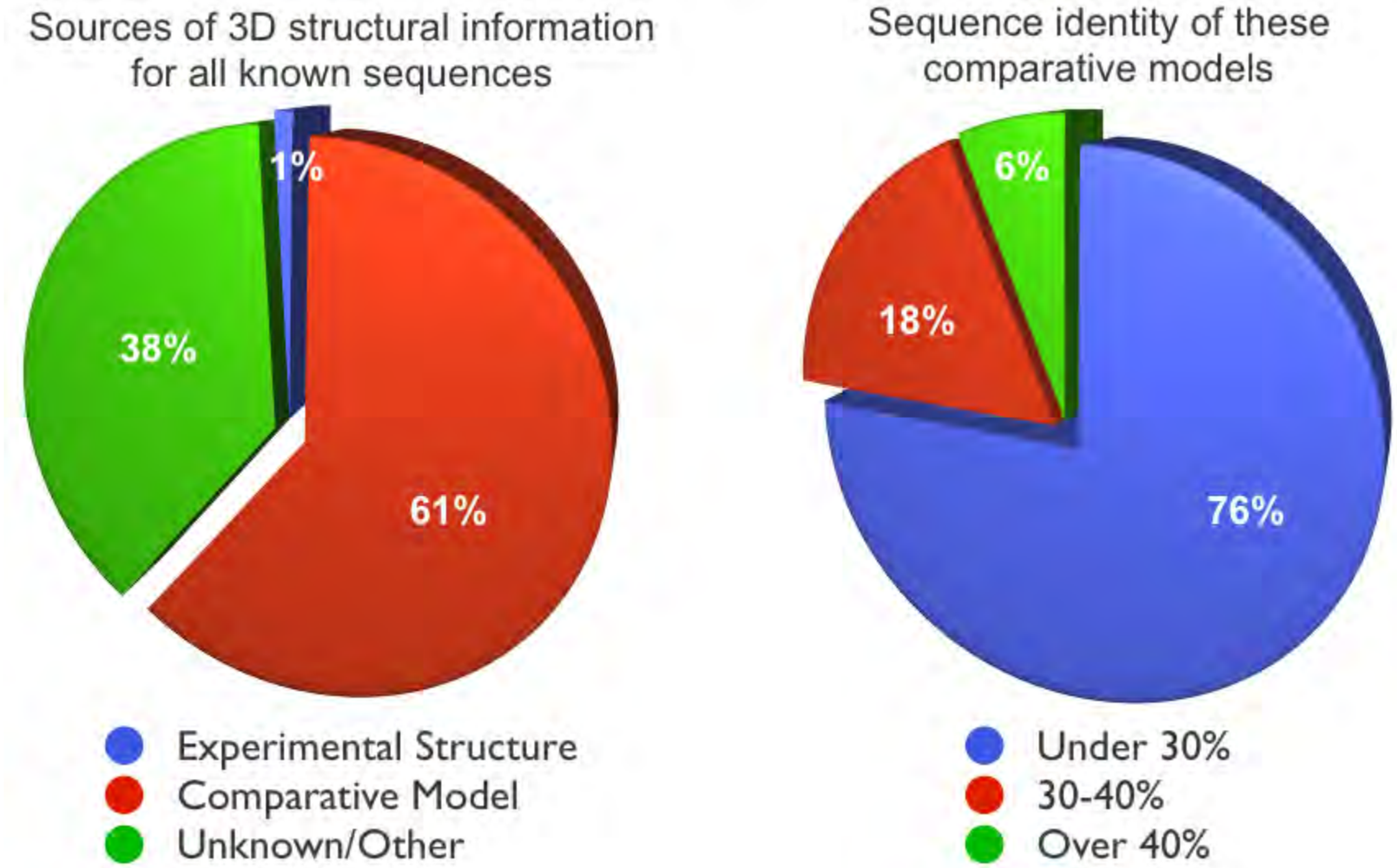
Comparative modeling of the UniProt database

Unique sequences processed: 2,130,404

Sequences with fold assignments or models: 1,273,766 (60%)

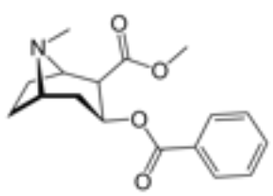
70% of models based on <30% sequence identity to template.

On average, only a domain per protein is modeled (an “average” protein has 2.5 domains of 175 aa).

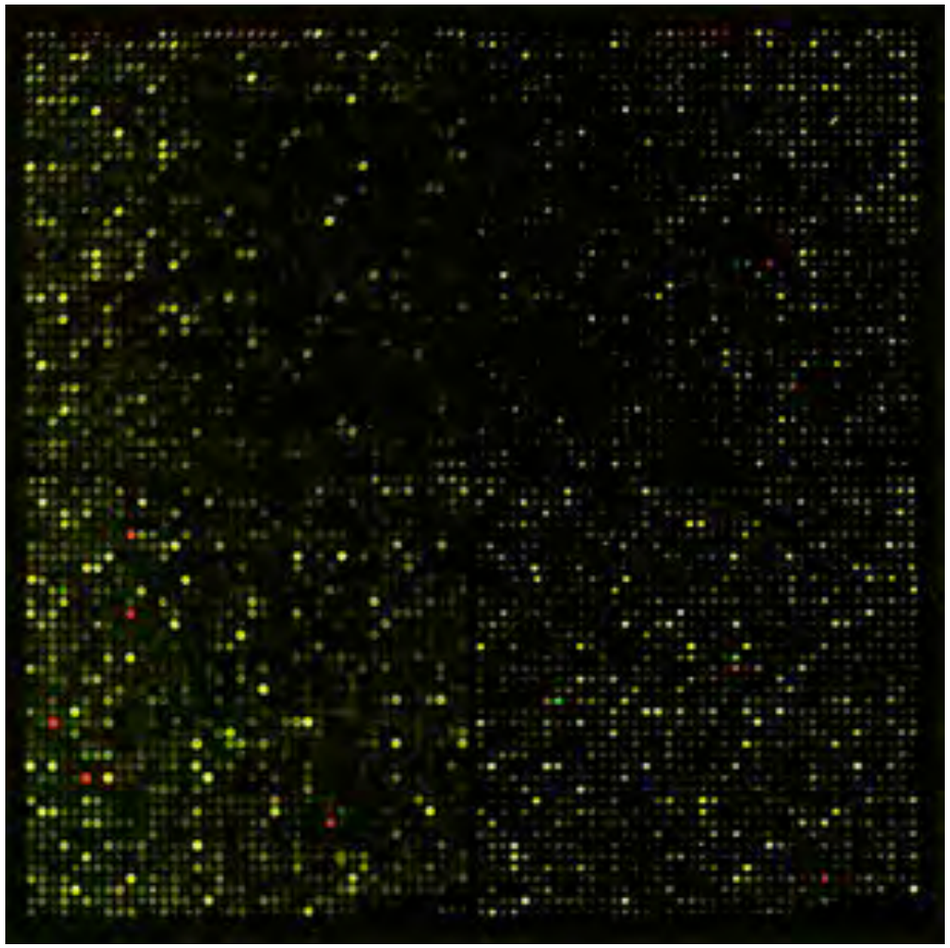


Pieper *et al. Nucleic Acids Research,* 2006, 2011.

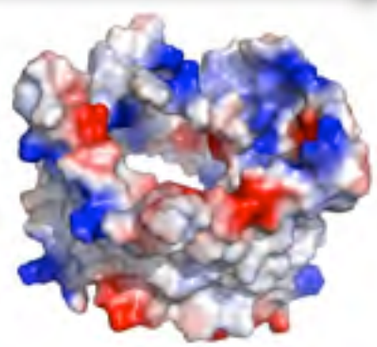
Comprehensive mapping of interactions between proteins and small ligands



all small molecules



all binding sites on all proteins



Genome-Wide Mapping of Protein-Ligand Interactions

All known protein structures
All known protein sequences

Comparative modeling

Refine protein models

Identify ligand
binding sites on models

Lists of small ligands

Virtual ligand libraries

Annotated
protein structure models

Build ligand-protein
complexes

Rescore ligand-protein
complexes

Central data warehouse

Graphical User Interface

Testing and Applications

Testing

Computational applications

Computer hardware and software environment

Software backplane

Global optimization

Information navigation and
search

Cluster hardware and
software

Contents

1. Vignettes:
   - Specificity of Brain Lipid-Binding Protein (BLBP)
   - Identifying binding sites on proteins
   - Comparative “docking” of small molecules to proteins
   - Overlap between binding sites for proteins and small molecules
2. Enzyme Function Initiative
3. Docking against comparative models
4. Application to norepinephrine transporter (NET)

What is the physiological ligand of Brain Lipid-Binding Protein?

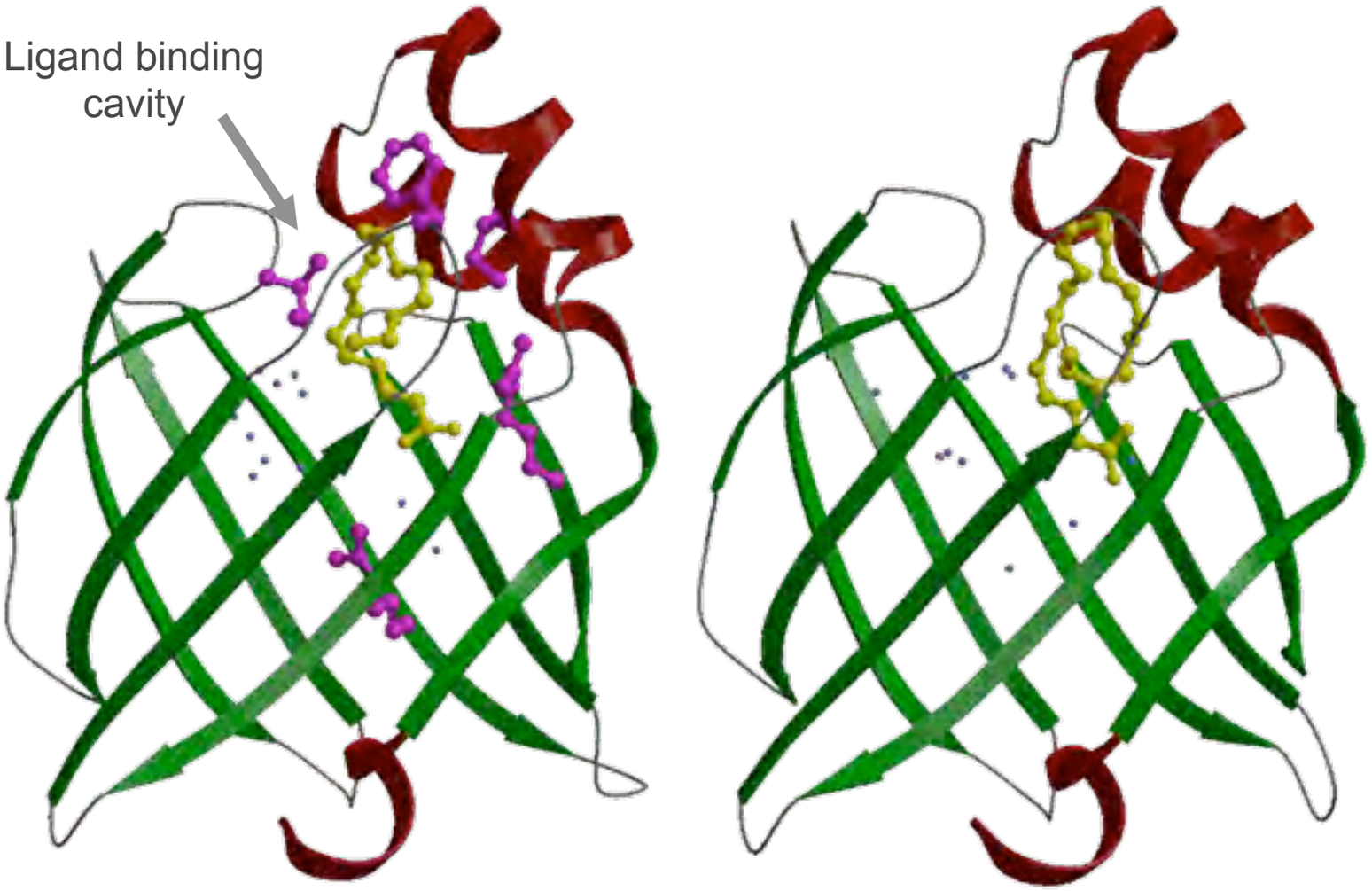
Predicting features of a model that are not present in the template

BLBP/oleic acid

Cavity is **not** filled

BLBP/docosahexaenoic acid

Cavity **is** filled



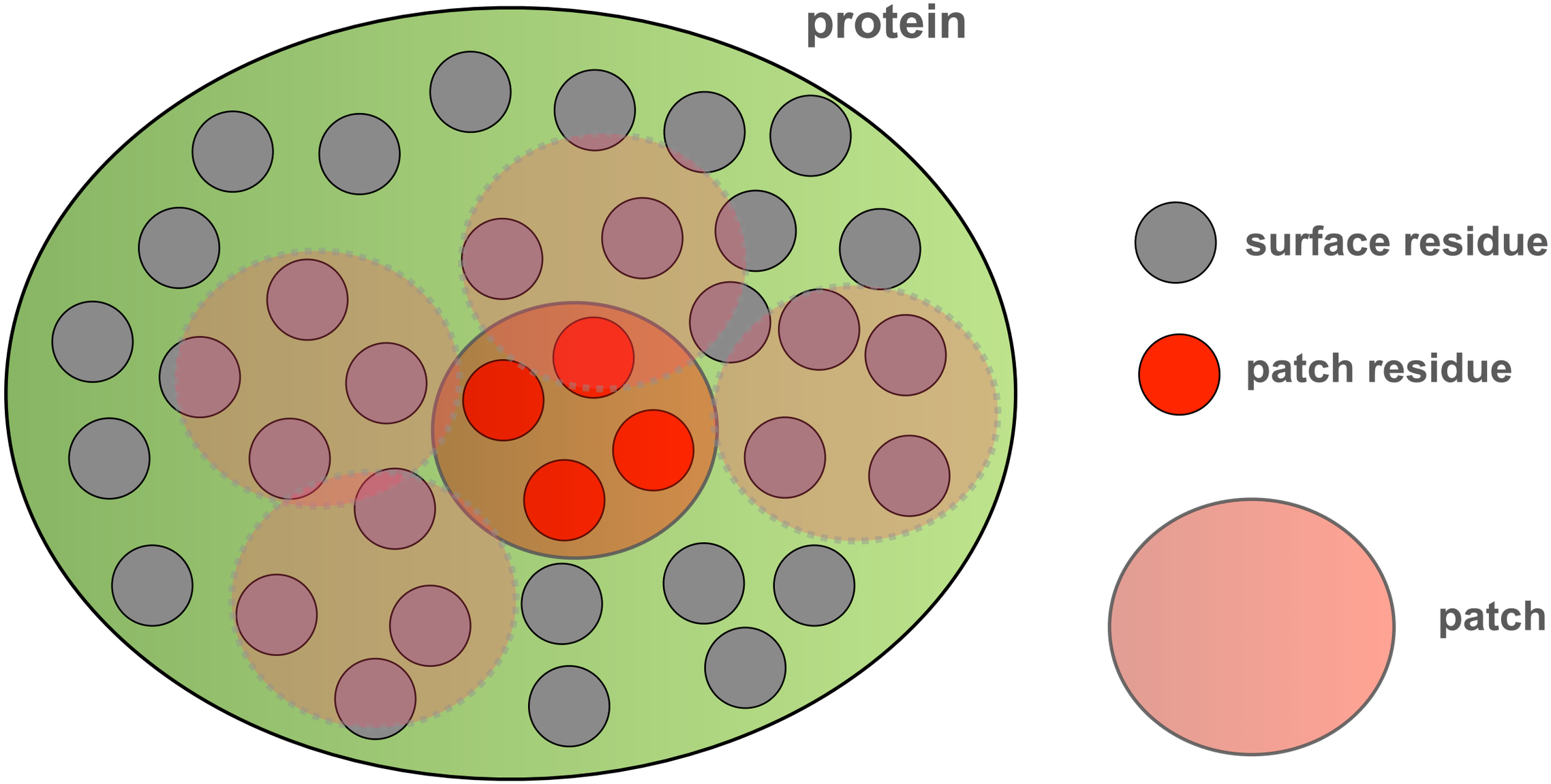
1. BLBP binds fatty acids.
2. Build a 3D model.
3. Find the fatty acid that fits most snuggly into the ligand binding cavity.

L. Xu, R. Sánchez, A. Šali, N. Heintz, *J. Biol. Chem.* **271**, 24711, 1996.

Contents

1. Vignettes:
   - Specificity of Brain Lipid-Binding Protein (BLBP)
   - Identifying binding sites on proteins
   - Comparative “docking” of small molecules to proteins
   - Overlap between binding sites for proteins and small molecules
2. Enzyme Function Initiative
3. Docking against comparative models
4. Application to norepinephrine transporter (NET)

Localization of a binding site of a given type by **optimizing** a scoring function that depends on properties of a surface residue patch



Rossi, Marti-Renom, Sali, *Prot Sci*, 2006.

Methods: Scoring Function

Our current scoring function assesses any patch based on these properties (requires examples of the binding site):

- Conservation (BLAST generated profiles)
- Compactness (average residue distance)
- Protrusion (nearest neighbor list)
- Convexity (exposure vectors)
- Rigidity (B-factor from crystallographic coordinates)
- Hydrophobicity (hydrophobicity scale)
- Charge density (CHARMM)
- Number of residues

Properties are transformed into Z-scores (scored patch *versus* random patches):

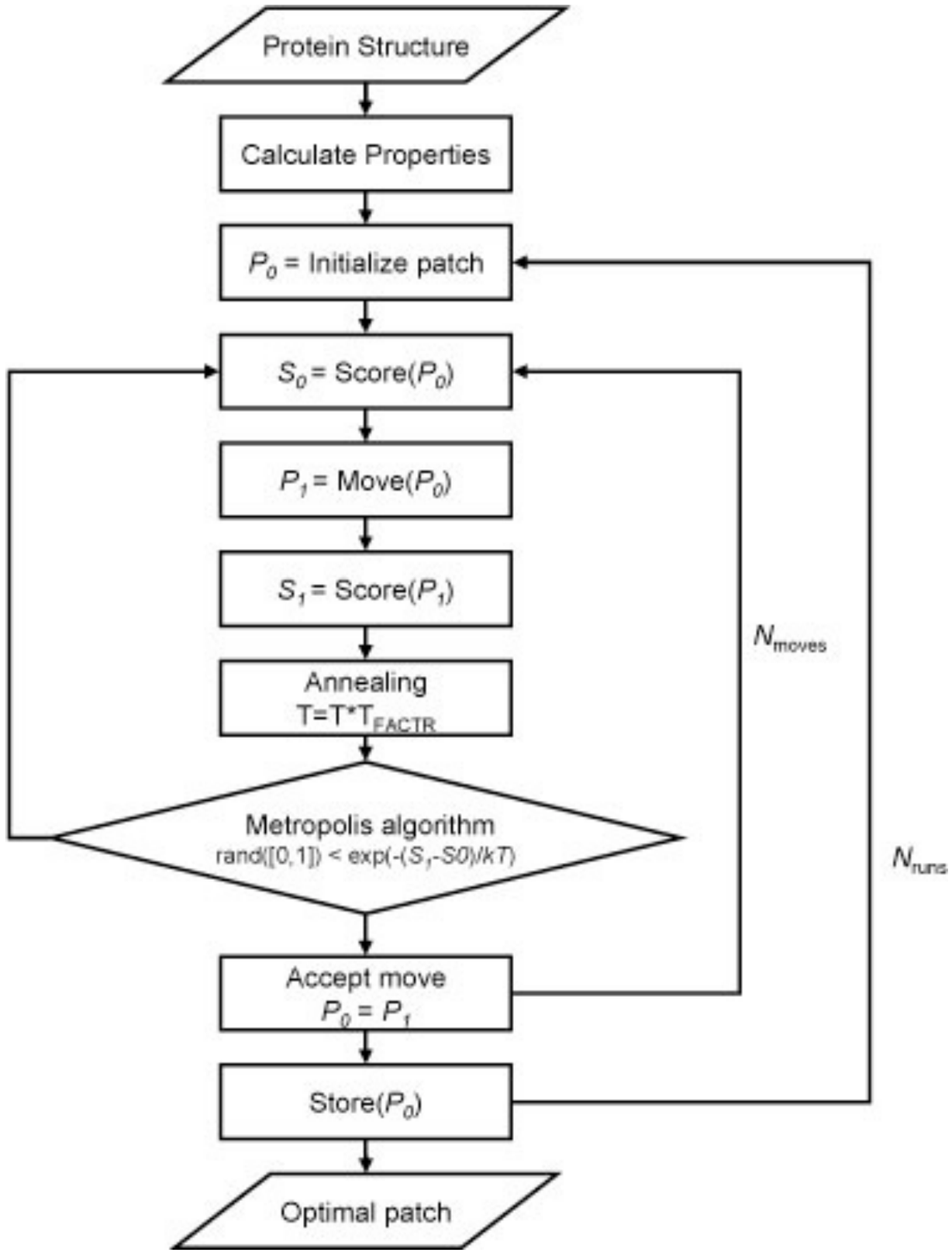
$$Z_k = \left(f_k - \bar{f}_k \right) / \sigma_k$$

The scoring function is a linear combination of property Z-scores:

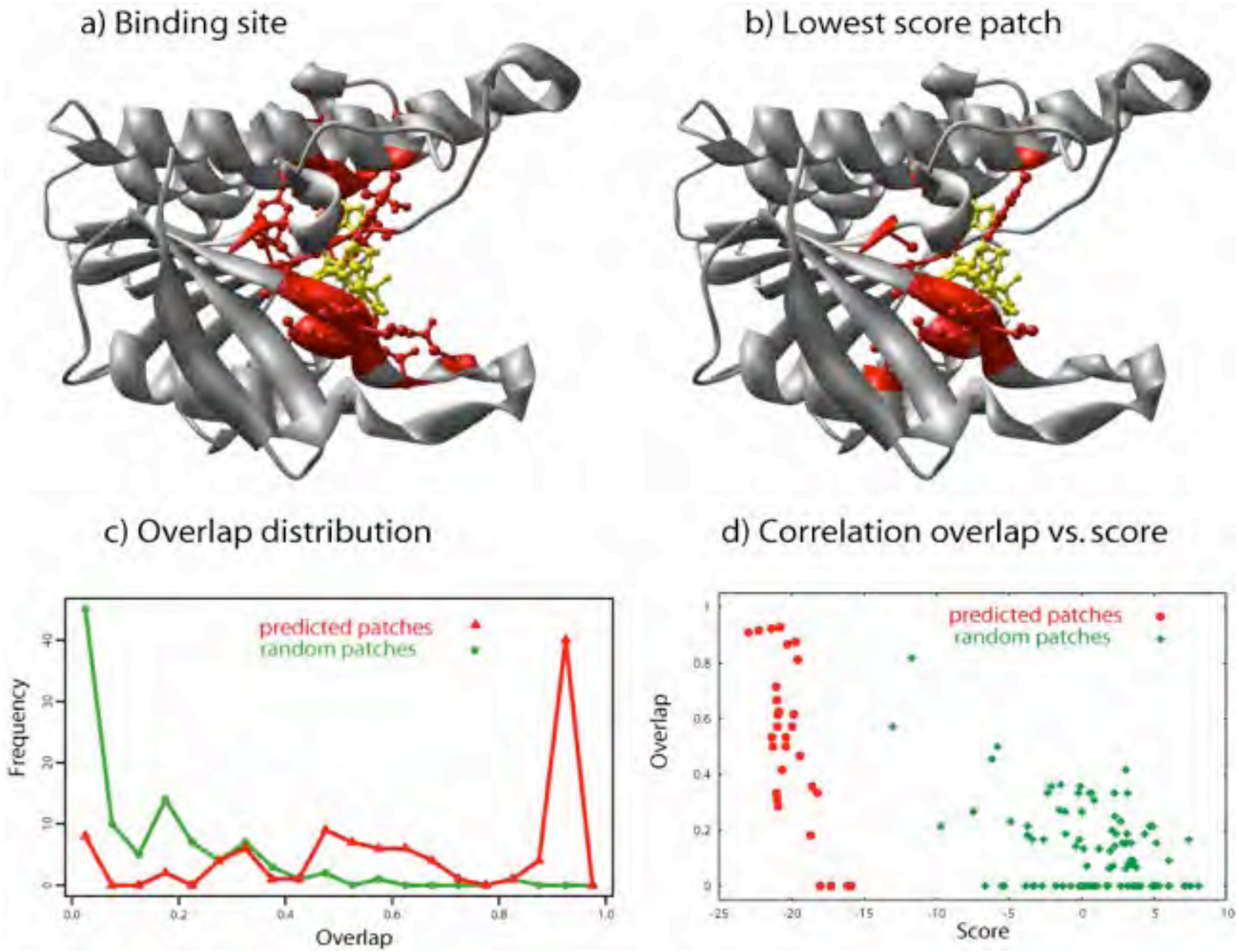
$$F(P;\{w_k\}) = \sum_{k=1}^{N_p} w_k Z_k(P)$$

Methods: Patch optimization

Monte Carlo Simulated Annealing



Example: NAD binding site localization on dihydropteridine reductase (1dhr)



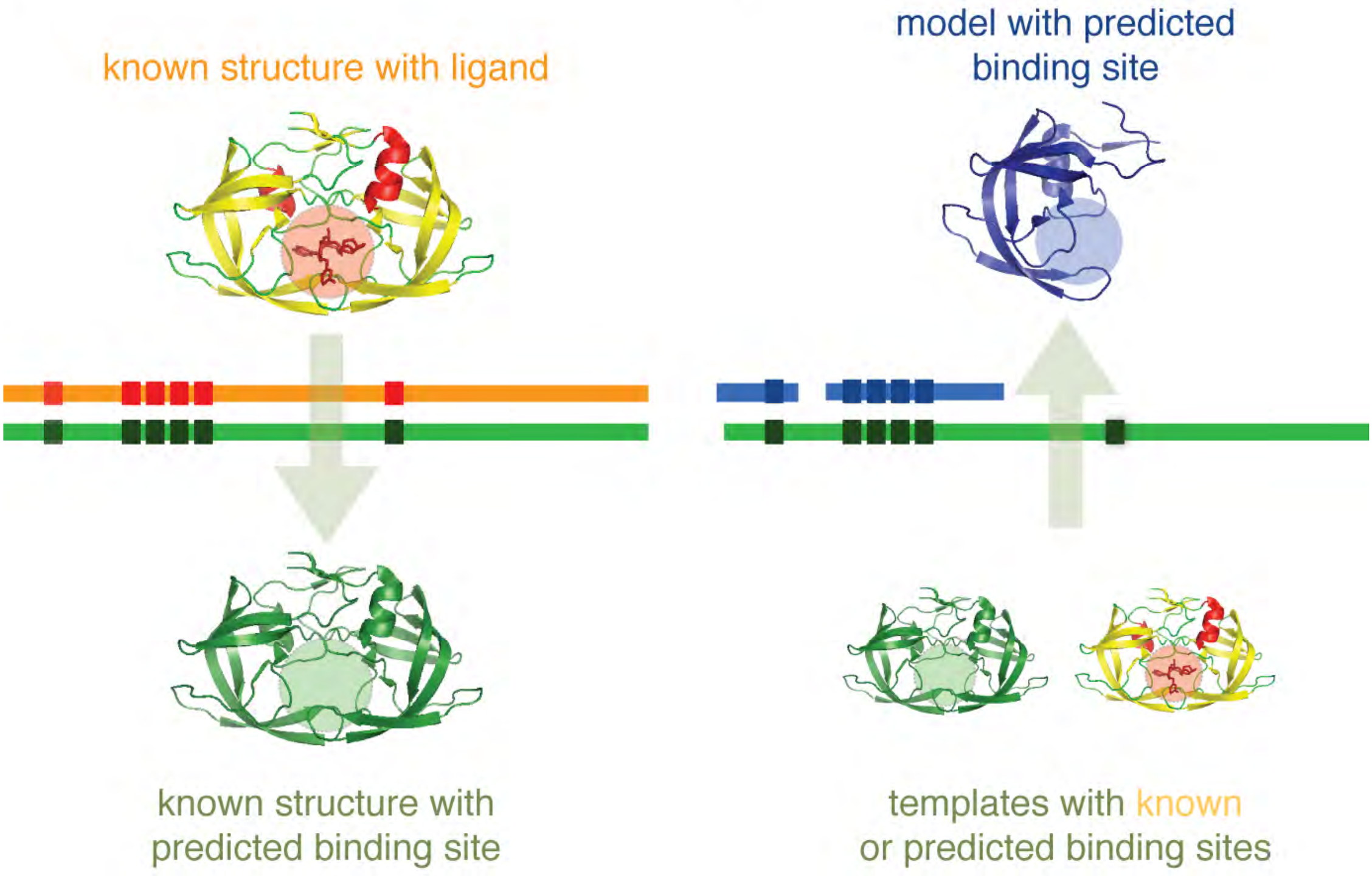
Large benchmark

For nonsugar ligands, such as various nucleotides, 20 different types of binding sites in 1008 structures were correctly identified in 55%–73% of the cases.

Contents

1. Vignettes:
   - Specificity of Brain Lipid-Binding Protein (BLBP)
   - Identifying binding sites on proteins
   - Comparative “docking” of small molecules to proteins
   - Overlap between binding sites for proteins and small molecules
2. Enzyme Function Initiative
3. Docking against comparative models
4. Application to norepinephrine transporter (NET)

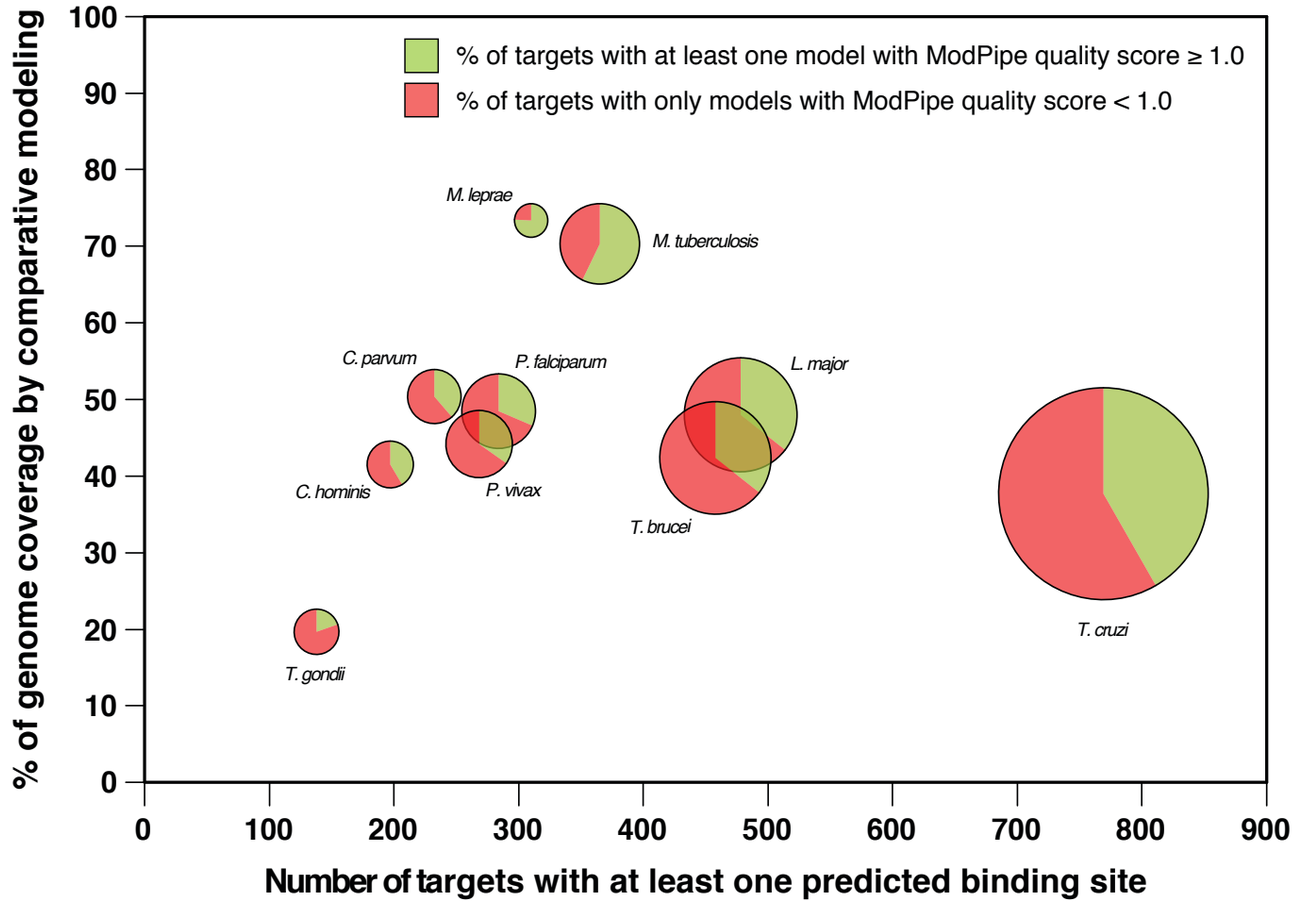
Prediction of a binding site and ligand by homology



Marti-Renom et al, Nucl Acids Res, 2007
Marti-Renom et al, BMC Bioinformatics 2006

Many others have explored these relationships (Thornton, Sternberg, Rost, ...)

A kernel for open source drug discovery in tropical diseases



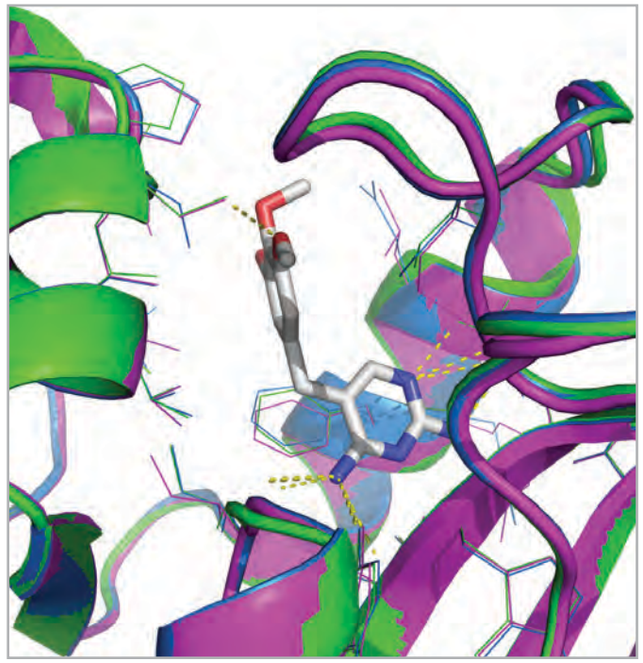
TrEMBL
PDB
ModPipe
ModBase
LigBase
DBAli
MSDChem
DrugBank

At least one binding site for a small molecule was predicted for 3499 proteins in 10 pathogen genomes, based on similarity to known binding sites. Relating ligands in the PDB to compounds in MSDChem and DrugBank predicts that 297 of these proteins bind a molecule similar to a known drug (143 are predicted to bind a known drug).

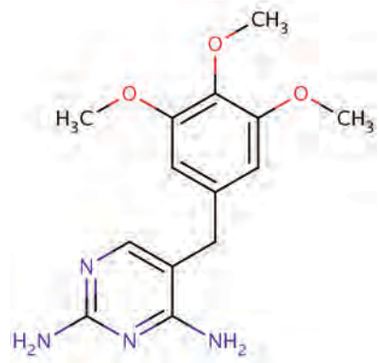
L. Orti, R.J. Carbajo, U. Pieper, N. Eswar, S. M. Maurer, A. K. Rai, G. Taylor, M. H. Todd, A. Pineda-Lucena, A. Sali, M. A. Marti-Renom, PLoS Negl Trop Dis, 2009

Examples of recovered known pathogen drugs

A) *M. leprae* dihydrofolate reductase



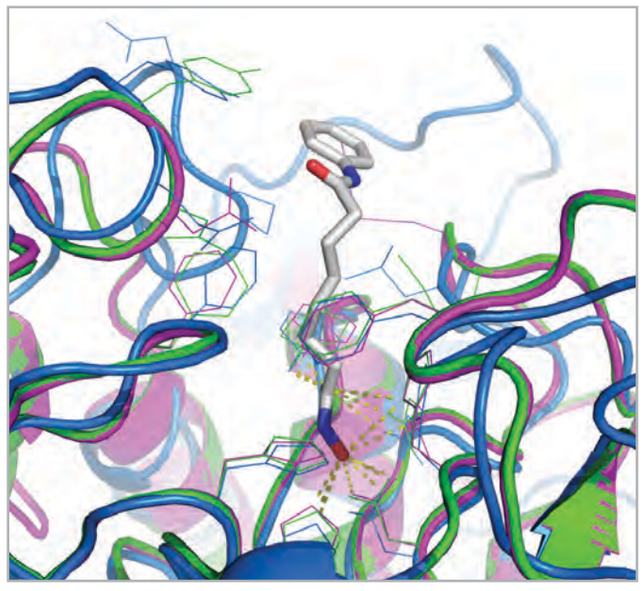
Trimethoprim
Small Molecule; Approved



Drug categories:
Antimalarials
Anti-Infectives

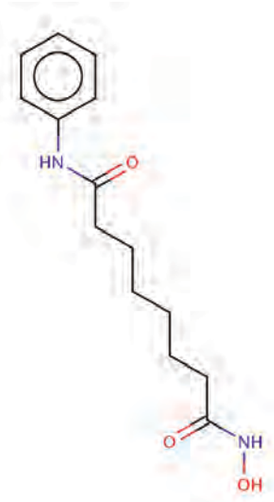
Drug indication:
For the treatment of initial episodes of uncomplicated urinary tract infections.

B)



L. major histone deacetylase

Vorinostat
Small Molecule; Approved; Investigational



Drug categories:
Anti-Inflammatory Agents, Non-Steroidal
Anticarcinogenic Agents
Antineoplastic Agents
Enzyme Inhibitors

Drug indication:
For the treatment of cutaneous manifestations in patients with cutaneous T-cell lymphoma who have progressive, persistent or recurrent disease on or following two systemic therapies.

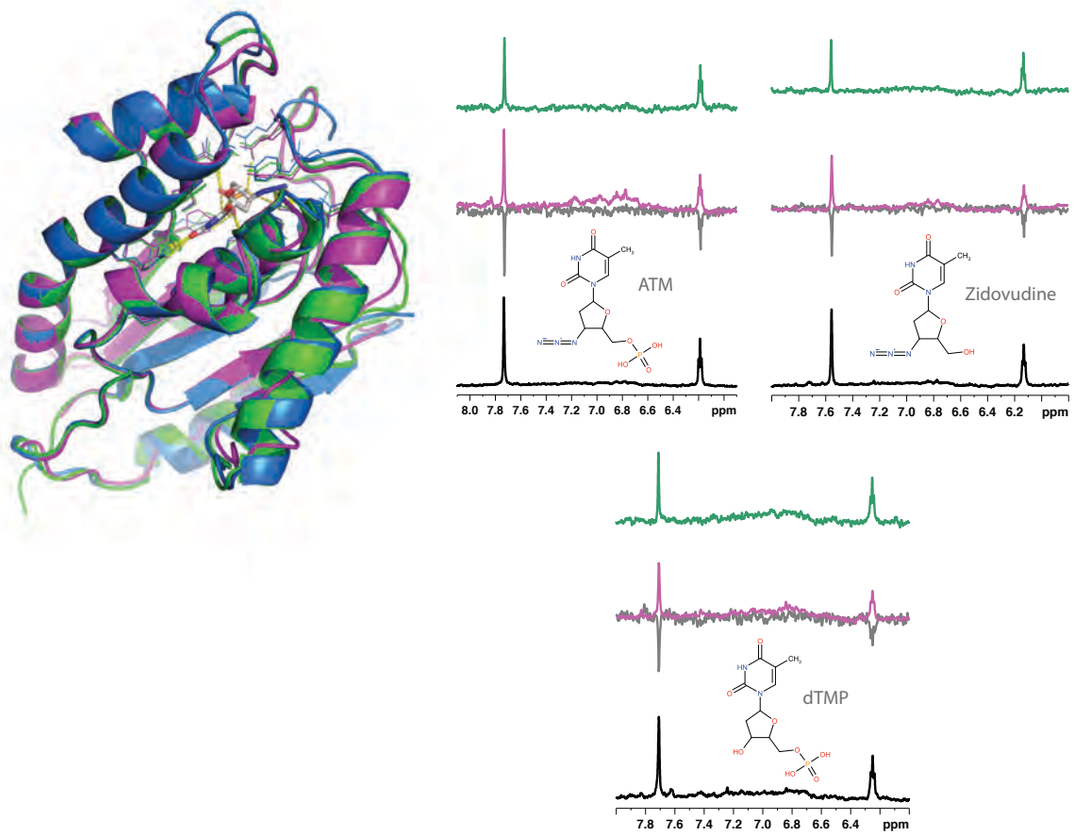
The original PDB structure with the ligand bound is shown in blue; the transferred binding site in the template structure is shown in green; and a comparative protein structure model of the target sequence is shown in magenta.

Orti et al, PLoS Negl Trop Dis, 2009

Testing of predicted pathogen protein - drug interactions

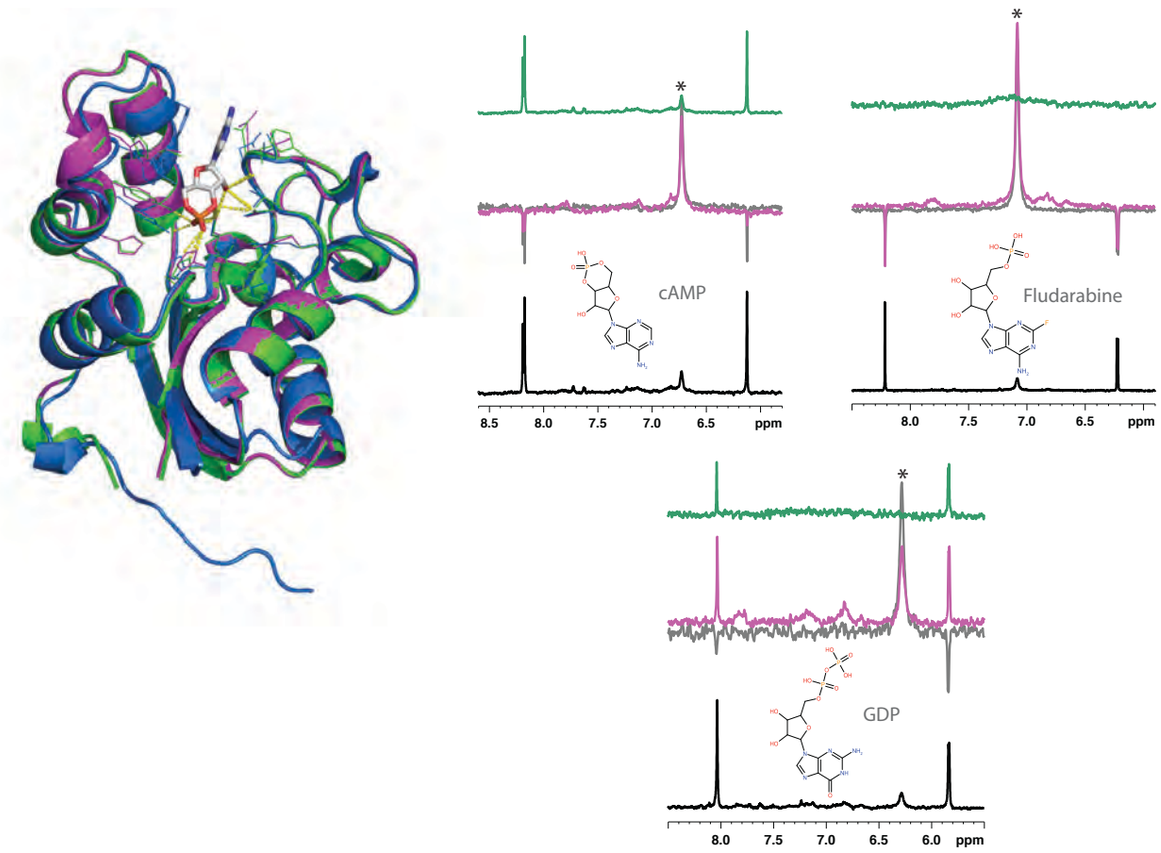
Water-LOGSY NMR spectroscopy

A)



Validated: *P. falciparum* thymidylate kinase interactions with dTMP, ATM and Zidovudine.

B)



Invalidated*: M. leprae* nucleoside diphosphate kinase interactions with GDP, cAMP and Fludarabine.

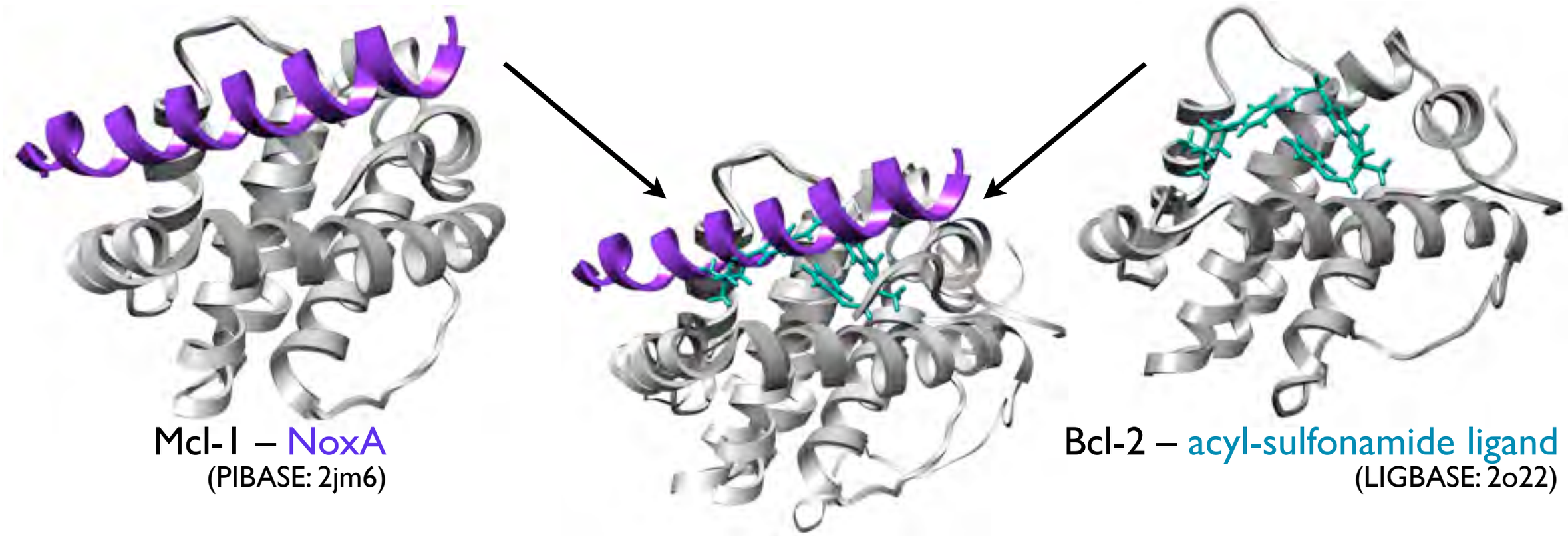
Each NMR spectrum shows a detail of the aromatic region for the interacting molecules, the bottom spectra corresponding to the reference 1D ^{1}H experiment (black line). In this experimental setting, a non-interacting compound results in negative resonances in the Water-LOGSY experiment and no signals in the STD spectrum. In contrast, protein-ligand interactions in the Water-LOGSY (magenta line) are characterized by positive signals or by a reduction in the negative signals obtained in the absence of the protein (reference spectrum, grey line). In the STD experiment, a positive interaction is recognized by the presence of positive signals (green line). Signals marked with an asterisk arise from exchangeable protons, and although positive, do not indicate an interaction between the protein and the ligand, as they also show the same behavior in the absence of protein.

Orti et al, PLoS Negl Trop Dis, 2009

Contents

1. Vignettes:
   - Specificity of Brain Lipid-Binding Protein (BLBP)
   - Identifying binding sites on proteins
   - Comparative “docking” of small molecules to proteins
   - Overlap between binding sites for proteins and small molecules
2. Enzyme Function Initiative
3. Docking against comparative models
4. Application to norepinephrine transporter (NET)

Do small molecule and protein binding sites overlap within families?



Mcl-1 – NoxA
(PIBASE: 2jm6)

Bcl-2 – acyl-sulfonamide ligand
(LIGBASE: 2o22)

Protein binding site

Overlapping (bi-functional)

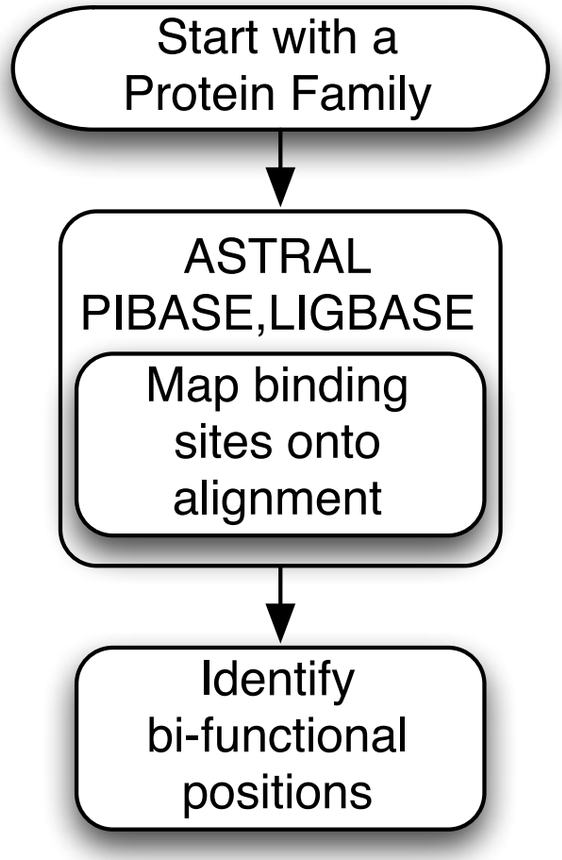
Ligand binding site

```
BDP87632-0_SCOP.d2jm6b1 g p l g s e d d l y r q s l e - - - i i s r y l r e q a t g s k d s k p l g e a g a a g r - - - - -
BDP89399-0_SCOP.d2o22a1 - - - - - - - h a g r t g y d n r e i v m k y i h - y k l s q r g y e w d a g d d v e e n r t e a p
BDP87632-0_SCOP.d2jm6b1 - - - - - - r a l e t l r r V g d g V q r N H e T A F q G M l r K L d i k n e g d v k S F s R V m v
BDP89399-0_SCOP.d2o22a1 e g t e s e v v h l T l r Q A G d D F S r R Y r r D F a e M s s q l h l t p f t a r g r f a t v v e
BDP87632-0_SCOP.d2jm6b1 H V F k D g V t N W G R i V T L i s F g a f v a k h l k s v n q e s f i e p l a e t i t d v l v r t
BDP89399-0_SCOP.d2o22a1 e L f r d g - v N W G R i V A f f e F g g v m c v e s v n r e m s p l v d n i a l w m t e y l n r h
BDP87632-0_SCOP.d2jm6b1 k r d w l v k q r g w d g f v E F F h V Q D L E g G
BDP89399-0_SCOP.d2o22a1 l h t w i q d n g g w d a F v e l Y g p s m r - - -
```

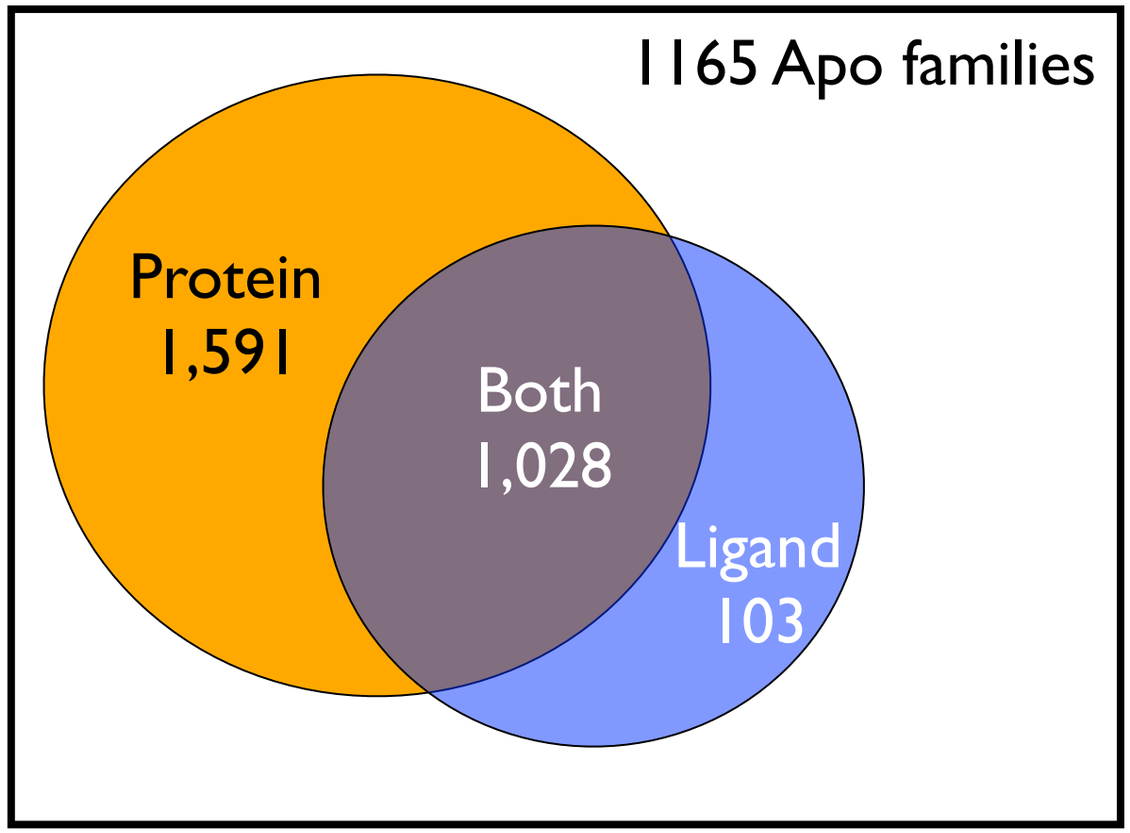
(ASTRAL alignment)

Davis and Sali, *PLoS Comp Biol* 2010.

Many families bind both small molecules and proteins



Stuart *et al.* ASTRAL, *Bioinformatics* 2002
Davis *et al*, PIBASE, *Bioinformatics* 2005
Chandonia *et al. LIGBASE, Nucl Acids Res* 2004



Ligand:
250-1000 Da

Total: 3,643 families (SCOP v1.73)

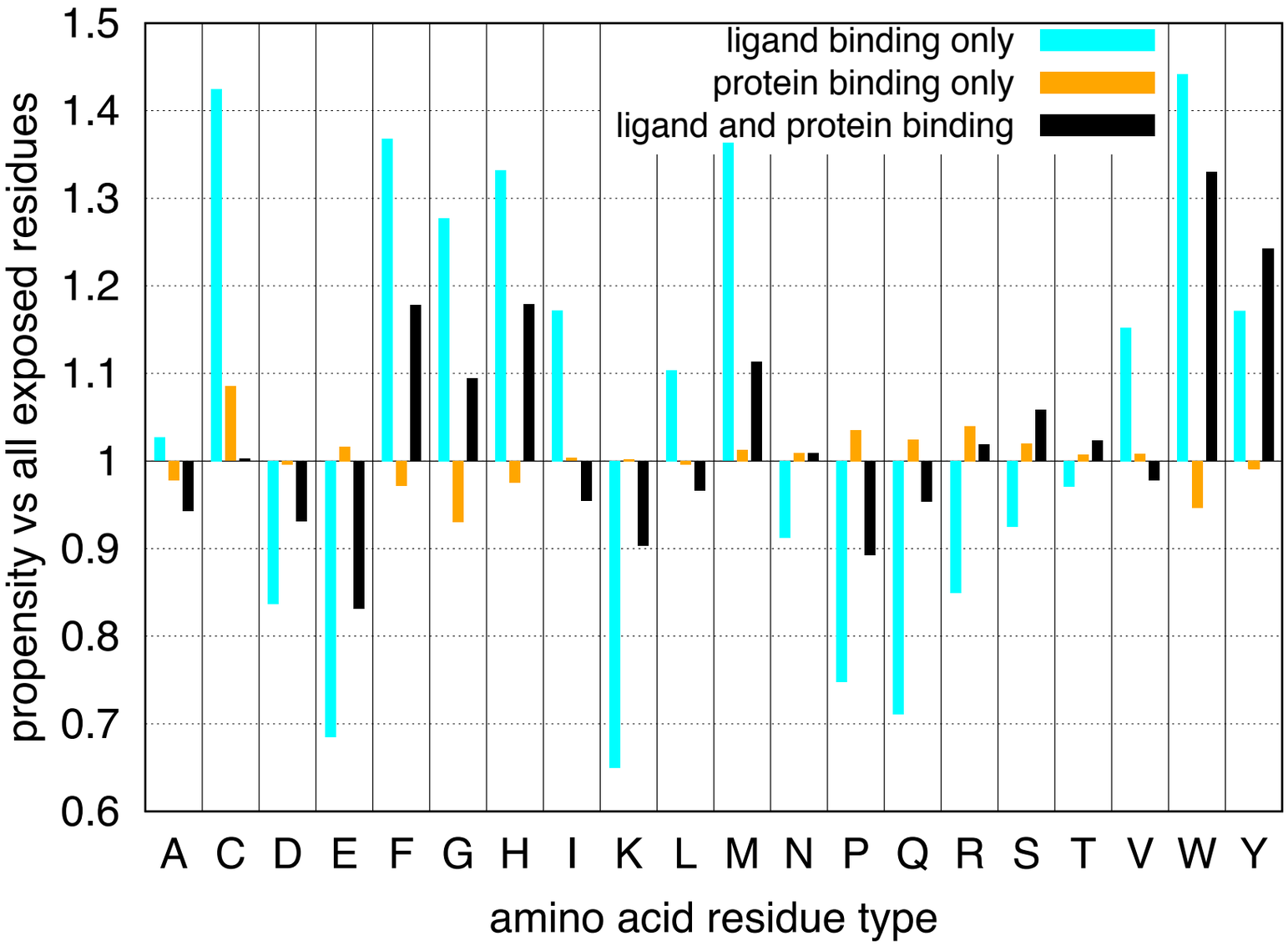
Number of families	Total	bind small molecules	≥ 5 bi-functional positions
Total	3,463	1,131	
Domain–peptide	469	232	150
Domain–domain, inter-chain	2,120	900	570
Domain–domain, intra-chain	1,189	562	356
Total protein-binding	2,619	1,028	736*

Davis and Sali, *PLoS Comp Biol* 2010.

*197 statistically significant

Properties of bi-functional positions: composition and conservation

- residue propensities are significantly different than mono-functional positions; similar to energetic hot-spots
- enriched for Trp and Tyr;
- less conserved than other surface positions



Davis and Sali, *PLoS Comp Biol* 2010.

Overlapping binding sites can suggest structural mechanisms for the effects of small molecules

- Bepridil was an FDA-approved Ca++-channel blocker for refractory angina

- Bepridil inhibits cellular entry of anthrax edema and lethal factors. (Sanchez, *et al. Antimicrob Agents Chemother* 2007)

Calmodulin – Anthrax Edema Factor (1K93; 2002)
Troponin C – bepridil (1MRW; 2002)

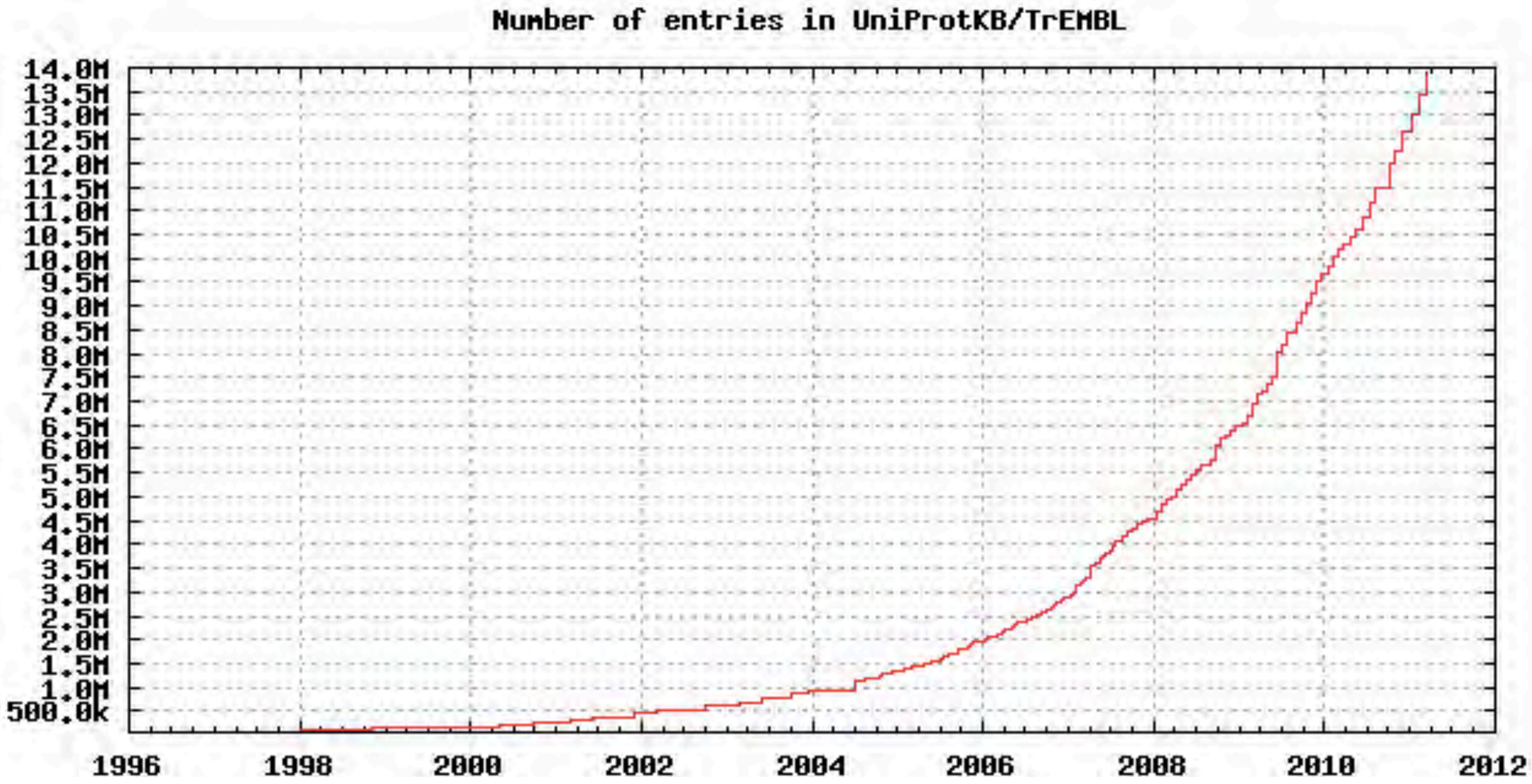
Davis and Sali, *PLoS Comp Biol* 2010.

Contents

1. Vignettes:
   - Specificity of Brain Lipid-Binding Protein (BLBP)
   - Identifying binding sites on proteins
   - Comparative “docking” of small molecules to proteins
   - Overlap between binding sites for proteins and small molecules
2. Enzyme Function Initiative
3. Docking against comparative models
4. Application to norepinephrine transporter (NET)

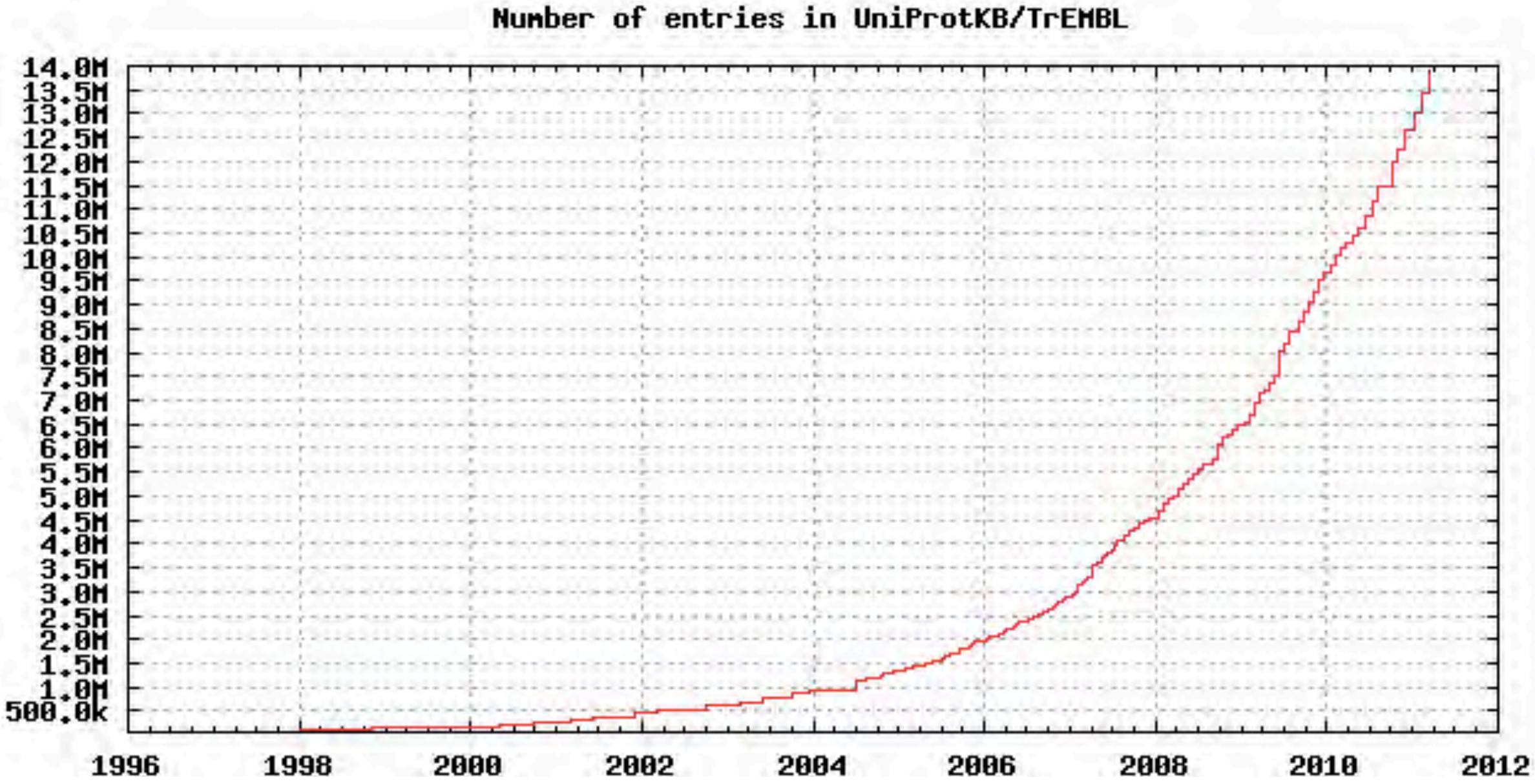
The number of protein sequences is “exploding” !

Release 2011_03 of 08-Mar-2011 of UniProtKB/TrEMBL contains 13897064 sequence entries, comprising 4465597779 amino acids .



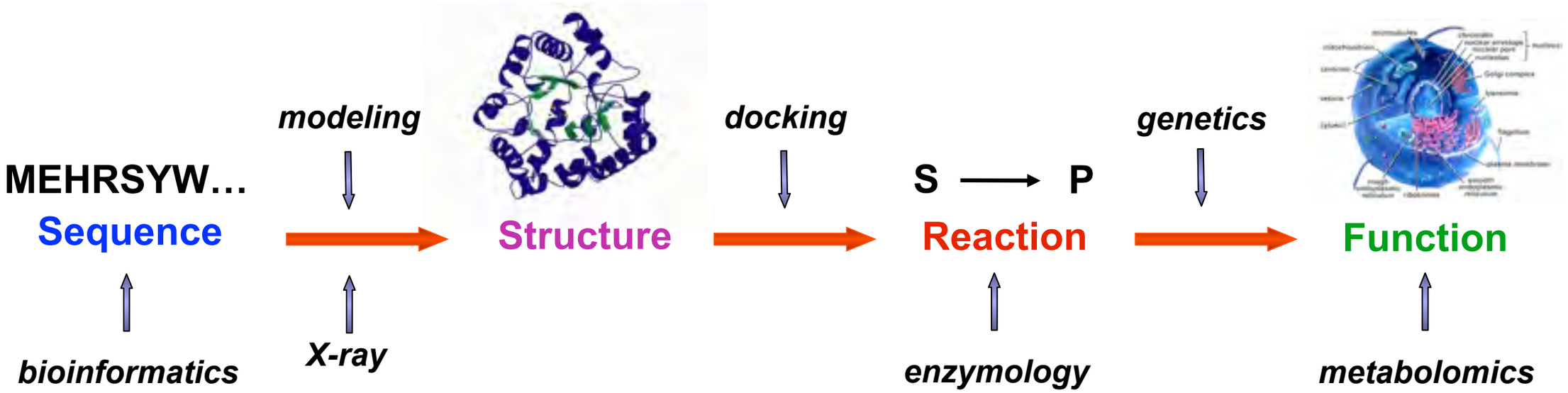
At least one-half have unknown/uncertain functions

Release 2011_03 of 08-Mar-2011 of UniProtKB/TrEMBL contains 13897064 sequence entries, comprising 4465597779 amino acids .



Functional assignment: high-throughput computation ?

U54 GM093342: "Enzyme Function Initiative" (EFI)



Illinois
John Gerlt
John Cronan
Jonathan Sweedler

Texas A&M
Frank Raushel

UCSF
Patricia Babbitt
Matthew Jacobson
Andrej Sali
Brian Shoichet

Albert Einstein
Steven Almo

University of Virginia
Wladek Minor

Boston University
Karen Allen

University of New Mexico
Debra Dunaway-Mariano

University of Utah
C. Dale Poulter

VanderbIlt University
Richard Armstrong

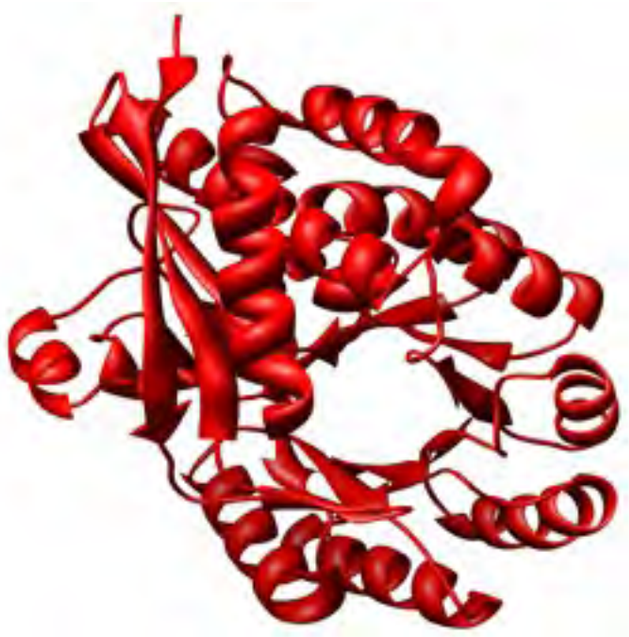
EFI: Deliverables

1. **Develop a robust sequence/structure-based strategy** for facilitating discovery of *in vitro* enzymatic and *in vivo* metabolic/physiological functions of unknown enzymes discovered in genome projects.

2. **Disseminate to the community** the intellectual, computational, and experimental tools, protocols, materials, and guidelines for determining *in vitro* and *in vivo* functions of unknown enzymes.

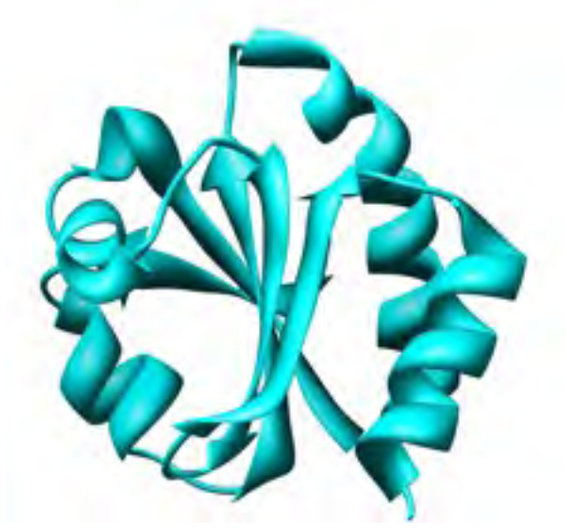
3. **Collaborate with the community** to facilitate sequence/ superfamily analyses as well as homology modeling and *in silico* docking of ligand libraries to unknown membes of other enzyme superfamilies.

EFI targets: 5 structurally diverse superfamilies



Enolase: > 7,000 sequences
Amidohydrolase: > 23,000 sequences

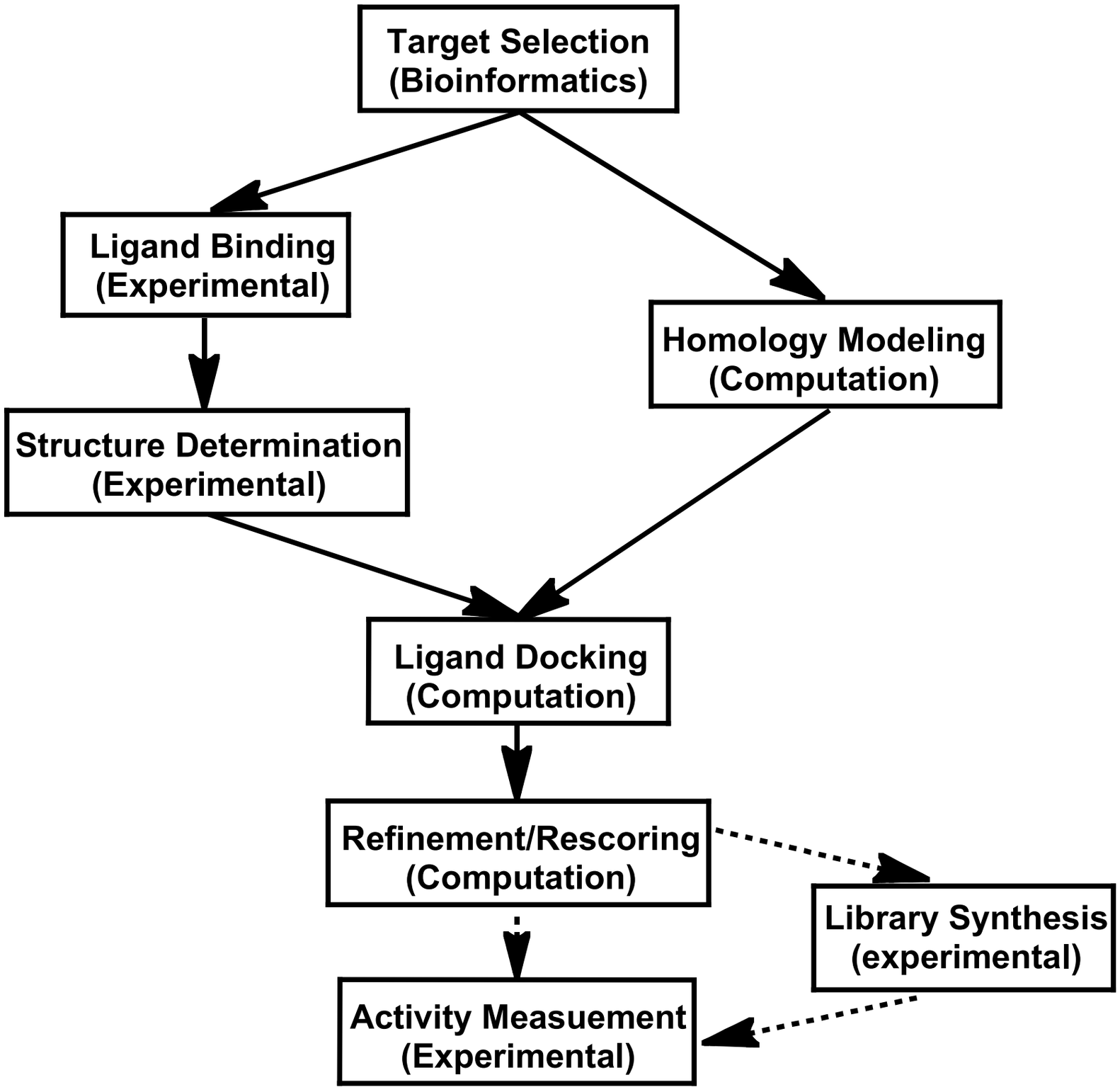
HAD: > 34,000 sequences



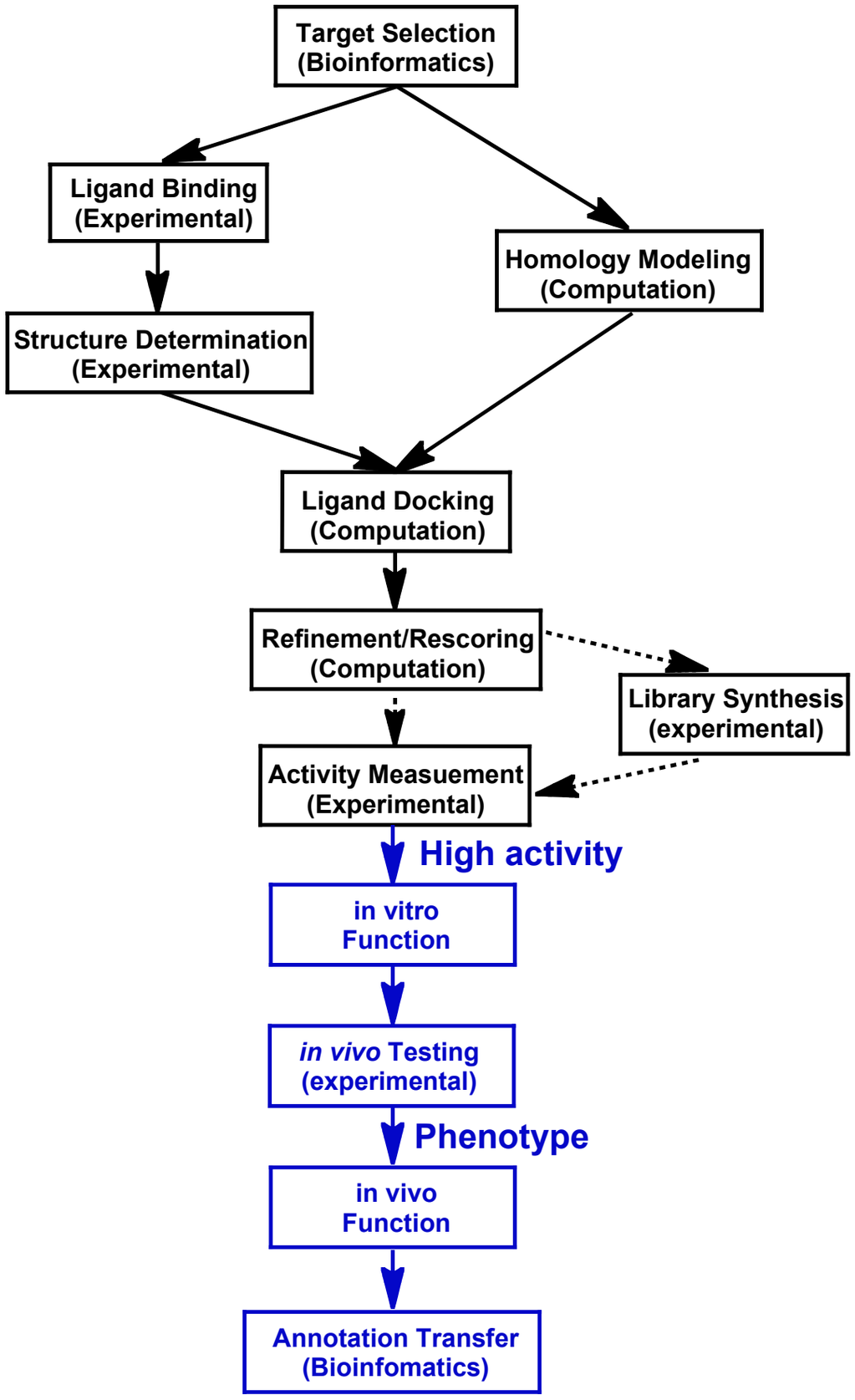
GST: > 8,000 sequences

Isoprene synthase: > 4,700 sequences

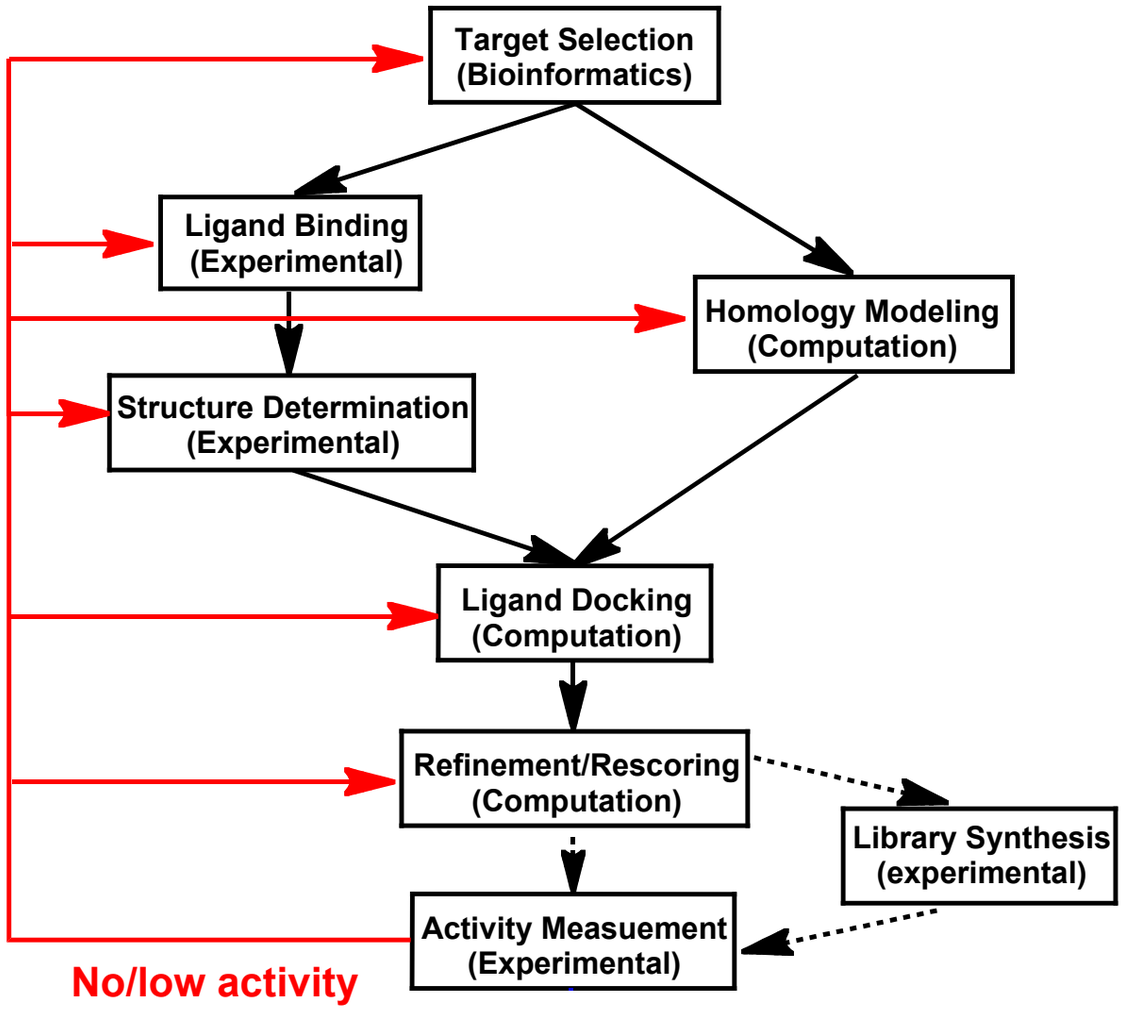
EFI pipeline: develop function assignment strategy



EFI pipeline: if correct, functional assignment

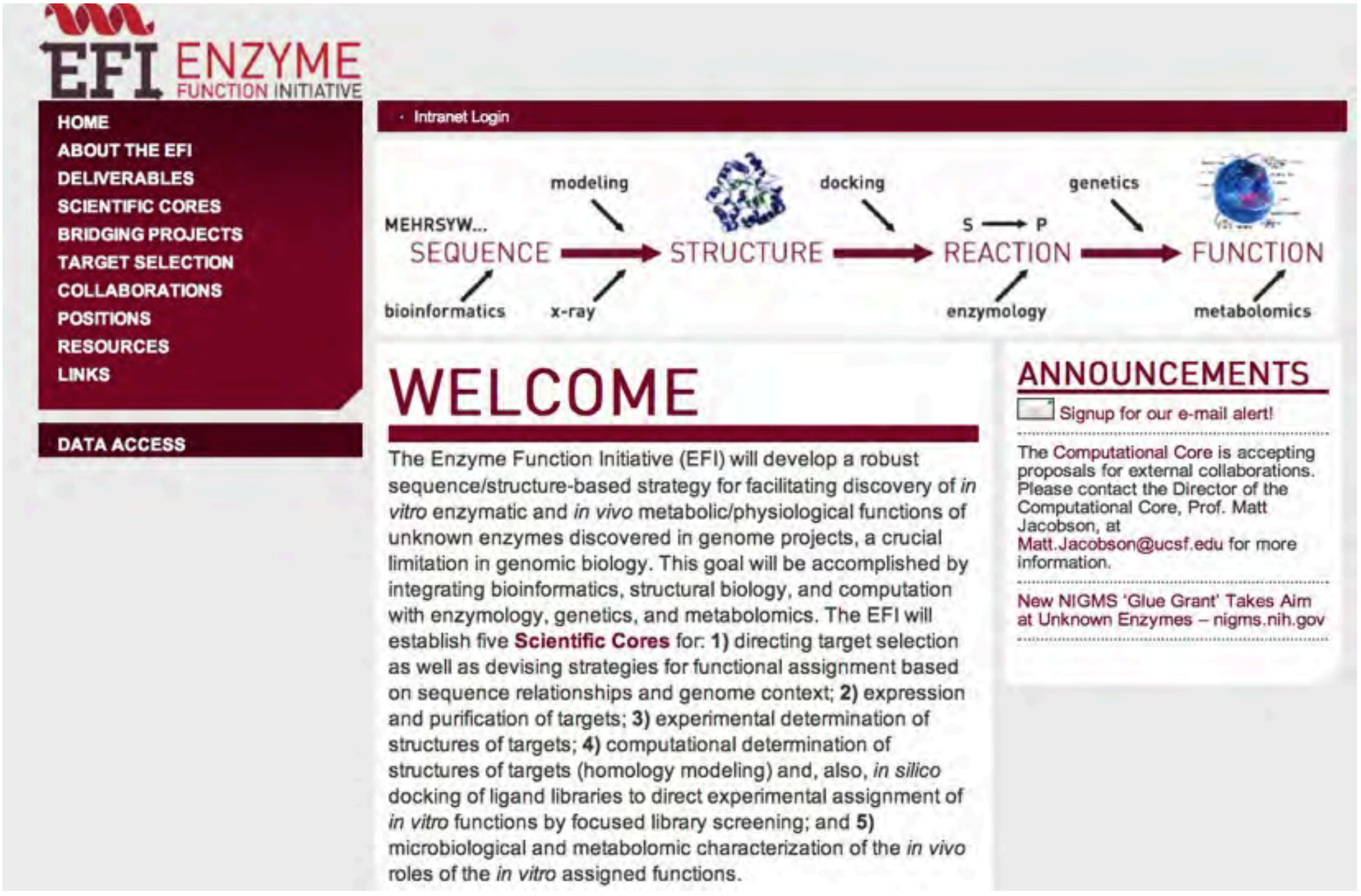


EFI pipeline: if incorrect, inform and improve strategy



Enzyme Function Initiative

http://enzymefunction.org



Genome-Wide Mapping of Protein-Ligand Interactions

All known protein structures
All known protein sequences

Comparative modeling

Refine protein models

Identify ligand
binding sites on models

Lists of small ligands

Virtual ligand libraries

Annotated
protein structure models

Build ligand-protein
complexes

Rescore ligand-protein
complexes

Central data warehouse

Graphical User Interface

Testing and Applications

Testing

Computational applications

Computer hardware and software environment

Software backplane

Global optimization

Information navigation and
search

Cluster hardware and
software

Benchmarking Comparative Models In Virtual Ligand Screening

Hao Fan, John J. Irwin, Benjamin M. Webb, Gerhard Klebe,

Brian K. Shoichet, and Andrej Sali

J. Chem. Inf. Model., 2009

Questions

1. How does docking against comparative models compare to random selection?

2. How does docking against comparative models compare to docking against the template structures?

3. If multiple models are calculated on the basis of different templates, can any of them outperform apo and even holo X-ray structures of the target?

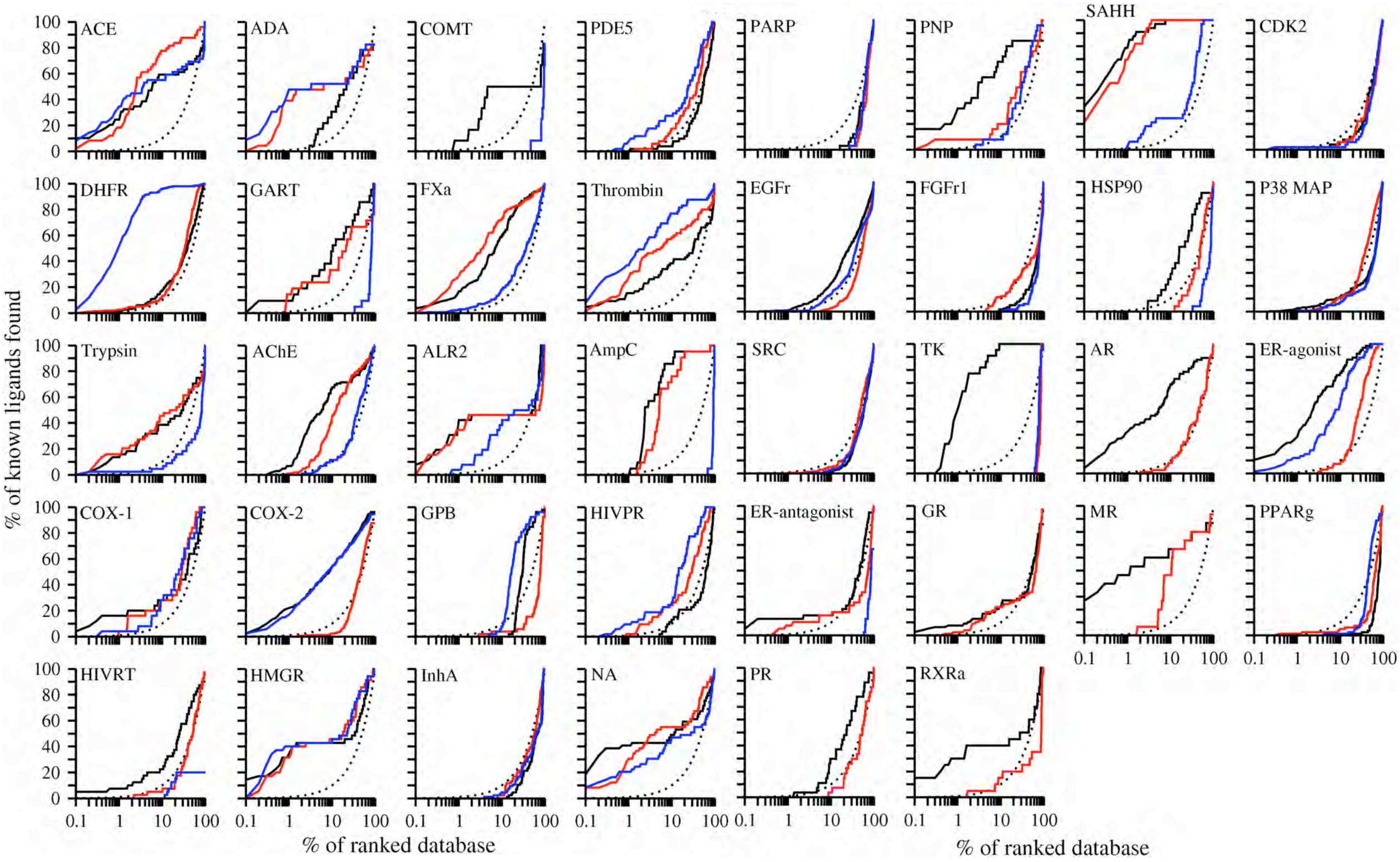
4. If so, can one reliably identify which model will be most enriching?

5. Can the docking screens be improved by employing multiple models?

Method

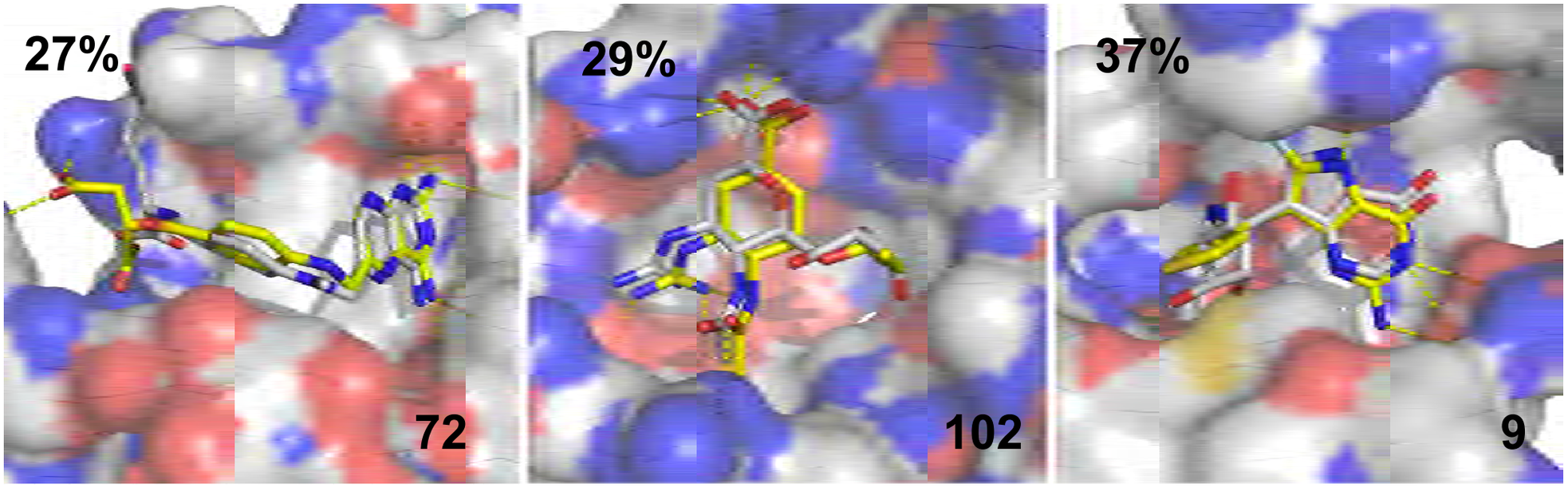
- Automated comparative modeling by MODELLER (Sali & Blundell, *J Mol Biol*, 1993).
- Automated virtual screening by DOCK (Meng, Shoichet, & Kuntz, *J Comp Chem*, 1992).
- “Directory of useful decoys” (DUD): 38 proteins with known ligands, ligand decoys, apo, holo, and related X-ray structures(Huang *et al*, *J Med Chem,* 2006).
- Consensus enrichment for virtual screening: For a given target protein, rank a compound by the best score against any of the alternative models or templates for the target (this project).

Results: Enrichment curves for DUD targets



Results: Sample accurate docking poses

Sequence identity and rank among 95,316 DUD decoys.



Both scoring and sampling need to be improved.

Answers

1. **How does docking against comparative models compare to random selection?** Comparative models typically outperform random selection significantly, doing so for 27 out of the 38 targets.
2. **How does docking against comparative models compare to docking against the template structures?** For the entire benchmark, comparative models are on average no more enriching than the corresponding templates. This measurement, however, is confounded by the likelihood of orthologous templates genuinely recognizing the ligands for the modeled target. Conversely, a modeled structure based on a paralogous template with at least 25% sequence identity to the target is 2.5 times more likely to be significantly more enriching than the template.
3. **If multiple models are calculated for a target, each one based on a different template, can any of them outperform apo and even holo X-ray structures of the target?** Typically, the holo X-ray structure returns the best enrichments, but the modeled structures are often competitive. For 15 of the 38 targets, the most enriching model is better for virtual screening than the holo X-ray structure; for nine targets, the most enriching model is as good as the holo X-ray structure. As compared to apo X-ray structures, the best model performance is better still.
4. **Can one reliably identify which model will be most enriching?** No, none of the tested sequence or structural attributes (i.e., the overall target-template sequence identity, the binding site target-template sequence identity, and the predicted accuracy of a model) can reliably predict the accuracy of ligand docking.
5. **Can the docking screens be improved by employing multiple models instead of a single model?** Yes. For the 38 targets, the enrichment of the model based on the highest sequence identity is better than or comparable to the enrichment for the apo and holo X-ray structures in 65% and 45% cases, respectively; in contrast, the consensus enrichment for multiple models (and templates) is better than or comparable to the enrichment for the apo and holo X-ray structures in 70% (79%) and 47% (50%) cases, respectively. For the 222 target-template pairs, the consensus enrichment is better and worse than the template enrichment in 23% and 10% of the cases, respectively. For the 87 paralogous target-template pairs related at more than 25% sequence identity, the consensus enrichment is better and worse than the template enrichment in 33% and 3% of the cases, respectively.

Structure-based discovery of prescription drugs that interact with the norepinephrine transporter (NET)

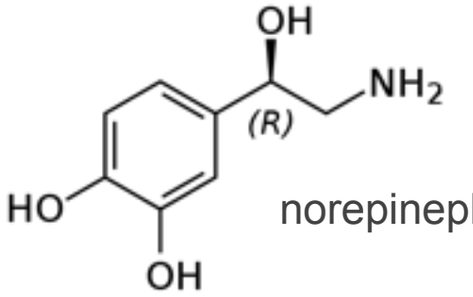
Avner Schlessinger, Ethan Geier, Hao Fan, John J. Irwin, Brian K. Shoichet,

Kathleen M. Giacomini, and Andrej Sali

The norepinephrine transporter (NET, SLC6A2)

Biological Function:

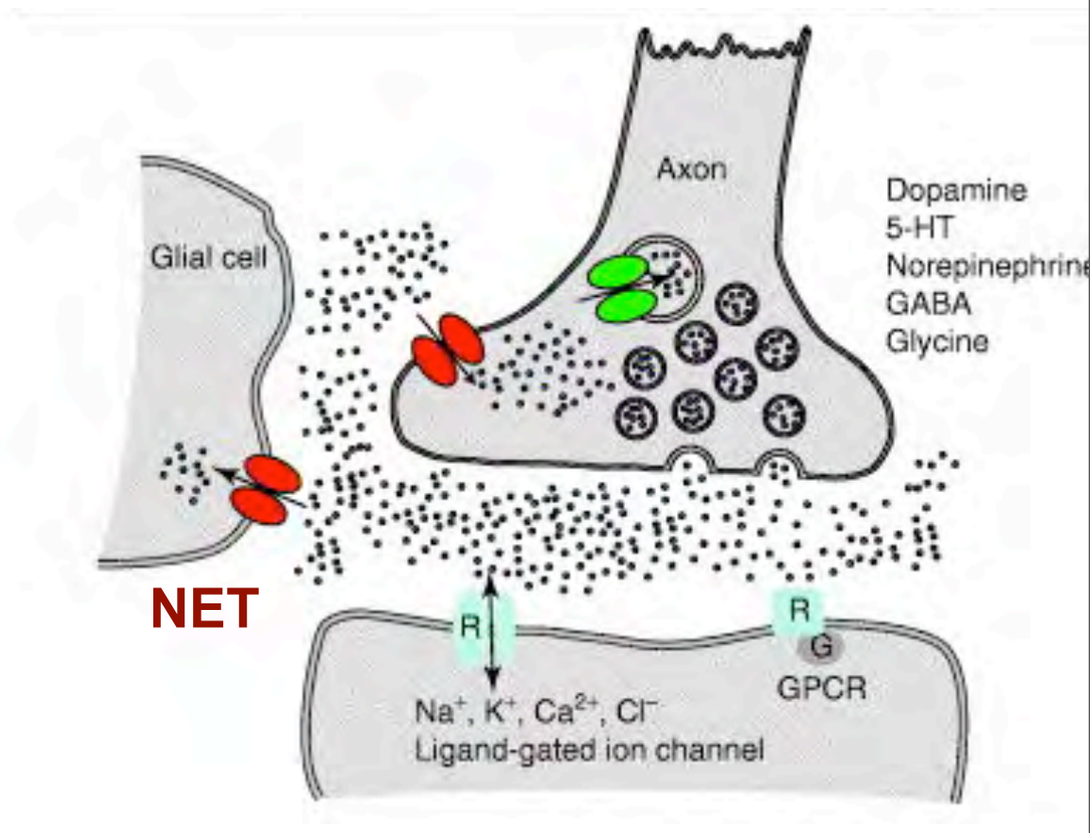
- Na^+- and Cl^-- dependent neurotransmitter transporter, from the synapse to presynaptic neurons
- Mutations are associated with attention deficit hyperactivity disorder (ADHD), panic disorder, and severe orthostatic hypotension



norepinephrine

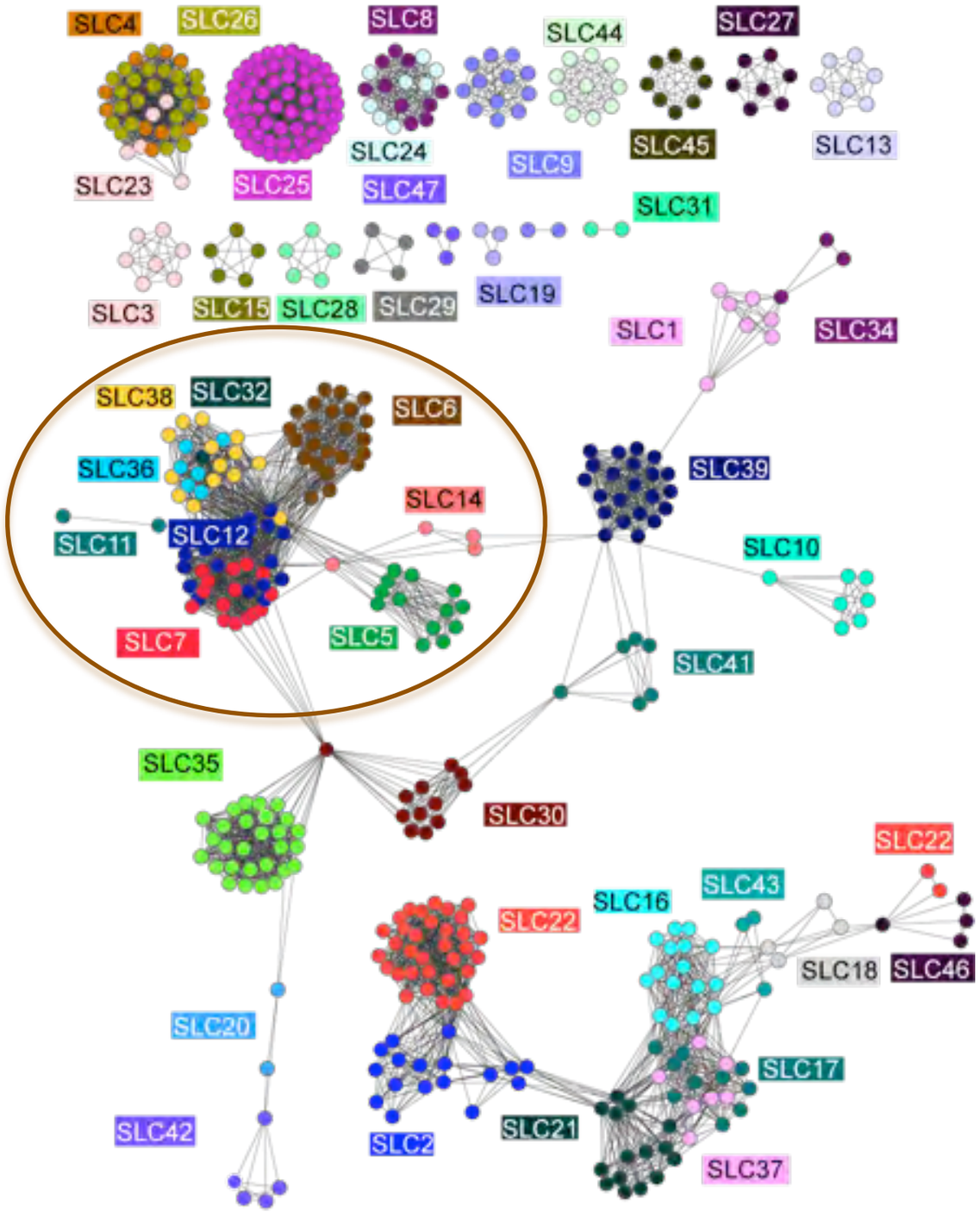
Pharmacology:

- Antidepressants, psychostimulants, ADHD, neuropathic pain, weight loss, nasal decongestion, hypotension

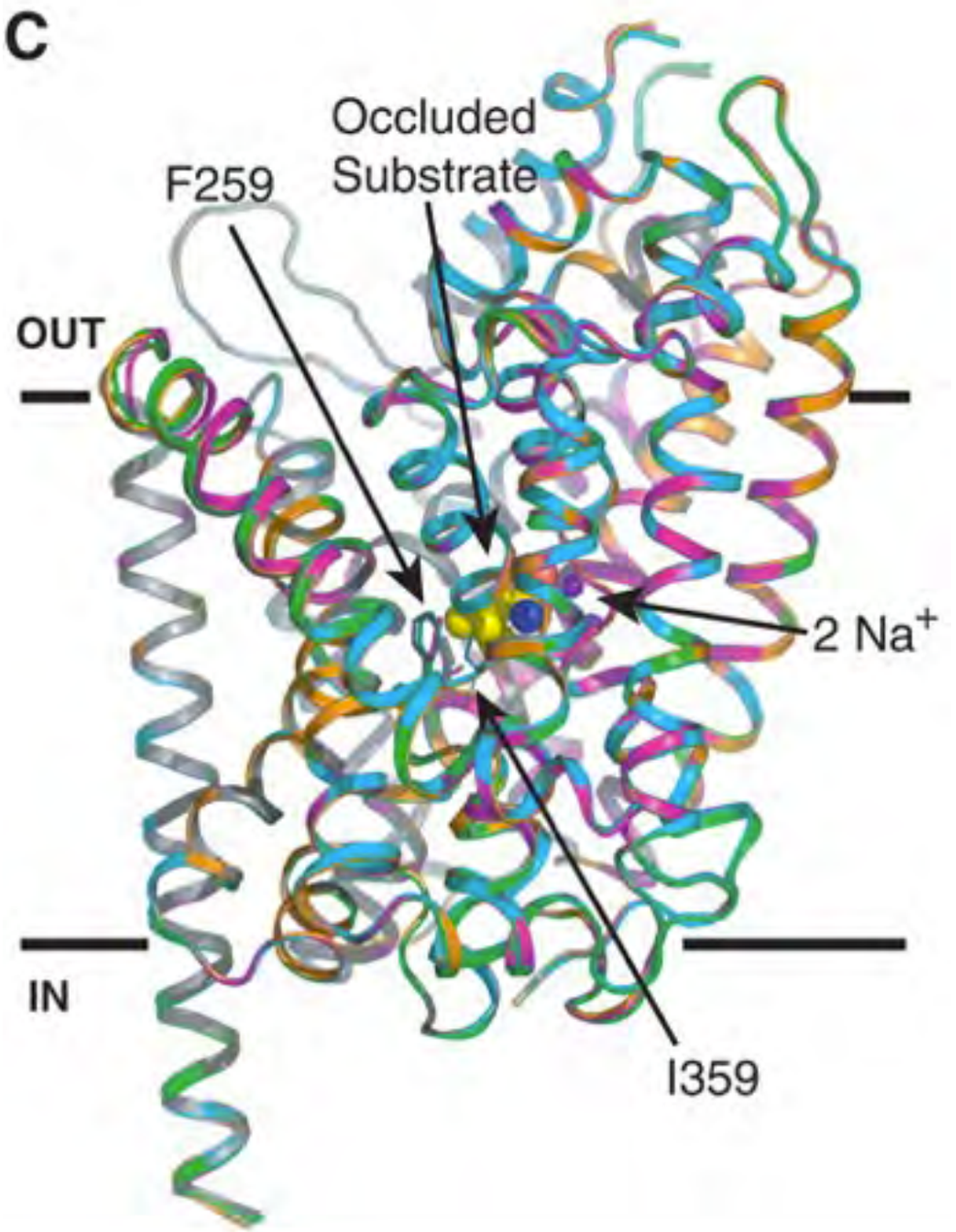


Schousboe *et al*. Trends Pharmacol Sci., 2006.

Solute Liquid Carrier (SLC) transporters
NET has the neurotransmitter: sodium symporter-like fold



Schlessinger *et al*. Protein Sci., 2010.



LeuT transporter from *Aquifex aeolicus*

Sing *et al*. Science, 2008.

Goals of NET ligand discovery

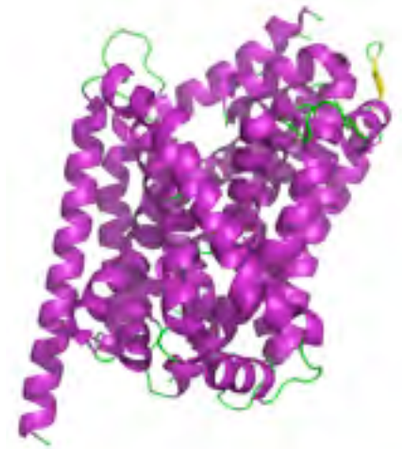
Pharmacogenetics of Membrane Transporters (K. Giacomini, UCSF)
Center for Structures of Membrane Proteins (R. Stroud, UCSF)

- Find unknown biological functions
- Find leads for drug development (*eg*, Pt derivative transport)
- Explain drug efficacy and side effects
- Rationalize impact of point mutations on the function
- Describe substrate specificity in the SLC6 family
- Aid crystallographic structure determination

Approach: comparative modeling, docking, and virtual screening

1. Search for template
PDB, OPM

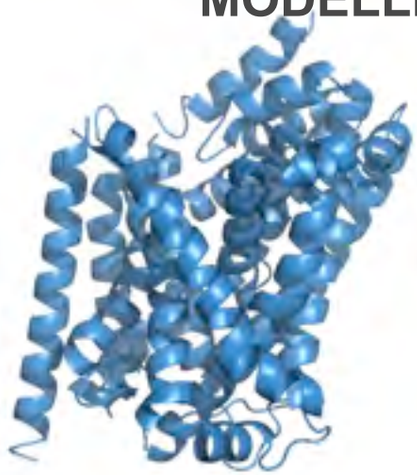
Target sequence:
GGMEAVITGLADDFQAA

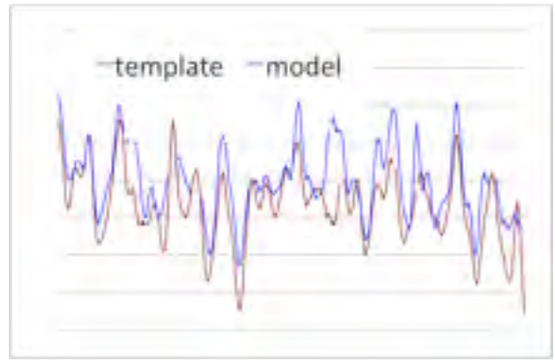


2. Align target and template
SALIGN, PROMALS3D

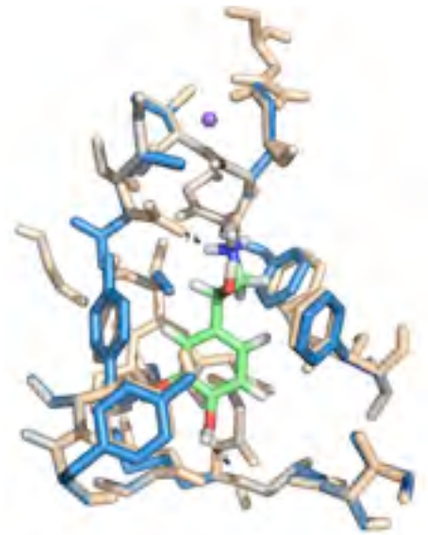
```
GGMEAVITGLADDFQAA
AIM--QPMIAFLEDELKL-
```

3. Construct and assess model
MODELLER, DOPE

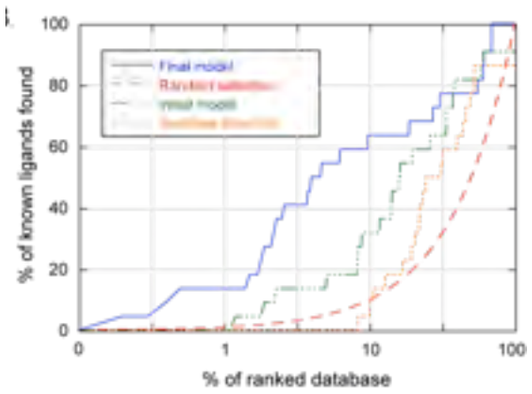




4. Refine side chains
SCRWL4

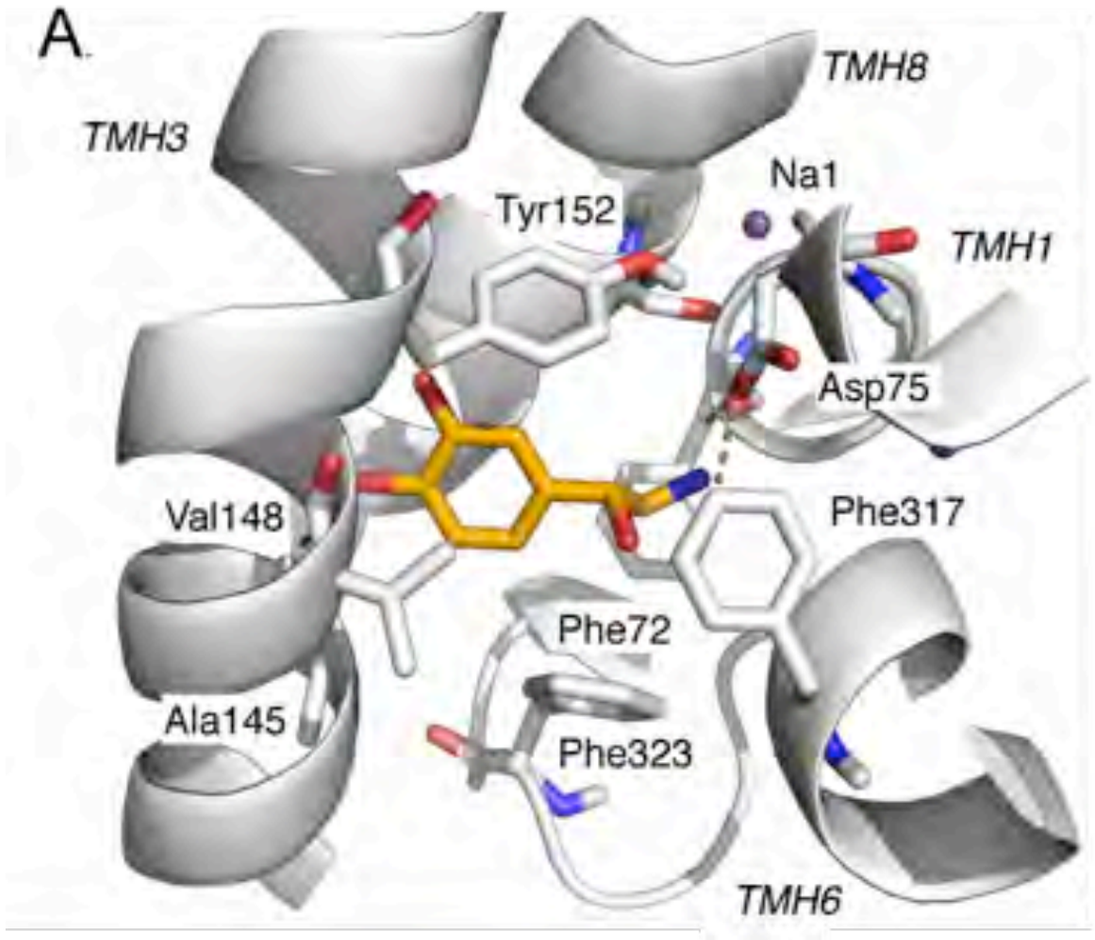


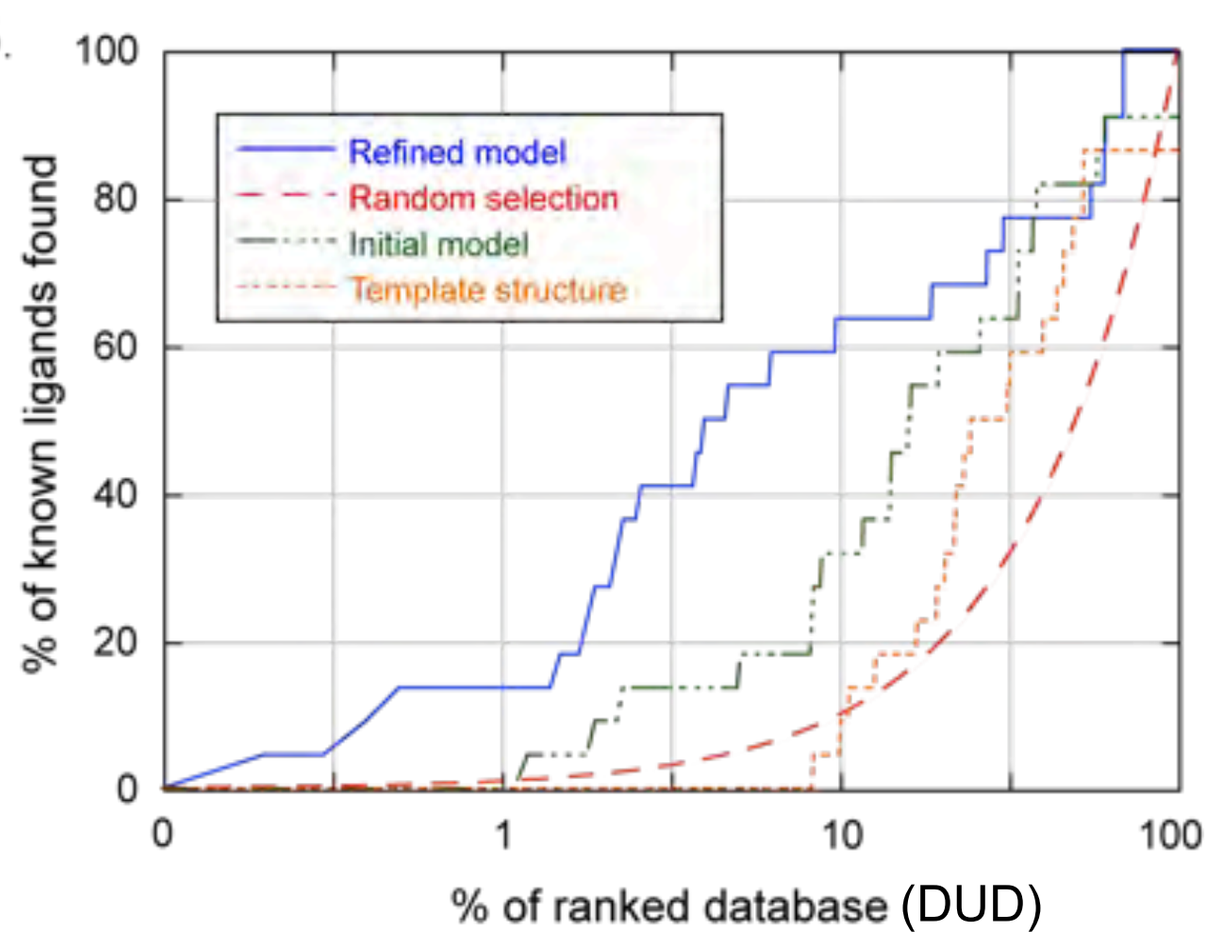
5. Validate binding site
DOCK, DUD



6. Virtual screening
DOCK

Final model can discriminate between ligands and non-ligands





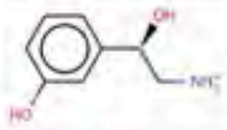
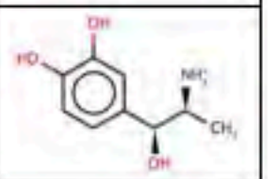
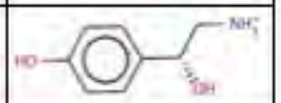
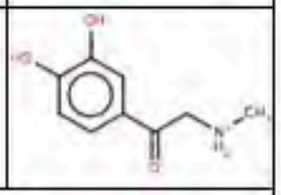
logAUC for final model 37.6
(without optimizing for enrichment)

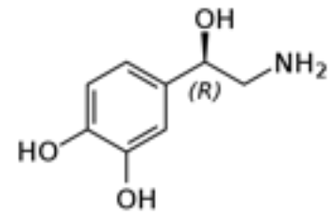
Virtual screening of 6,436 drugs from the Kyoto Encyclopedia of Genes (KEGG DRUG) database against the NET model

- 200 highest-ranked drugs (the top 3.1% of the library) were analyzed manually for the similarities of their predicted poses to those in structurally defined complexes, frequent scaffolds, and common pharmacological function.

- **5** high-confidence (similar to NE) and **13** medium-confidence (dissimilar to NE) compounds were selected for validation in the lab.

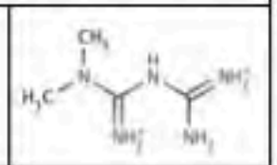
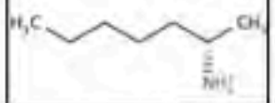
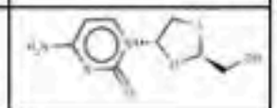
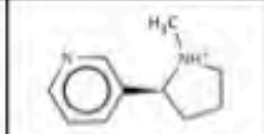
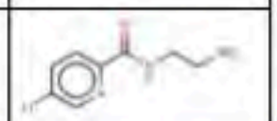
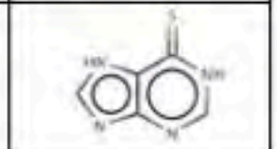
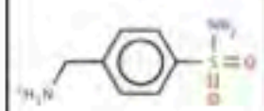
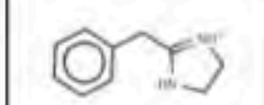
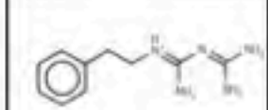
High-confidence hits

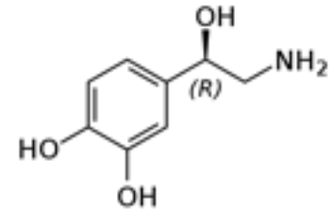
High-confidence predictions					
Rank[a]	ZINC id[b]	Name[c]	Indication[d]	Tc[e]	Structure[f]
3	119286	Norfenefrine	Cardiac stimulant	0.34 (0.94)	
4	34159	Levonordefrin	Vasoconstrictor; hypotensive; topical nasal decongestant	0.47 (0.90)	
10	388198	Octopamine	Cardiac stimulant; hypotensive	0.34 (0.90)	
25	1482197	Tranylcypromine	Antidepressant	0.37 (0.38)	
112	57542	Adrenalone	Hemostatic; vasoconstrictor	0.40 (0.61)	



norepinephrine

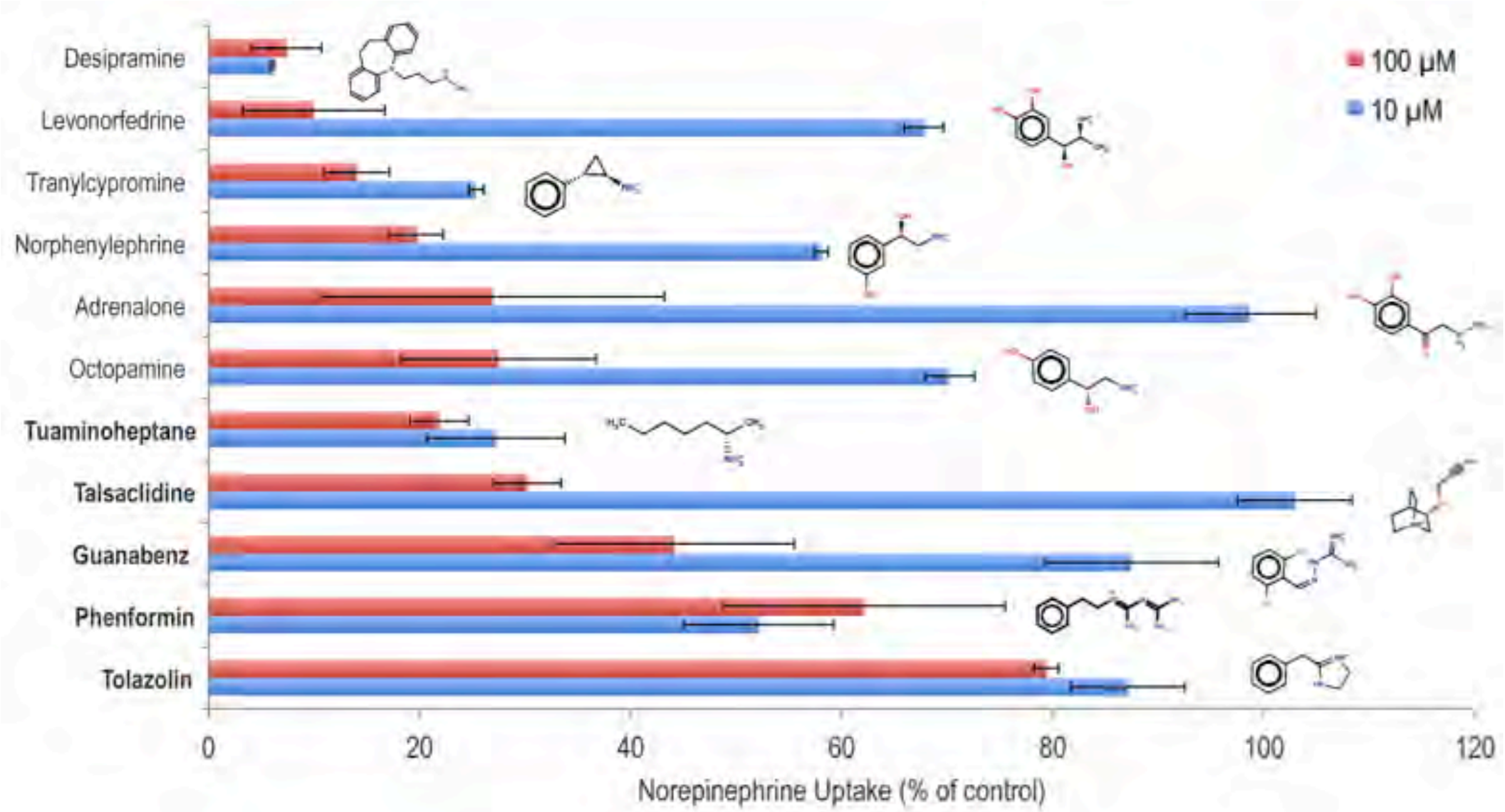
Medium-confidence hits (10 of 13)

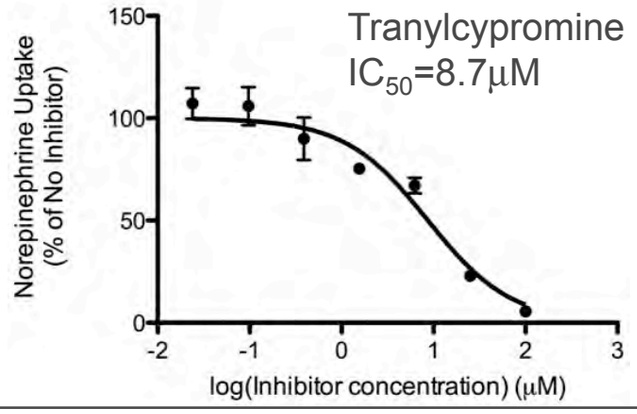
Medium-confidence predictions					
16	896658	Metformin	Antidiabetic	0.13 (0.09)	
18	2015538	Tuaminoheptane	Nasal decongestant; stimulant; vasoconstrictor	0.23 (0.18)	
40	16952920	Lamivudine	Antiviral (HIV and HBV).	0.18 (0.14)	
46	391812	Nicotine	Stimulant	0.28 (0.46)	
79	1542229	Lazabemide	Alzheimer disease	0.25 (0.20)	
92	18119521	6-mercaptopurine	Antineoplastic; immunosuppressive; antimetabolite	0.13 (0.16)	
93	1535336	Talsaclidine	Alzheimer's Disease	0.17 (0.26)	
98	1644	Mafenide	Severe burns; synthetic antimicrobial agent	0.35 (0.56)	
104	125006	Tolazolin	Vasodilator (peripheral)	0.29 (0.45)	
124	14768667	Phenformin	Antidiabetic	0.29 (0.35)	



norepinephrine

Experimental validation of 18 hits
(positives only shown)





NET uptake inhibition measured in transfected human embryonic kidney 293 (HEK293) cells
Chen *et al*, J Pharmacol Exp Ther., 2007

Clinical implications

Membrane transporters in drug development

*The International Transporter Consortium**

Abstract | Membrane transporters can be major determinants of the pharmacokinetic, safety and efficacy profiles of drugs. This presents several key questions for drug development, including which transporters are clinically important in drug absorption and disposition, and which *in vitro* methods are suitable for studying drug interactions with these transporters. In addition, what criteria should trigger follow-up clinical studies, and which clinical studies should be conducted if needed. In this article, we provide the recommendations of the International Transporter Consortium on these issues, and present decision trees that are intended to help guide clinical studies on the currently recognized most important drug transporter interactions. The recommendations are generally intended to support clinical development and filing of a new drug application. Overall, it is advised that the timing of transporter investigations should be driven by efficacy, safety and clinical trial enrolment questions (for example, exclusion and inclusion criteria), as well as a need for further understanding of the absorption, distribution, metabolism and excretion properties of the drug molecule, and information required for drug labelling.

Giacomini *et al.* Nat Rev Drug Discov. 2010.

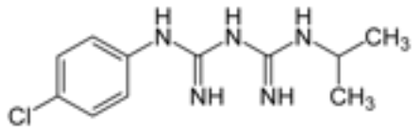
Clinical implications of NET positives

1. Possible efficacy of drugs for other primary targets (enzymes, receptors):

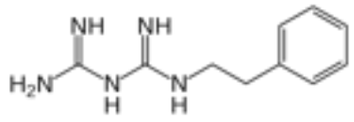
- Sympathetic drugs: Epirenor, Stryphnasal, Zondel, Corbadrin
- Antidepressants: Parnate (monoamine oxidase inhibitor)

2. **Possible side effects of drugs with other primary targets:**

- Anorexia and high blood pressure
- Phenformin (anti diabetic)

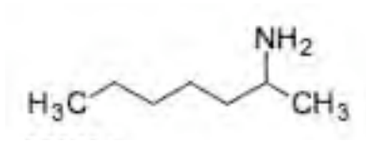


- Anti-malarial drug proguanil

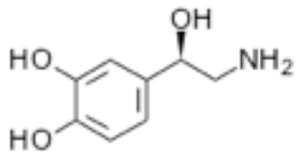


3. **Novel NET ligand scaffolds discovered:**

- *eg*, Tuaminoheptane – no aromatic ring



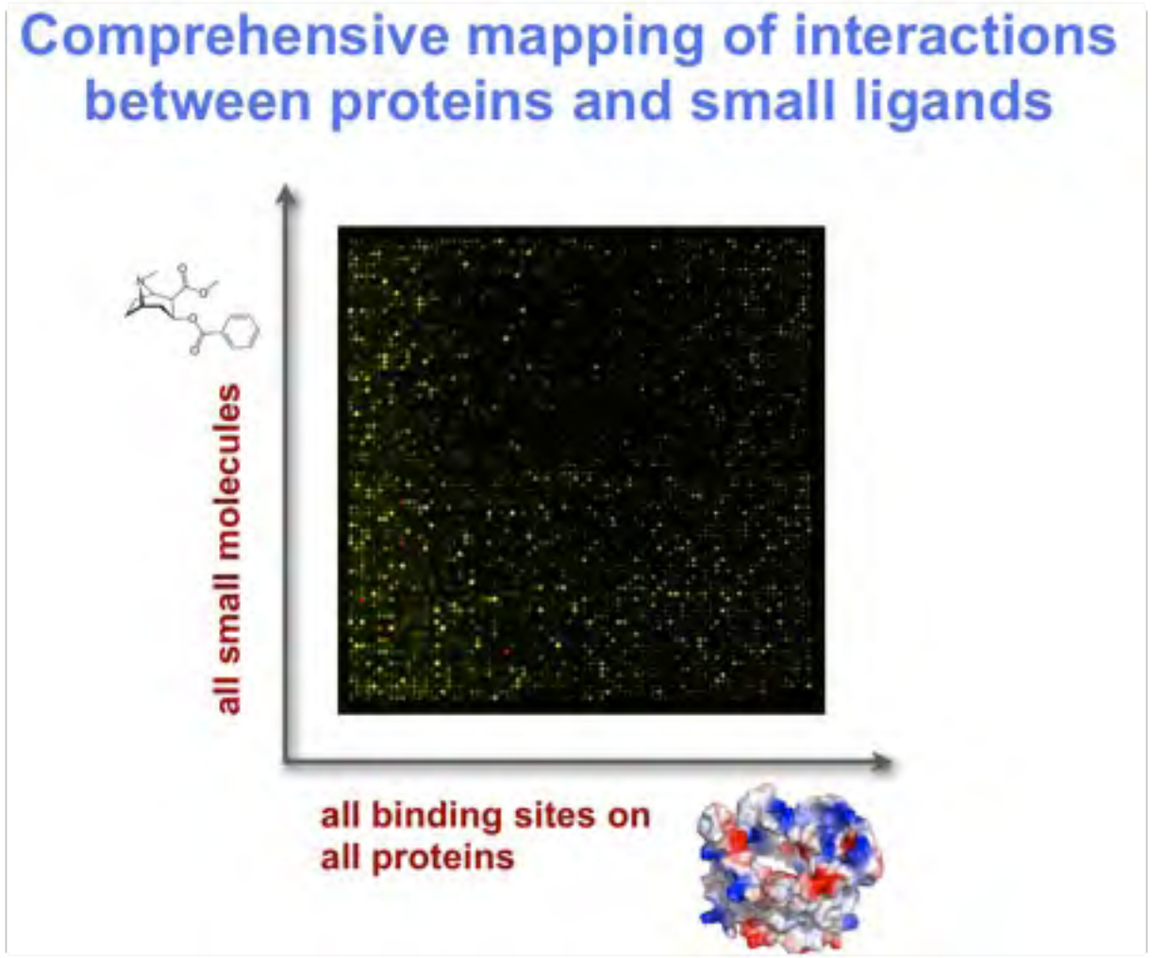
versus



Summary

Even when the target is a membrane protein sharing only 27% sequence identity and a dissimilar binding profile to the template structure, comparative modeling, docking, and virtual screening can be informative.

Future



It is difficult to imagine how significant progress towards this goal can be achieved without virtual screening against comparative models, though there are also many other bottlenecks. As we progress, an optimal, integrative approach involving a variety of techniques will evolve.

For docking against models, both scoring and sampling need to be improved.

Acknowledgments

http://salilab.org

QB3 @ UCSF:

Hao Fan
Avner Schlessinger
Patrick Weinkam
Quanqiang Dong
Jeremy Horst
Backy Chen
Ben Webb
Ursula Pieper
Elina Tjioe

Alumni:

Narayanan Eswar (Du Pont)
Marc Marti-Renom (Valencia)
Fred Davis (HHMI)
Roberto Sanchez (MSSM)
Andrea Rossi (Pfizer)
M.S. Madhusudhan (Singapore)
David Eramian (UCSF)
Min-Yi Shen (Applied Biosys)
Mark Peterson (BCG)
Francisco Melo (Catholic U.)
Ash Stuart (Rampallo Coll.)
Eric Feyfant (Pfizer)
Valentin Ilyin (NEU)
Andras Fiser (AECOM)
Ranyee Chiang (NYU)

Collaborators:

Brian Shoichet (UCSF)

Patsy Babbitt (UCSF)
Matt Jacobson (UCSF)
Frank Raushel (TAMU)
Steven Almo (AECOM)
Stephen Burley (Lilly)
Gerhard Klebe (Marburg U)
Robert Stroud (UCSF)
Kathy Giacomini (UCSF)
Wah Chiu (Baylor)
Judith Frydman (Stanford)
Charly Craik (UCSF)
Jim Wells (UCSF)
Tom Ferrin (UCSF)

NIH, NSF
The Sandler Family Foundation
Human Frontiers Science Program
IBM, Intel, Hewlett-Packard, NetApps
Pfizer, Structural Genomix
Pharmaceuticals, Mike Homer,
Ron Conway

10/15/2006

Statistical Potentials for Modeling and Ranking of Protein-Ligand Interactions

Hao Fan, Dina Schneidman, John J. Irwin, Guangqiang Dong, Brian K. Shoichet, and Andrej Sali

- idea of stat pot - DOPE joint pdf
- app to prot-lig eq, sample, atm types, reference, optimization on training (criteria)
- sample distribution
- testing on testing set
- results: absolute, relative
- discussion points: glass ceiling

DOPE philosophy

Equation

Conditional probability - statistical preference

$$P\left(C \mid d_{ij}\right) = P(C) * \frac{P\left(d_{ij} \mid C\right)}{P\left(d_{ij}\right)}$$

$P(C|d_{ij})$, the probability that a structure is correct, given distances $\{d_{ij}\}$. The score.

$P(d_{ij}|C)$ - the probability of observing $\{d_{ij}\}$ in a correct structure. Given by the sample of known structures.

$P(d_{ij})$ - the probability of observing $\{d_{ij}\}$ in any structure. The reference state - remains to be determined.

$P(C)$ - the probability of observing a correct structure. Constant for a given ligand (but not in virtual screening).

Variables

Protein atom type: 158 DOPE types.

Small molecule atom type: 26 Sybyl types

X-ray structure resolution: 2.0 Å, 2.5 Å.

Distance cut-off: 6 Å, 8 Å, 10 Å, 12 Å, 14 Å.

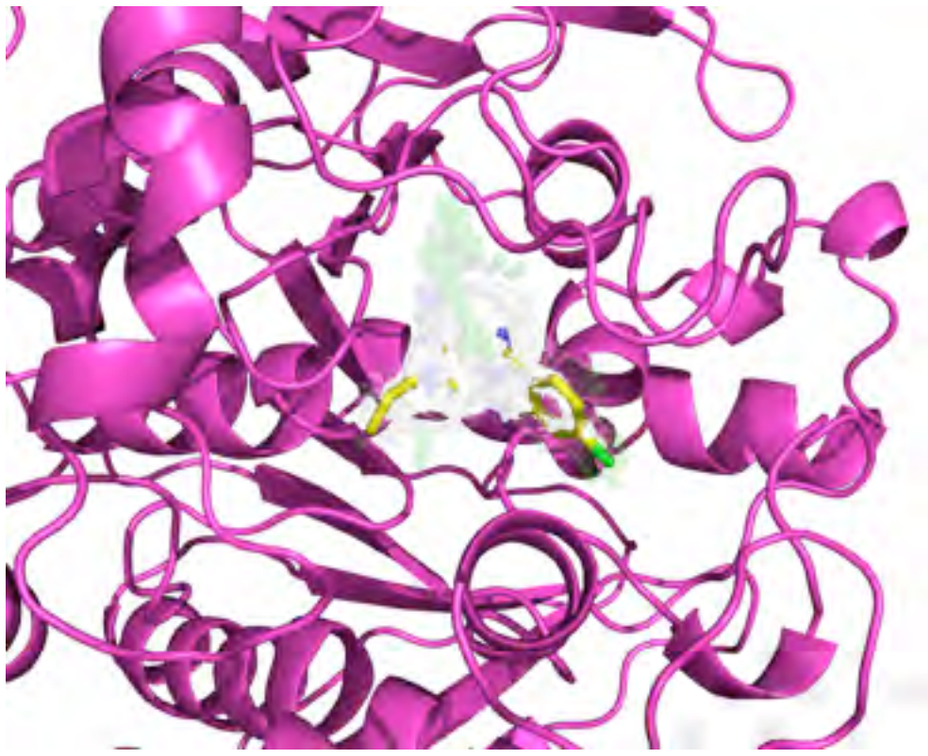
Reference equation: Sippl, DFIRE.

Reference resource: PDB,

PDB + DOCK decoys,

Random distribution.

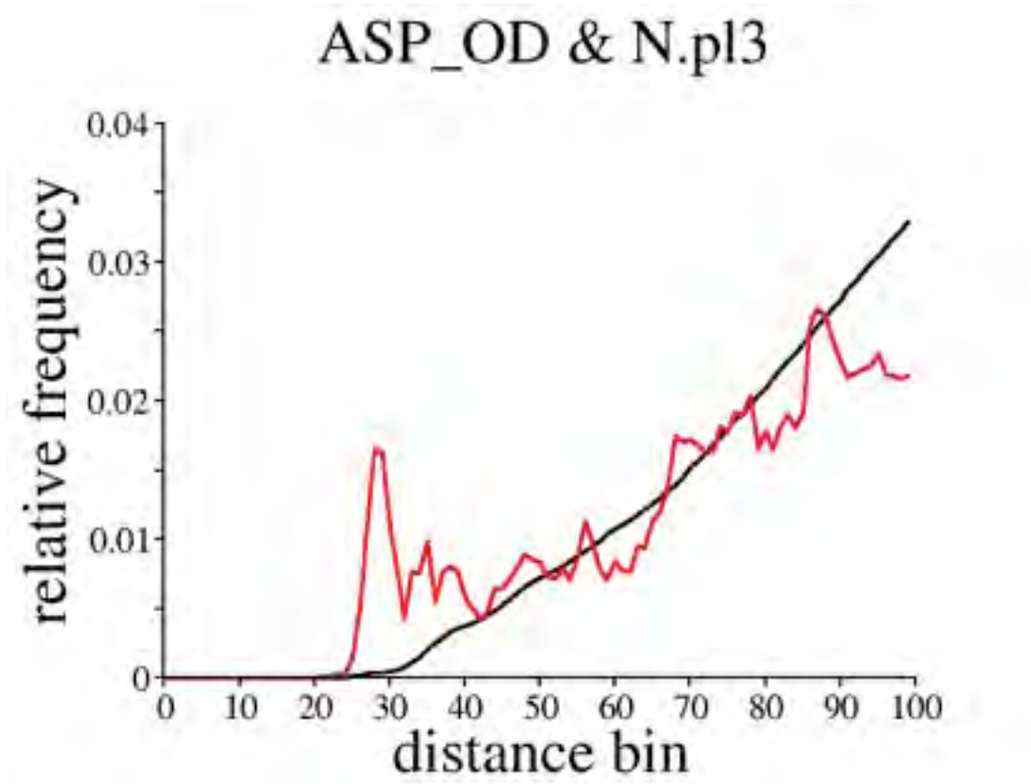
***P*(*C*) in virtual screening.**

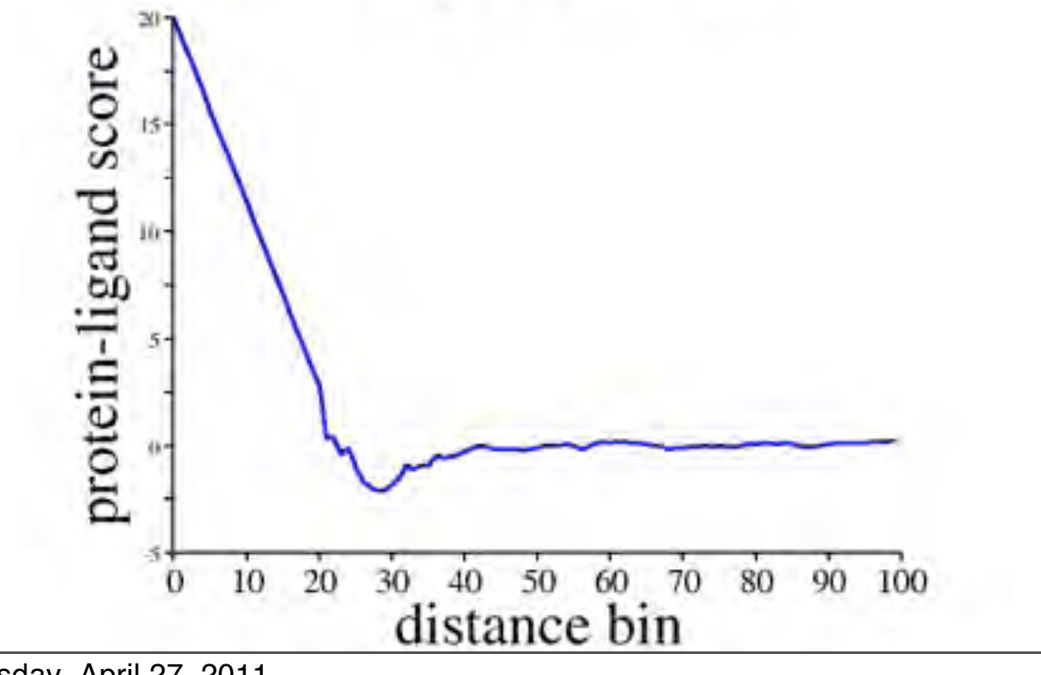


04/07/11

Pairwise score

One example for ASP_OD & N.pl3 (10, 0.1)



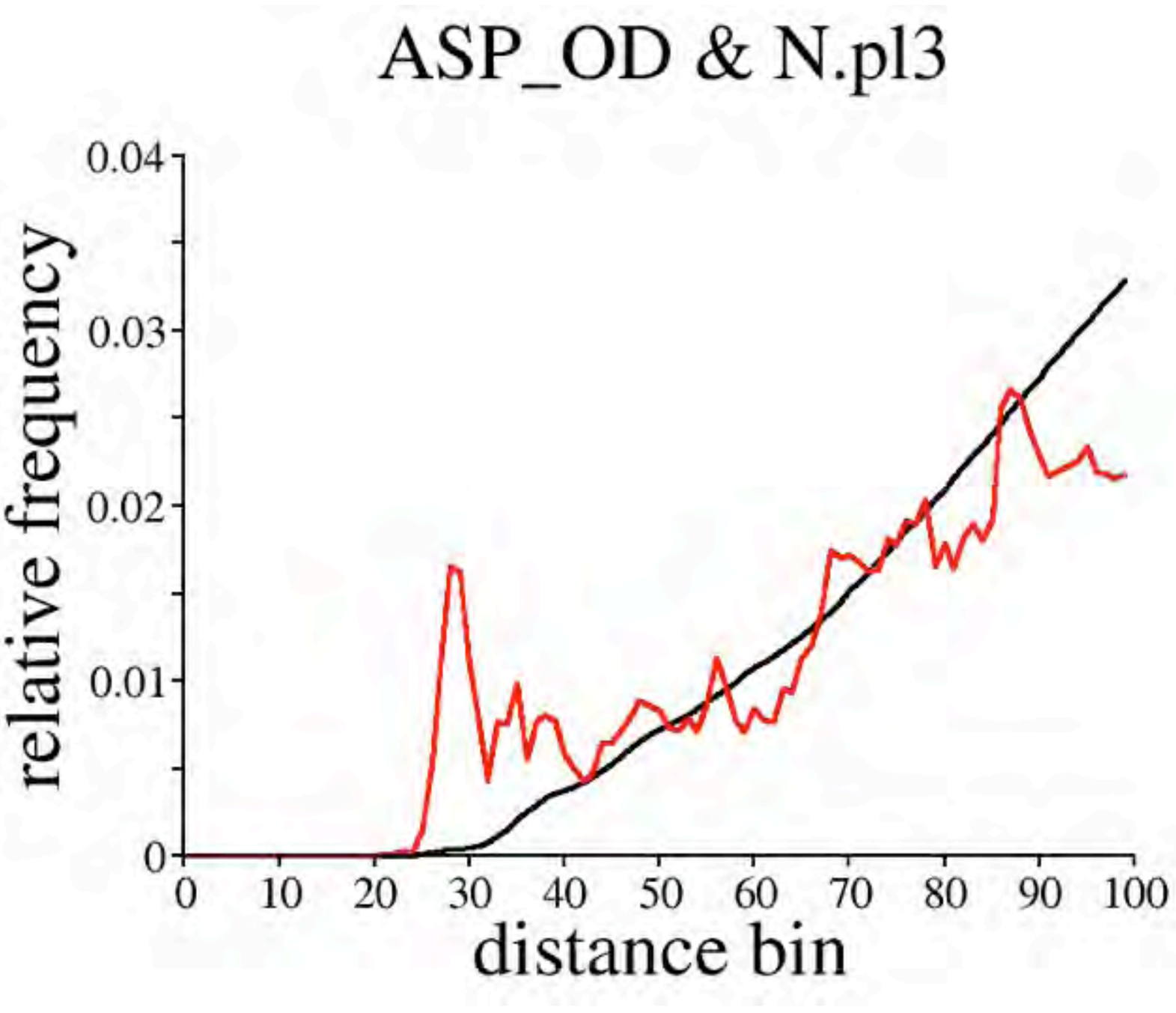


Sippl reference: assume no atom type difference

Sippl, J Mol Biol 1990

Pairwise score

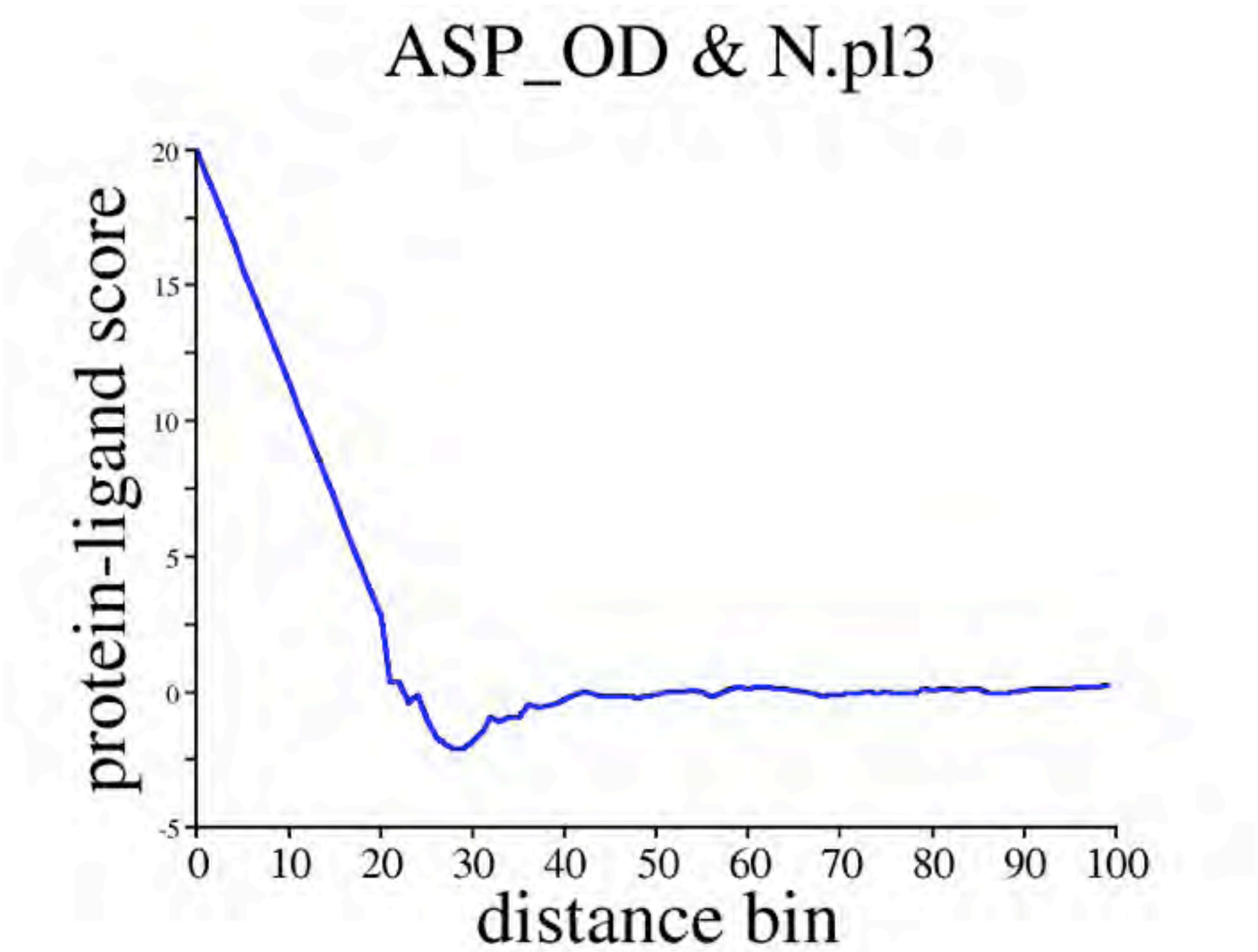
One example for ASP_OD & N.pl3 (10, 0.1)



Sippl reference: assume no atom type difference

Pairwise score

One example for ASP_OD & N.pl3 (10, 0.1)



Training and Validation

Geometry (single ligand, X-ray and decoy poses)

X-ray pose or decoy pose ≤ 2.0 Å ranked first

- **Training set on the basis of Astex, generated by DOCK**

  70 targets, 100 decoys

  For all targets, at least 1 decoy ≤ 2.0 rmsd Å to X-ray

- **Validation set - Wang's dataset, generated by autodock**

  100 targets, 100 decoys

  For 91 targets, at least 1 decoy ≤ 2.0 rmsd Å to X-ray

Training and Validation

Enrichment (ligands and decoy compounds)

compare rescoring LogAUC to DOCK LogAUC

- **Training set : DUD-1**

ACE	ALR2	PNP	P38 MAP
COMT	COX-1	SAHH	AR
PDE5	GPB		$ER_{antagonist}$
FXa	HIVRT	CDK2	MR
Trypsin	InhA	FGFr1	PR

- **Validation set : DUD-2**

ADA	AmpC	PARP	TK
DHFR	COX-2	HSP90	$ER_{agonist}$
GART	HIVPR		GR
Thrombin	HMGR	EGFr	PPARg
AChE	NA	SRC	RXRa

Result

- **Geometry validation**

Scores	Native	Native / Decoy<2 Å	Decoy < 2 Å
Gscore	70	88	69
$Gscore_{random}$	73	89	69
$DrugScore_{CSD}$	77	87	66
$DrugScore_{PDB}$	49	72	65
PMF	32	52	48
PLP	52	76	70
AutoDock	8	62	66

Result

- **Enrichment validation - rescoring vs DOCK**

Protein	DOCK	Escor	Protei	DOCK	Escor	Protein	DOCK	Escor
ADA	22.7	45.8	HIVP	11.9	33.1	SRC	9.5	26.6
DHFR	18.9	62.0	HMG	40.9	35.3	TK	63.5	75.4
GART	35.3	40.0	NA	47.6	58.4			
Thrombi	29.4	22.1	PARP	8.2	40.7	ER_{agoni}	55.4	61.9
AChE	38.5	39.8	HSP9	24.6	29.6	GR	20.5	28.2
AmpC	47.4 (53.8)	10.3 (19.3)				PPARg	4.4	17.6
COX-2	40.8	19.2	EGFr	21.5	17.0	RXRa	37.9	45.1

13 out of 19, score > DOCK (1.4 Å desolvation radius)

Result

- **Enrichment validation - rescoring vs DOCK**

Protein	Rank of ligand		Protein	Rank of ligand		Protein	Rank of ligand	
	DOCK	Escor		DOCK	Escor		DOCK	Escor
ADA	2989	74	HIVP	5200	24	SRC	7536	2
DHFR	166	2	HMG	19	3	TK	319	40
GART	123	72	NA	15	1			
Thrombi	21	1	PARP	15976	292	ER_{agoni}	3	4
AChE	304	107	HSP9	2967	108	GR	9	1
AmpC	1098 (628)	27680 (14)				PPARg	16898	462
COX-2	12	74	EGFr	257	103	RXRa	1	5

AmpC A chain is broken in X-ray structure

The rank of the best ranked ligand is < 500 in all cases

04/07/11

Result

- **Enrichment validation - rescoring vs rescoring**

Escore	FLEXX	PMF	PLP	Screen	PLOP
Better	10	11	11	10	13
Equal	3	3	6	4	1
Worse	6	5	2	5	5

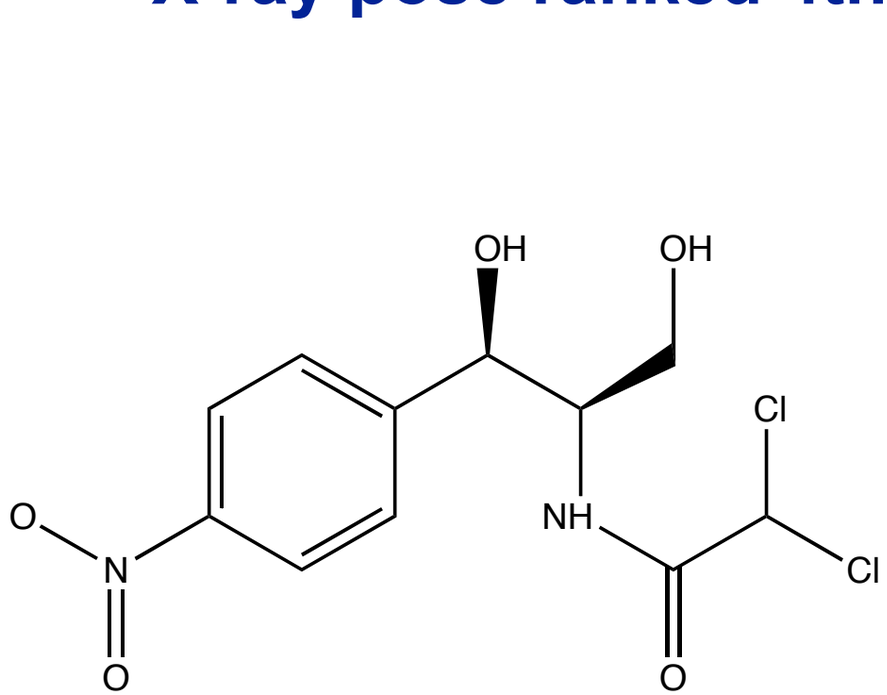
Discussion

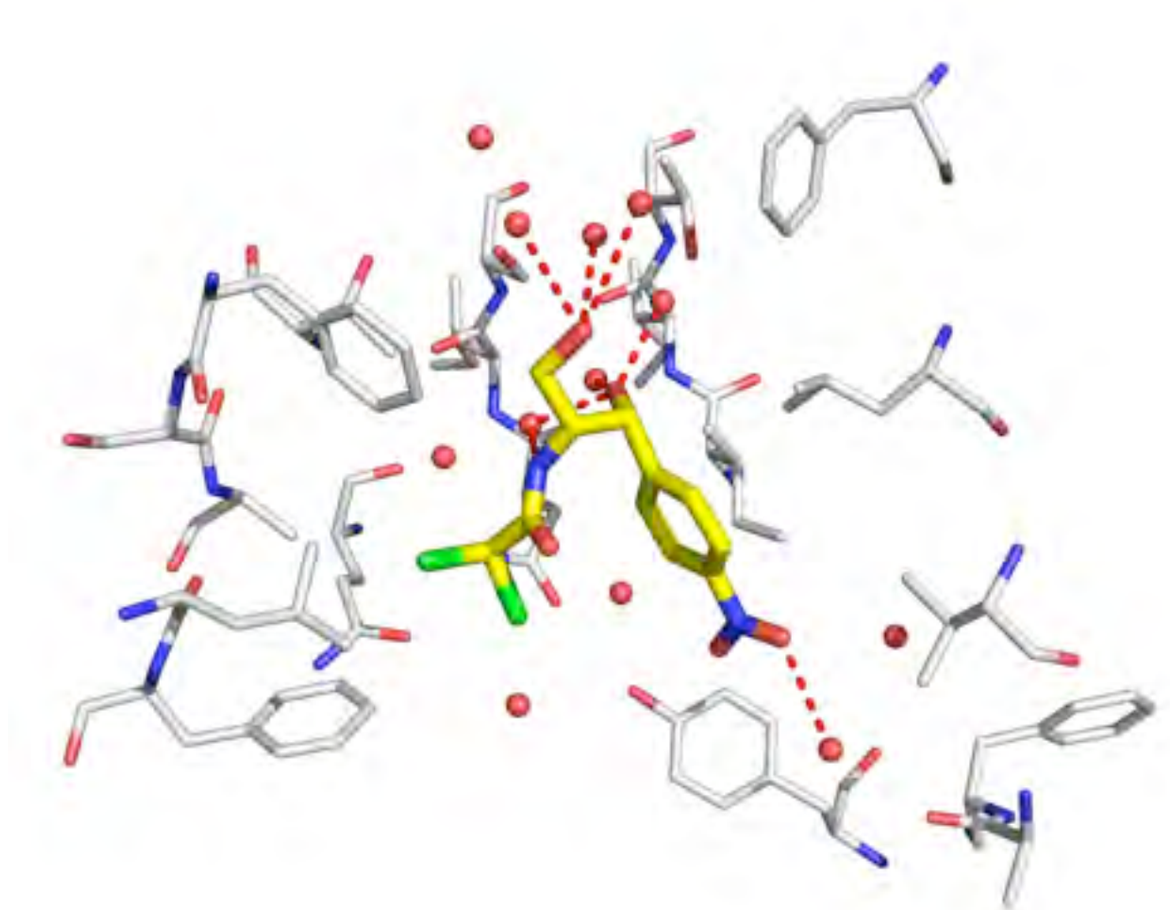
- **Geometry validation - failed cases**

  3CLA, 1.75 Å; type III chloramphenicol acetyltransferase (1CLA)

  Only H-bond with crystal water, No decoy < 2.0 Å

  X-ray pose ranked 4th





04/07/11

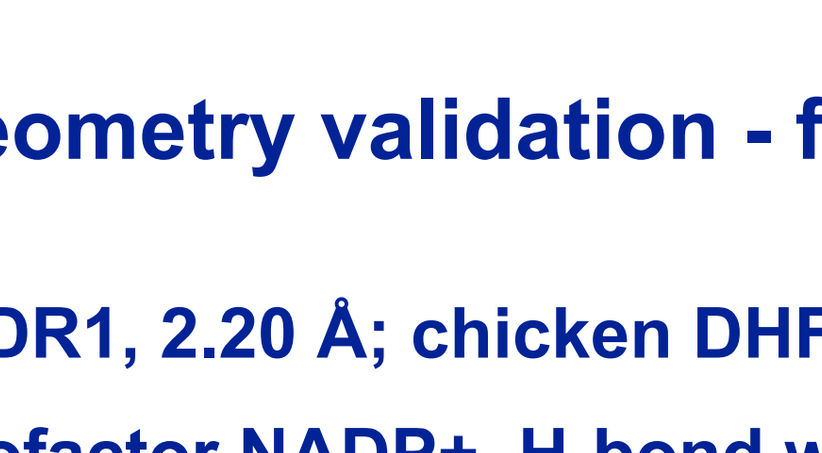
Discussion

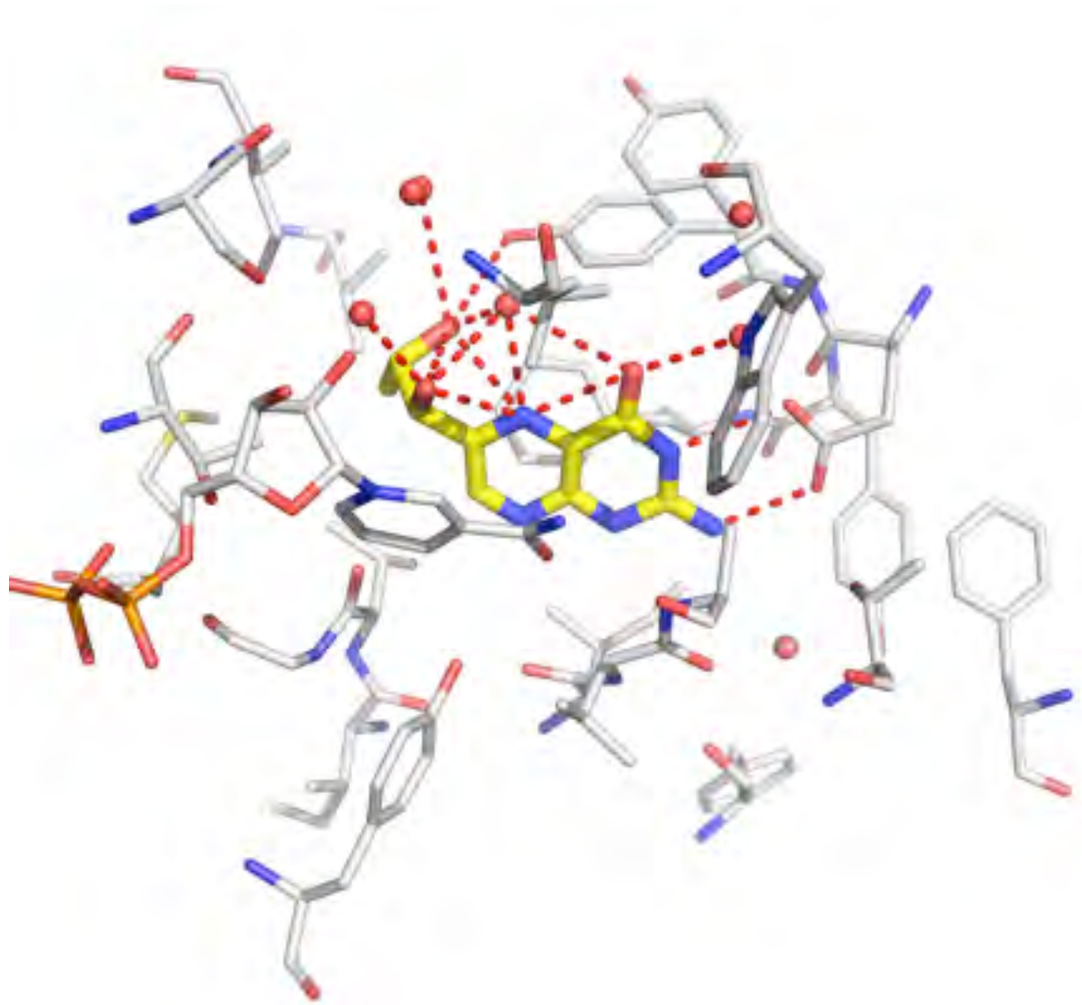
- **Geometry validation - failed cases**

1DR1, 2.20 Å; chicken DHFR

cofactor NADP+, H-bond with crystal water;

X-ray pose ranked 2nd





07/11

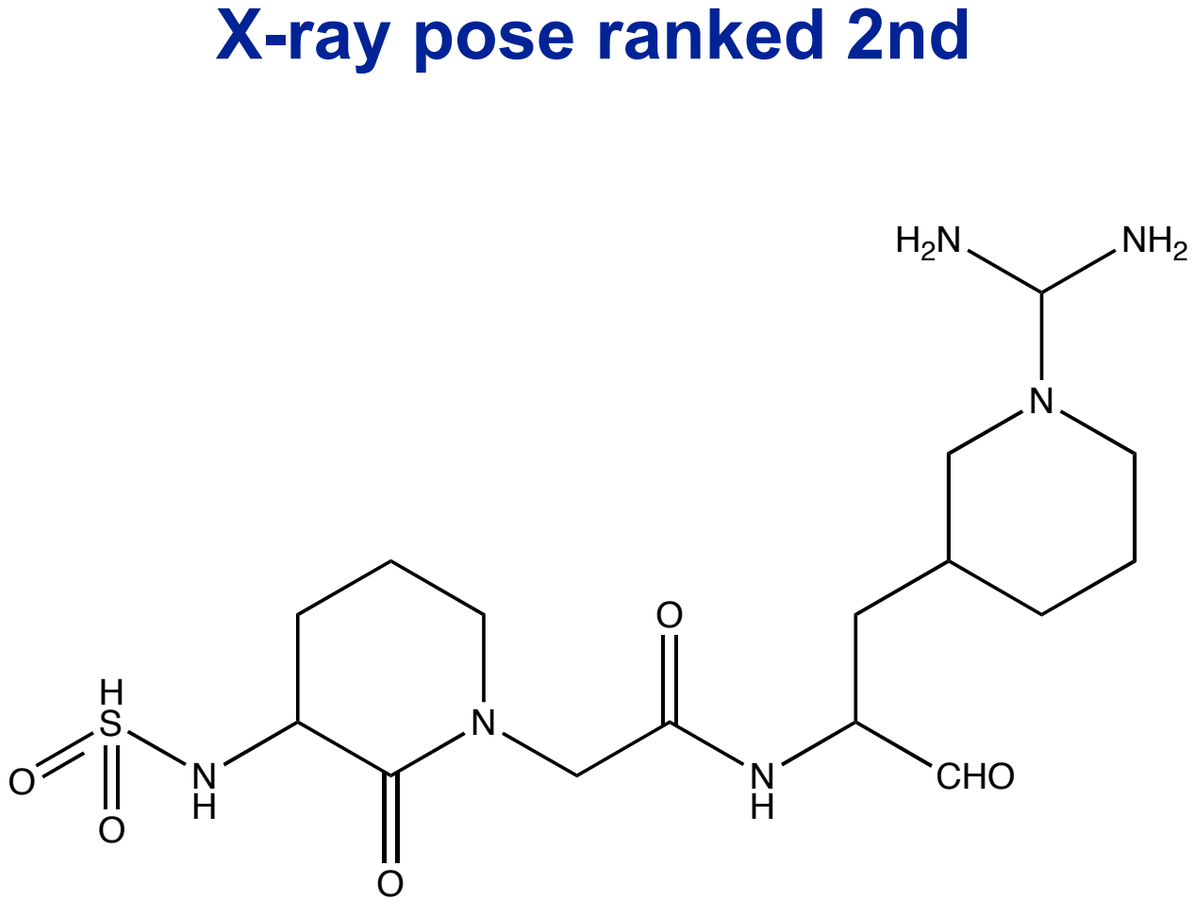
Discussion

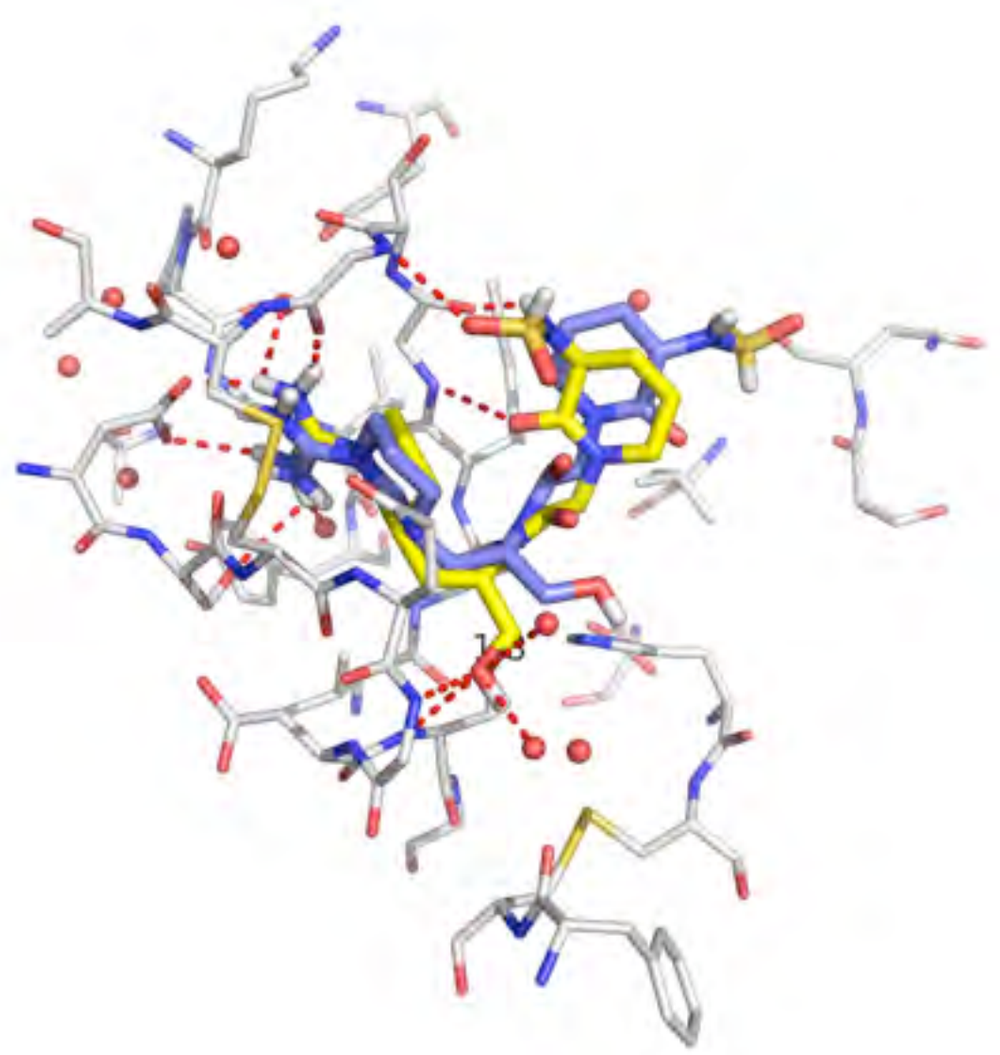
- **Geometry validation - failed cases**

1ZZZ, 1.90 Å; Thrombin

Transition state with Ser195, H-bond with crystal water;

X-ray pose ranked 2nd





04/07/11

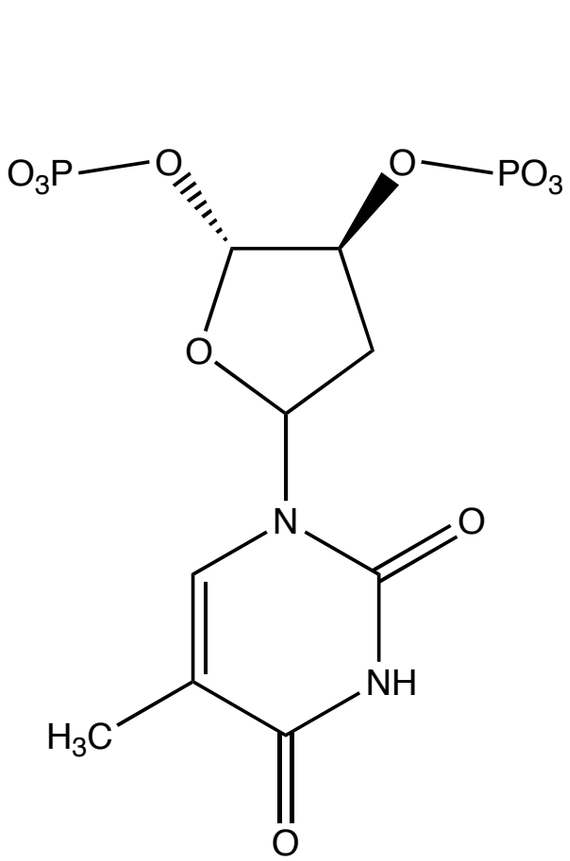
Discussion

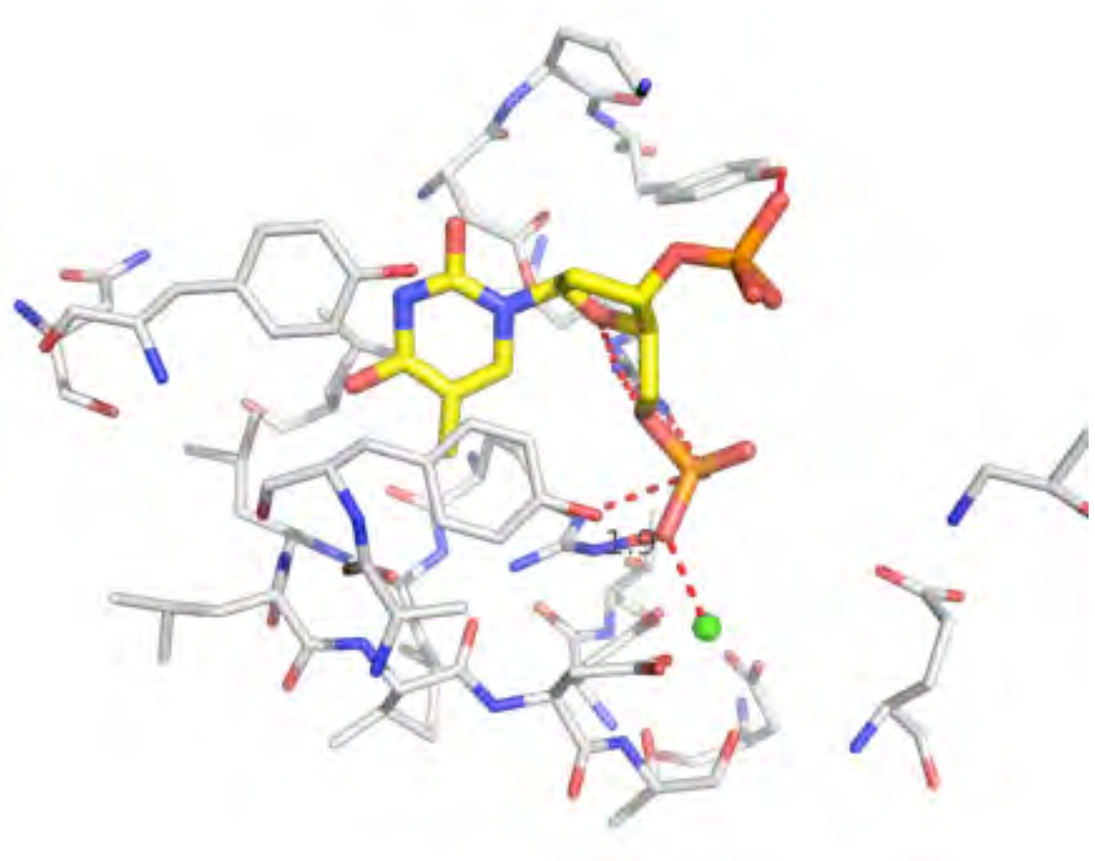
- **Geometry validation - failed cases**

  2SNS, 1.50 Å; Staphylococcal nuclease

  Close distance (1.9 Å) with Arg35 and Ca;

  Decoy < 2 Å (1.44 Å) ranked 2nd





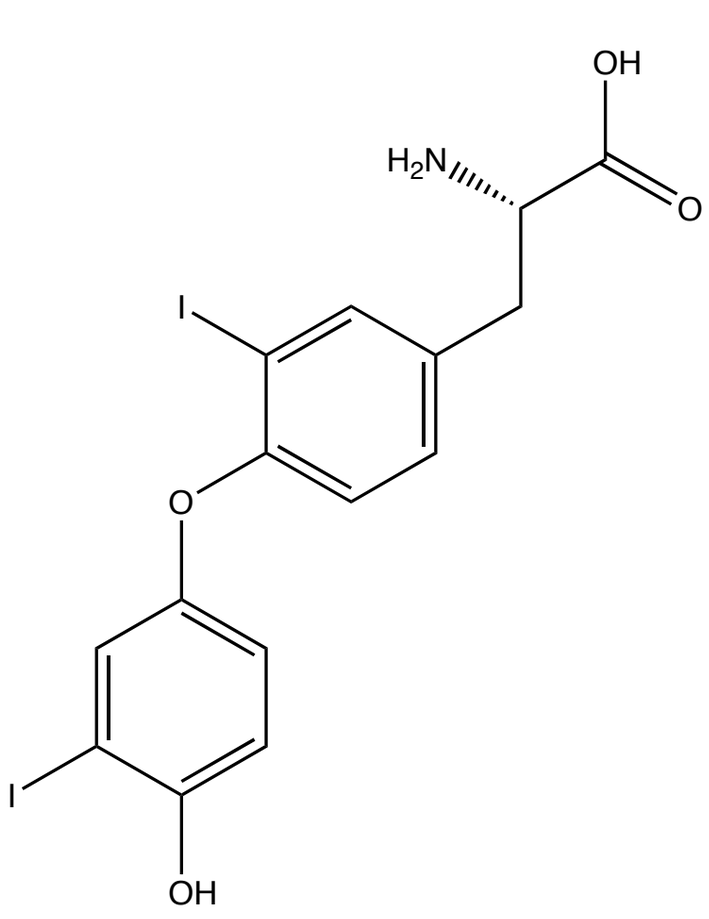
Discussion

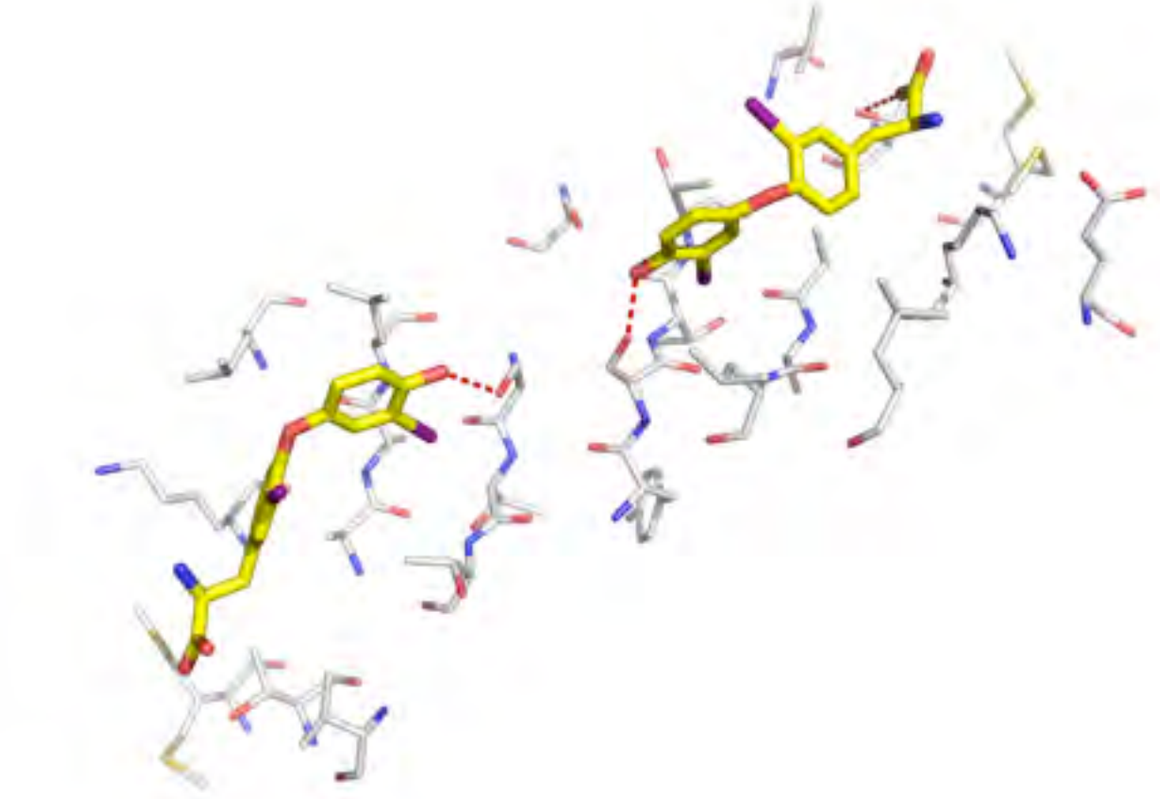
- **Geometry validation - failed cases**

  1tha, 2.00 Å; human serum transthyretin

  Protein co-crystallized with two ligands;

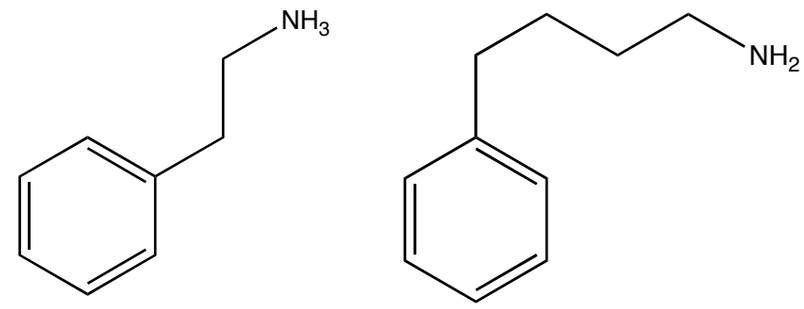
  X-ray pose ranked 3rd



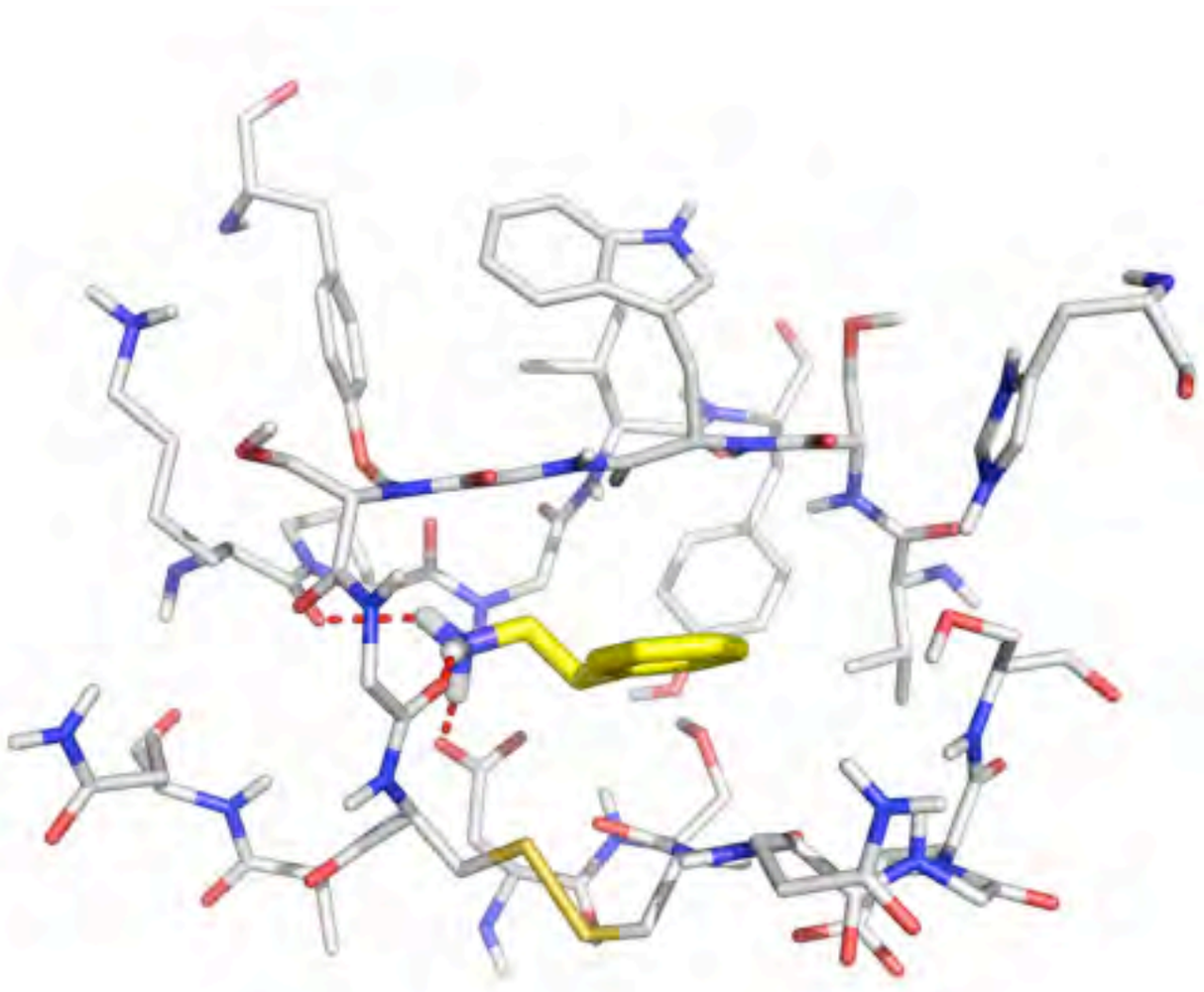


Discussion

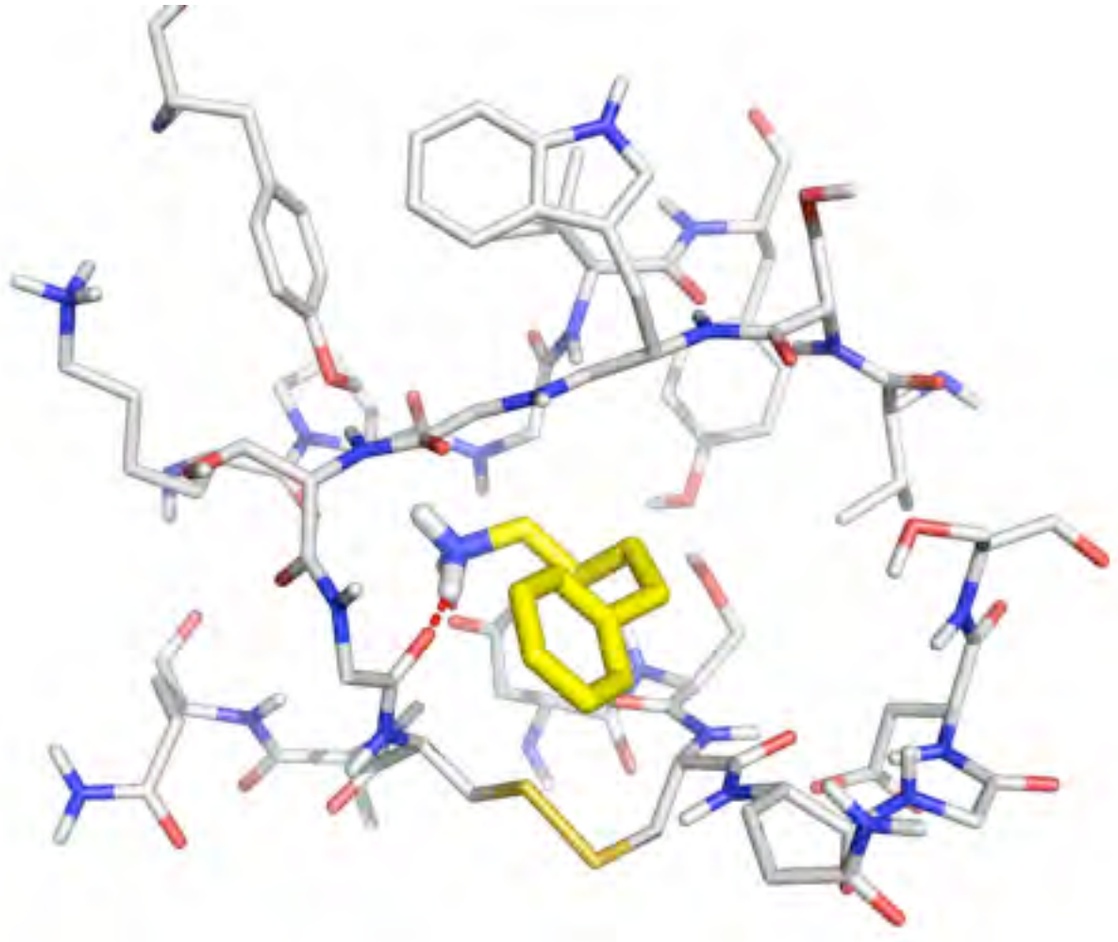
- **Geometry validation - failed cases**



1tnj, 1.80 Å; serine proteinase trypsin;
X-ray pose ranked 2nd



1tni, 1.80 Å; serine proteinase trypsin
X-ray or decoy < 2.0 Å not top 5



04/07/11

Conclusion

- When interested in maximum accuracy, it may be worth considering a statistical potential.