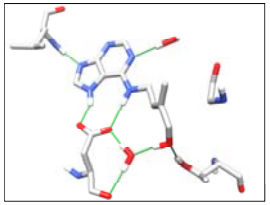
Center for Bioinformatics

The Long and Rocky Road from a PDB File to a Protein-Ligand Docking Score

Matthias Rarey
Stefan Bietz
Nadine Schneider
Sascha Urbaczek

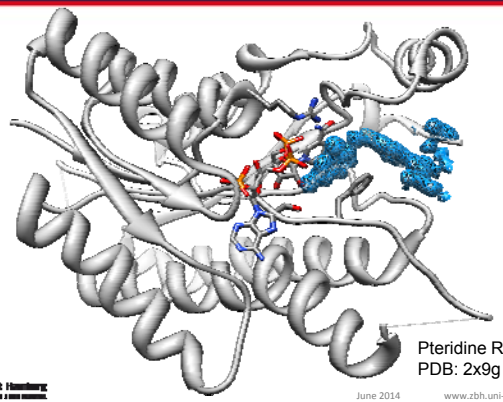
University of Hamburg
Center for Bioinformatics
Bundesstraße 43
20146 Hamburg, Germany

www.zbh.uni-hamburg.de
info@zbh.uni-hamburg.de



Center for Bioinformatics

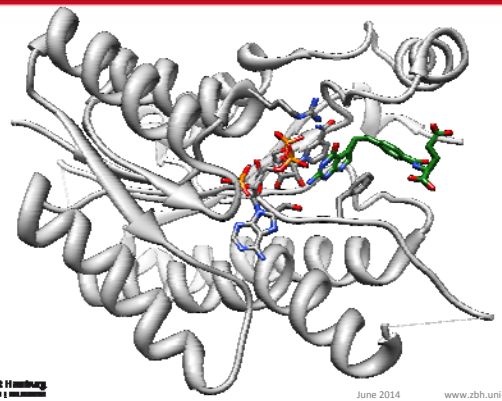
Protein Structures: The Starting Point for New Drugs



Pteridine Reductase
PDB: 2x9g

Center for Bioinformatics

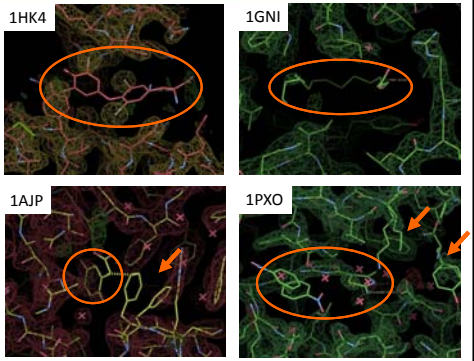
Protein Structures: The Starting Point for New Drugs



Center for Bioinformatics

Detailed Structural Analysis of PDBbind 2007 by Consulting Electron Density Maps

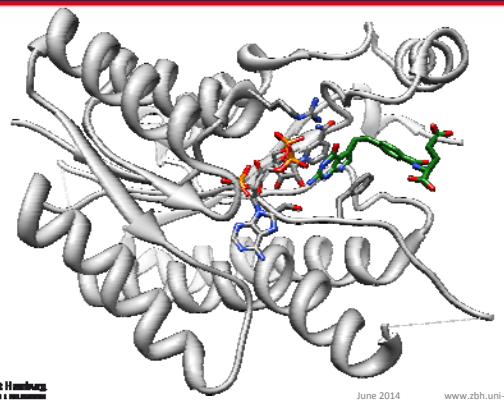
- 107 complexes with structural deficiencies
  - Missing electron density
  - Alternative conformations
  - High temperature factors
  - Contact with symmetry related subunits



Protein Structures: The Starting Point for New Drugs

Center for Bioinformatics

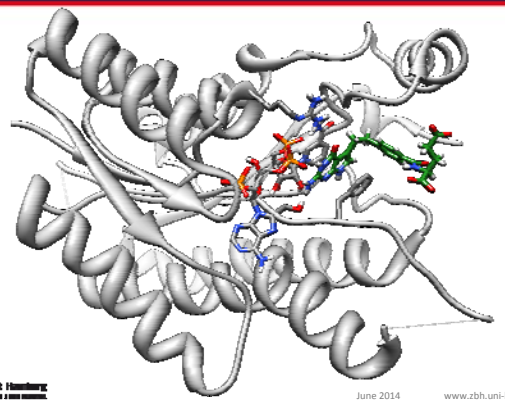
5



Protein Structures: The Starting Point for New Drugs

Center for Bioinformatics

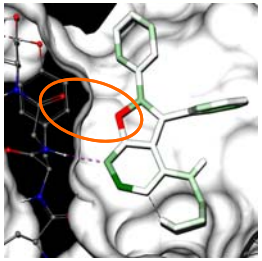
6



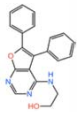
Does it Matter?

Center for Bioinformatics

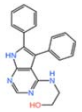
7



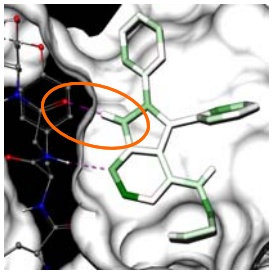
2BRB



ΔG= -28 kJ/mol



ΔG= -34 kJ/mol



* Foloppe N. et al., JMedChem 2005

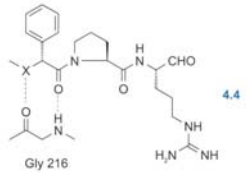
Contribution of a Single Hydrogen Bond to Binding Affinity

Center for Bioinformatics

- Example: Thrombin/Trypsin-inhibitor complexes
- Exchange NH → CH_2:
  - slight loss in IC_{50}
- Exchange NH → O:
  - loss of two hydrogen bonds upon dehydration
  - no formation of a new hydrogen bond
  - IC_{50} drops by about three orders of magnitude

Abb.: G. Klebe, Wirkstoffdesign Spektrum Akad. Verlag, 2009

Tabelle 4.4 Bindung von **4.4** an die Serinproteasen Thrombin bzw. Trypsin



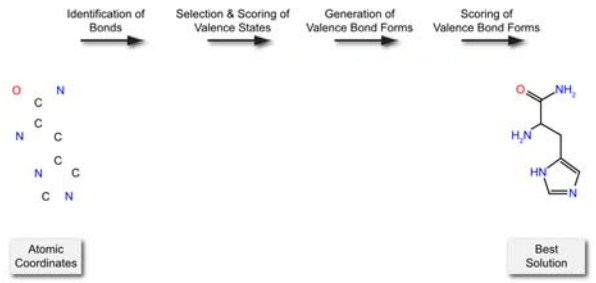
Enzym	IC_{50}-Werte in mg/ml X = -NH-	-O-	$-CH_2-$
Thrombin	0,009	52	0,07
Trypsin	0,009	43	0,018

Urbaczek et al, JCIM (2013), 53, 76-87

Center for Bioinformatics

PHASE 1: Perception of Chemical Structure

9



June 2014 www.zbh.uni-hamburg.de

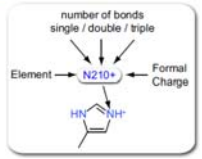
Urbaczek et al, JCIM (2013), 53, 76-87

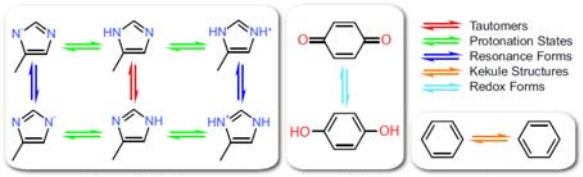
Center for Bioinformatics

Valence States

10

- Valence state: Chemically valid combination of bond orders and formal charge for a particular element.
- Aim: Find a valid valence state combination (VSC)





Forms of Histidine A | Forms of Chinon B | Kekule Structures C

June 2014 www.zbh.uni-hamburg.de

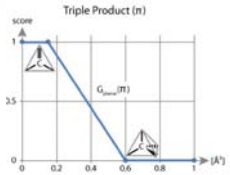
Urbaczek et al, JCIM (2013), 53, 76-87

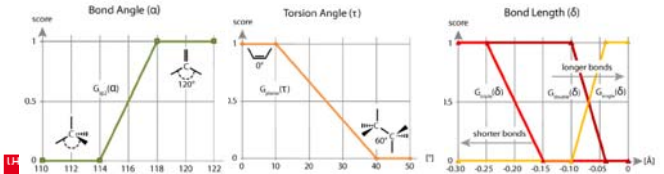
Center for Bioinformatics

Valence States

11

- Atoms get sets of all valence states compatible to the detected bond structure
- Each valence state is scored by the geometry of surrounding covalent bonds





June 2014 www.zbh.uni-hamburg.de

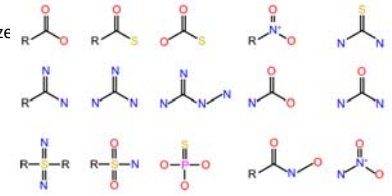
Urbaczek et al, JCIM (2013), 53, 76-87

Center for Bioinformatics

From Valence States to Valence Bond Structures

12

- Valence states get probabilities converted to scores
- Iterative refinement by removing valence states incompatible to those of neighboring atoms
- Partitioning of the molecules into zones of covalently bound atoms with yet undefined valence state
- Branch&Bound procedure enumerating and selecting highest scoring valence states
- Consider aromaticity, favorize certain functional group topologies in case of nearly equal scores



June 2014 www.zbh.uni-hamburg.de

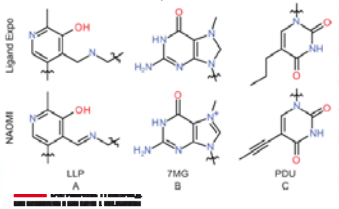
Urbaczek et al, JCIM (2013), 53, 76-87

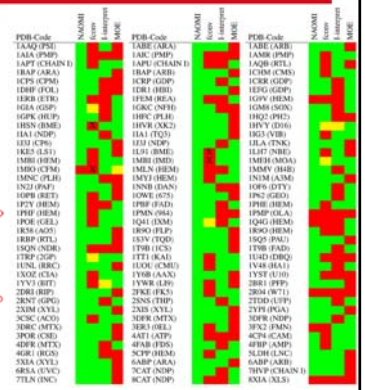
Center for Bioinformatics

Results: Quality of Chemical Structure Perception

13

- 363 PDB files with hand-curated structures, 563 ligands in total
- NAOMI vs. Ligand Expo:
  - 92% identical structures
  - differences found: (NAOMI structure validated by literature):

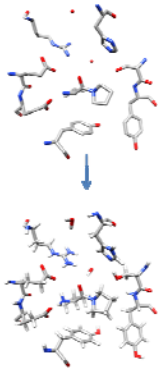




Center for Bioinformatics

PHASE 2: Prediction of Protonation

14

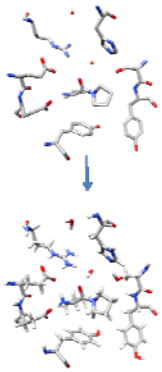


Center for Bioinformatics

Prediction of Hydrogen Coordinates

15

- Existing hydrogen placing tools:
  - WHAT IF[1]
  - Protonate3D (MOE)[2]
  - HINT Comp. Titration[3]
  - Protoss[4]
  - YASARA[5]



[1] R.W.W. Hooft *et al.* Proteins, 26(4):363–376, 1996.

[2] P. Labute *et al.* Proteins, 75(1):187, **2009**.

[3] A.S. Bayden *et al.* Journal of Computer-Aided Molecular Design, 23(9): 621–632, **2009**.

[4] **T. Lippert *et al.* Journal of Cheminformatics, 1(1):13, 2009.**

[5] E. Krieger *et al.* Computational Drug Discovery and Design, 819:405–421, **2012**.

Center for Bioinformatics

Prediction of Hydrogen Coordinates

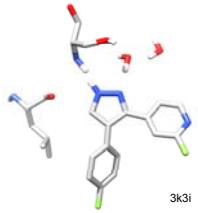
16

	proteins: rotatable hydrogens	proteins: tautomers	proteins: protonation states	proteins: side chain flips	proteins: water orientations	ligands: rotatable hydrogens	ligands: tautomers	ligands: protonation states	ligands: side chain flips
WHAT IF	✓	✓	✓	✓	✓				
Protonate3D	✓	✓	✓	✓	✓	✓	(✓)	(✓)	
HINT CT	✓	✓	✓	✓	✓	✓	(✓)	(✓)	
Protoss v2009	✓	✓	✓	✓	✓	✓	(✓)	(✓)	
YASARA	✓	✓	✓	✓	✓	✓	(✓)	(✓)	✓

Error Case Examples

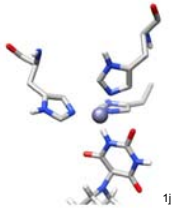
Center for Bioinformatics

17



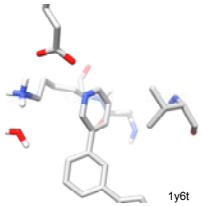
Repulsive donor interactions

YASARA



Hydrogen-metal clashes

Protonate3D



Unsaturated donors / acceptors

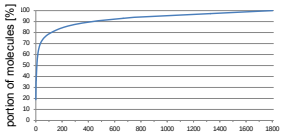
Protoss v2009

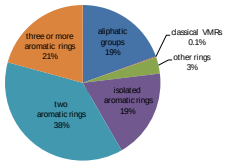
Frequency of Tautomerism

Center for Bioinformatics

18

- Ligand Expo[6]: 17,563 small molecules[7]
- According to the ProToss protonation model
  - 81 % exhibit alternative tautomers or / and protonation states
  - 17 % only contain commonly used groups (primary amines, carboxylats, imidazoles)
  - Only 19 % do not show variability
  - 1802 structurally different variable mode regions (VMRs)





different VMR type portions

[6] Z. Feng *et al.* Bioinformatics, 20(13):2153-5, 2004
[7] Downloaded from http://ligand-expo.rcsb.org (01/03/2014)

Bietz et al, JCheminf (2014), 6 (12):1-12

Challenges to an Advancement of ProToss

Center for Bioinformatics

19

- Find the tautomer / protonation state that leads to the best hydrogen bonding network in the protein-ligand complex
- Keep the runtime low

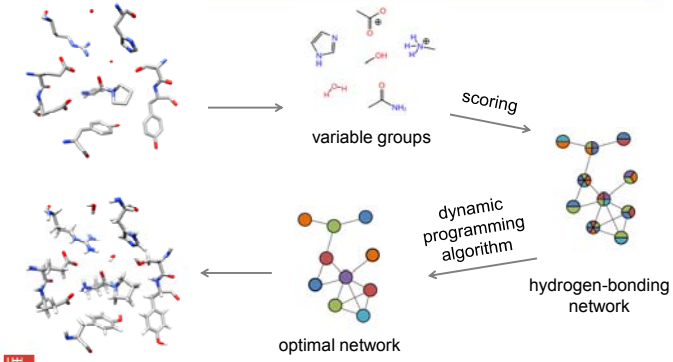
	protein: rotatable hydrogens	protein: tautomers	protein: protonation states	protein: side chain flips	protein: water orientations	ligands: rotatable hydrogens	ligands: tautomers	ligands: protonation states	ligands: side chain flips
WHAT IF	✓	✓	✓	✓	✓				
Protonate3D	✓	✓	✓	✓	✓	✓	(✓)	(✓)	
HINT CT	✓	✓	✓	✓	✓	✓	(✓)	(✓)	
Protoss v2009	✓	✓	✓	✓	✓	✓	(✓)	(✓)	
YASARA	✓	✓	✓	✓	✓	✓	(✓)	(✓)	✓
Protoss v2013	✓	✓	✓	✓	✓	✓	✓	✓	(✓)

Bietz et al, JCheminf (2014), 6 (12):1-12

Protoss Workflow

Center for Bioinformatics

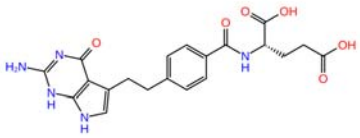
20



Center for Bioinformatics

Generation of Tautomers and Protonation States

21



Pemetrexed (PTR1 inhibitor)

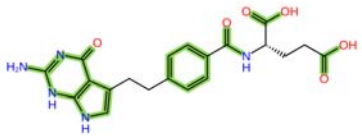
Urbaczek et al, J Chem Inf Model (2014), 45 (3):756-766

Center for Bioinformatics

Generation of Tautomers and Protonation States

22

Step 1: Conjugated zone identification



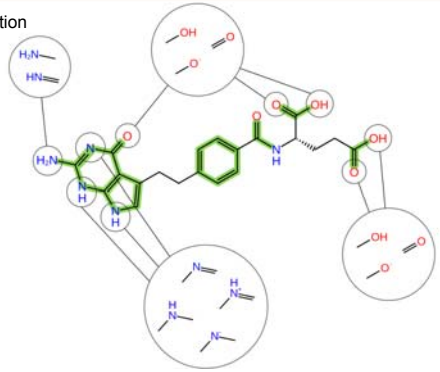
Urbaczek et al, J Chem Inf Model (2014), 45 (3):756-766

Center for Bioinformatics

Generation of Tautomers and Protonation States

23

Step 2: Atom state selection



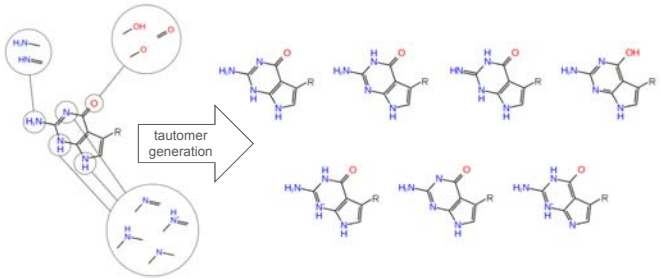
Urbaczek et al, J Chem Inf Model (2014), 45 (3):756-766

Center for Bioinformatics

Generation of Tautomers and Protonation States

24

Step 3: Enumeration of tautomers and protonation states



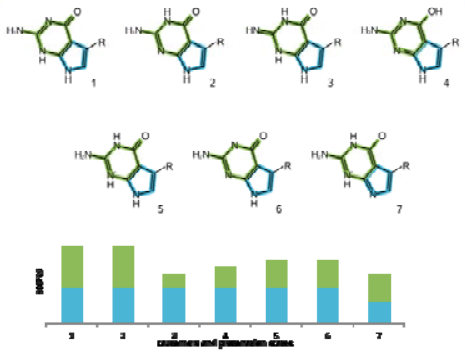
Urbaczek et al, J Chem Inf Model (2014), 45 (3):756-766

Center for Bioinformatics

Generation of Tautomers and Protonation States

25

Step 4: State scoring



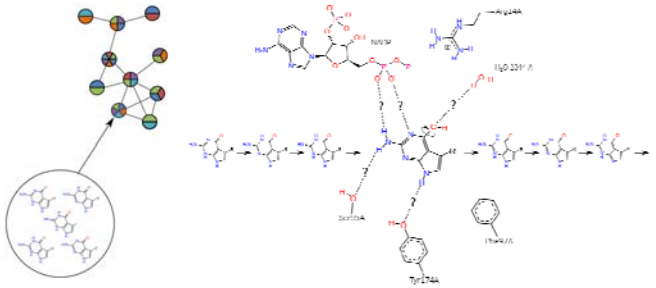
Bietz et al, JCheminf (2014), 6 (12):1-12

Center for Bioinformatics

ProToss Workflow Extended

26

Step 5: Inclusion into ProToss' optimization procedure



Bietz et al, JCheminf (2014), 6 (12):1-12

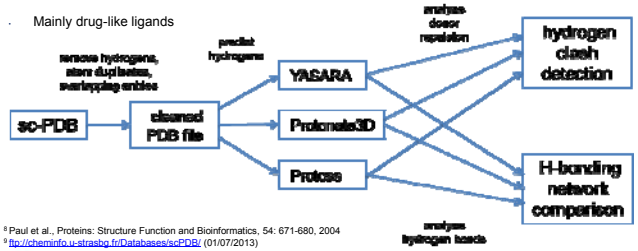
Center for Bioinformatics

Results: Comparison on the sc-PDB

27

Dataset: sc-PDB[8-9]

- 8077 protein-ligand complexes (37 with errors excluded)
- Druggable binding-sites
- Mainly drug-like ligands



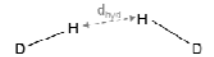
[8] Paul et al., Proteins: Structure Function and Bioinformatics, 54: 671-680, 2004
[9] ftp://cheminfo.u-strasbg.fr/Databases/scPDB/ (01/07/2013)

Bietz et al, JCheminf (2014), 6 (12):1-12

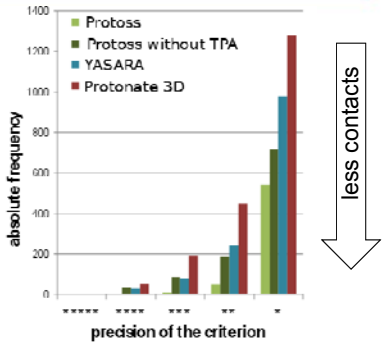
Center for Bioinformatics

Undesirable Donor Contacts in the Protein-Ligand Interface

28



precision	d_{hyd} [Å]
*****	0.7
****	1.0
***	1.3
**	1.6
*	1.9

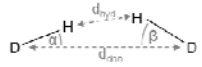


Bietz et al, JCheminf (2014), 6 (12):1-12

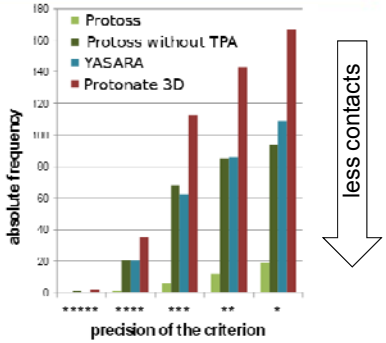
Center for Bioinformatics

Undesirable Donor Contacts in the Protein-Ligand Interface

29



presision	d_{hyd} [A]	d_{don} [A]	α / β [°]
*****	0.7	2.6	30
****	1.0	2.9	30
***	1.3	3.2	30
**	1.6	3.5	30
*	1.9	3.8	30

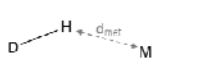


Bietz et al, JCheminf (2014), 6 (12):1-12

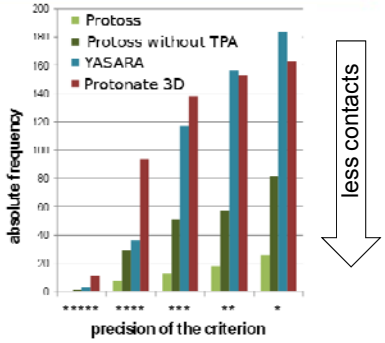
Center for Bioinformatics

Undesirable Donor-Metal Contacts in the Protein-Ligand Interface

30



precision	d_{met} [A]
*****	1.0
****	1.3
***	1.6
**	1.9
*	2.2

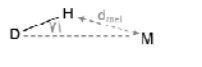


Bietz et al, JCheminf (2014), 6 (12):1-12

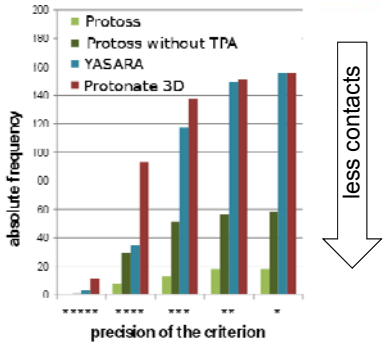
Center for Bioinformatics

Undesirable Donor-Metal Contacts in the Protein-Ligand Interface

31



precision	d_{met} [A]	γ [°]
*****	1.0	60
****	1.3	60
***	1.6	60
**	1.9	60
*	2.1	60

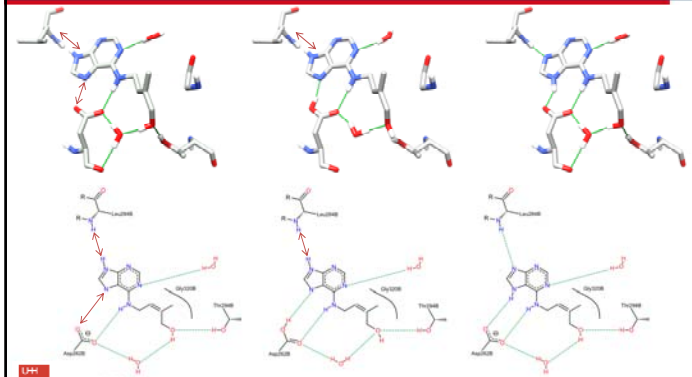


Bietz et al, JCheminf (2014), 6 (12):1-12

Center for Bioinformatics

Undesirable Contacts in the Protein-Ligand Interface

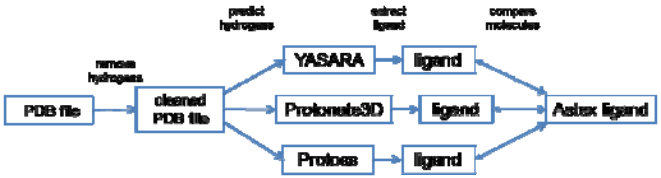
32



Astex Diverse Set[9]: Protonation Comparison

Center for Bioinformatics

33

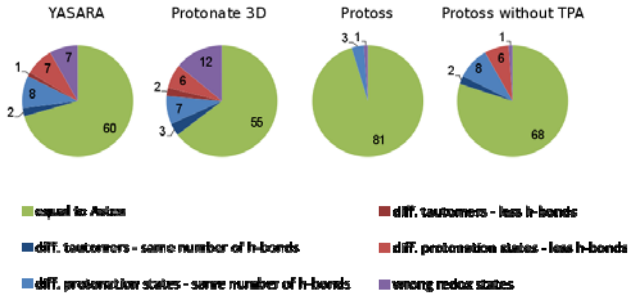


[9] Hartshorn et al., J.Med.Chem, 50(4): 726-741, 2007

Astex Diverse Set: Protonation Comparison

Center for Bioinformatics

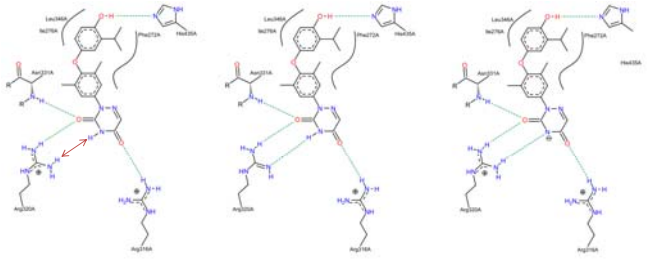
34



Astex Diverse Set: Protonation Comparison

Center for Bioinformatics

35

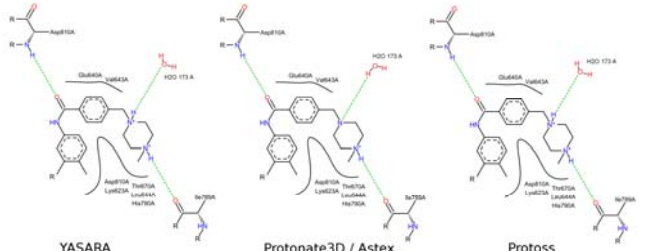


PDB code: 1n46

Astex Diverse Set: Protonation Comparison

Center for Bioinformatics

36

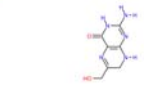


PDB code: 1t46

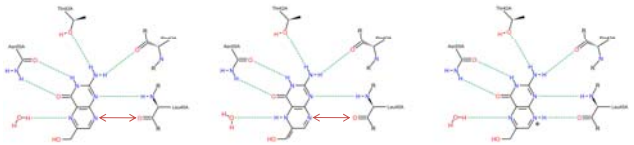
Astex Diverse Set: Protonation Comparison

Center for Bioinformatics

37



6-hydroxymethyl-7,8-dihydropterin



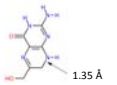
YASARA Protonate3D Protoss

PDB code: 1hq2

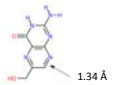
Astex Diverse Set: Protonation Comparison

Center for Bioinformatics

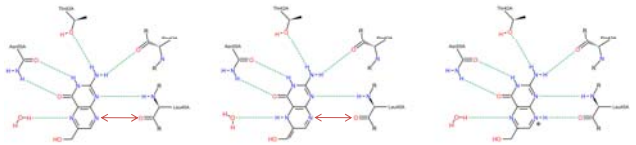
38



6-hydroxymethyl-7,8-dihydropterin (1hq2)



6-hydroxymethylpterin (3ip0)



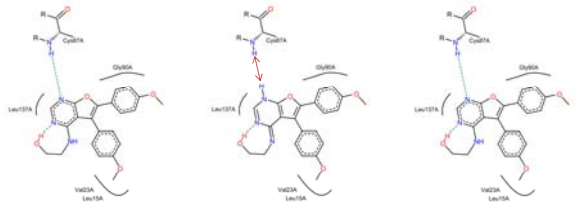
YASARA Protonate3D Protoss

PDB code: 1hq2

Astex Diverse Set: Protonation Comparison

Center for Bioinformatics

39



YASARA / Astex Protonate3D Protoss / Astex

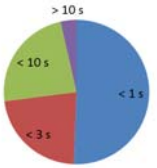
PDB code: 2br1

ProToss Runtime Evaluation

Center for Bioinformatics

40

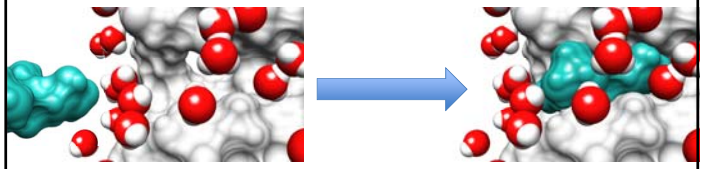
- Whole complex optimization on the Astex Diverse Set
  (including all ligands, co-factors and water molecules)
- Runtimes include
  - File IO
  - Complex preprocessing
  - Network optimization
- Intel® Core™ i7-2600 CPU (3.4 GHz)
  and 8 GB memory
- Mean: 2.47 s
- Median 0.93 s



PHASE 3: Scoring Protein-Ligand Complexes

Center for Bioinformatics

41



Schneider et al, *J. Comput. Aided Mol. Des.*, **27(1)**, 15-29 (2013)

HYDE – Towards a Consistent Description of Hydrogen Bonding, Dehydration and the Hydrophobic Effect

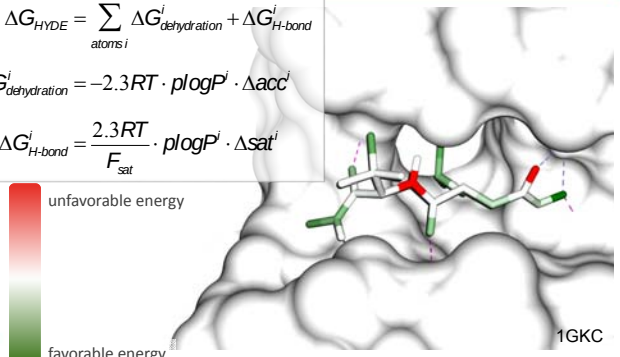
Center for Bioinformatics

42

$$\Delta G_{HYDE} = \sum_{atoms\ i} \Delta G^{i}_{dehydration} + \Delta G^{i}_{H\text{-}bond}$$

$$\Delta G^{i}_{dehydration} = -2.3RT \cdot plogP^{i} \cdot \Delta acc^{i}$$

$$\Delta G^{i}_{H\text{-}bond} = \frac{2.3RT}{F_{sat}} \cdot plogP^{i} \cdot \Delta sat^{i}$$



Schneider et al, J.Comput.-Aided Mol. Des. **26** 2011: 701-723

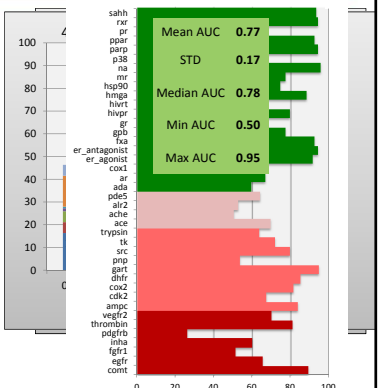
HYDE Benchmarks

Center for Bioinformatics

43

- PDBbind 2007 coreset[1] ➔ Binding affinity prediction
- Astex diverse set[2] ➔ Redocking experiments
- Dataset of Useful Decoy[3] ➔ Virtual screening study

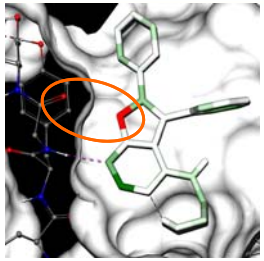
[1] Wang R. et al, JMedChem, 2004, 2005
[2] Hartshorn M. J. et al , JMedChem, 2007
[3] Huang N. et al, JMedChem, 2006
[4] Schneider N. et al, JCAMD, 2011



Estimating the Cost of a Single Hydrogen Bond

Center for Bioinformatics

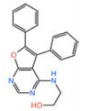
44

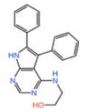


2BRB

ΔG= -28 kJ/mol

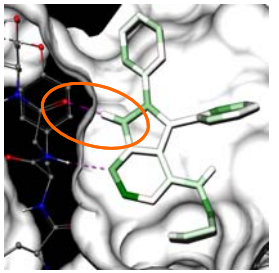
ΔG_{HYDE} = -22 kJ/mol





ΔG_{HYDE} = -32 kJ/mol

ΔG= -34 kJ/mol



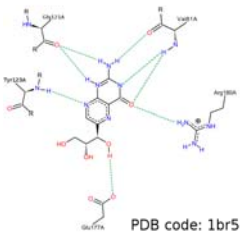
* Foloppe N. et al., JMedChem 2005

A Few Things to Remember

Center for Bioinformatics

45

- PDB files are models – consult electron density.
- Perception of chemical structures from PDB is difficult – cross-check with literature.
- Assigning protonation and tautomeric states is not simpler
  - cross-check the results of automatic assignment procedures
- Tools like ProToss help in large scale applications in which manual curation is not possible.



Contributions / Acknowledgements

Center for Bioinformatics

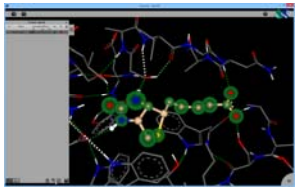
46

- **Stefan Bietz**
  ProToss development
- **Sascha Urbaczek**
  Tautomer generation
- **Tobias Lippert**
  First version of ProToss
- **Nadine Schneider**
  Hyde development
- **Benjamin Schulz**
  ProToss development, testing and validation
- **Holger Claussen**
  Testing, testing …
- **Christian Lemmen**

Software Availability

Center for Bioinformatics

47



More on ProToss
http://protoss.zbh.uni-hamburg.de/
ProToss in action:
http://www.biosolveit.de/SeeSAR/

More on tools and servers from ZBH:
http://www.zbh.uni-hamburg.de/AMD
http://www.biosolveit.de/techtransfer