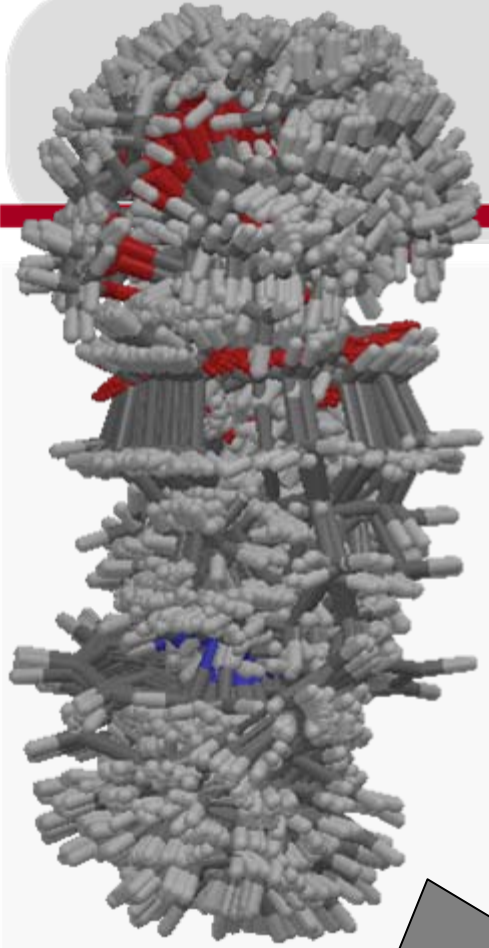
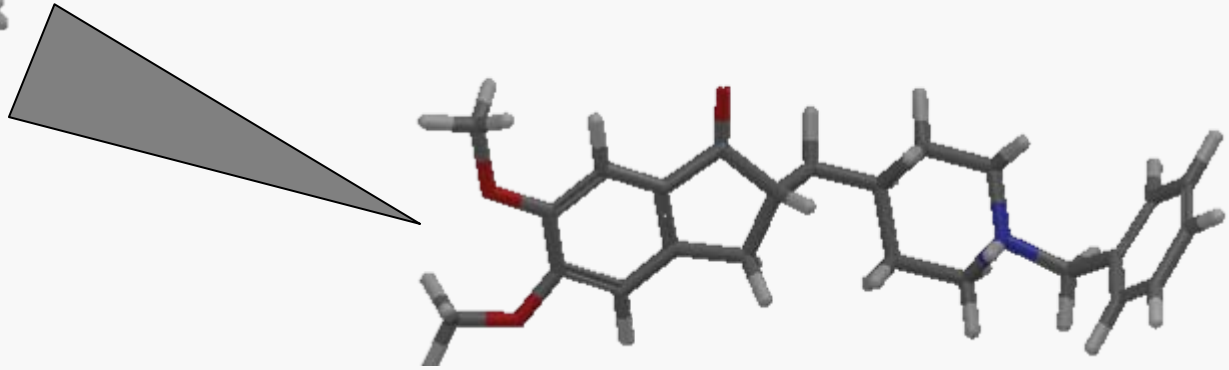


Focusing Conformational Ensembles on Bioactive-Like Conformations

Hannah H. Avgy, Boaz Musafia, Hanoch Senderowitz

Department of Chemistry

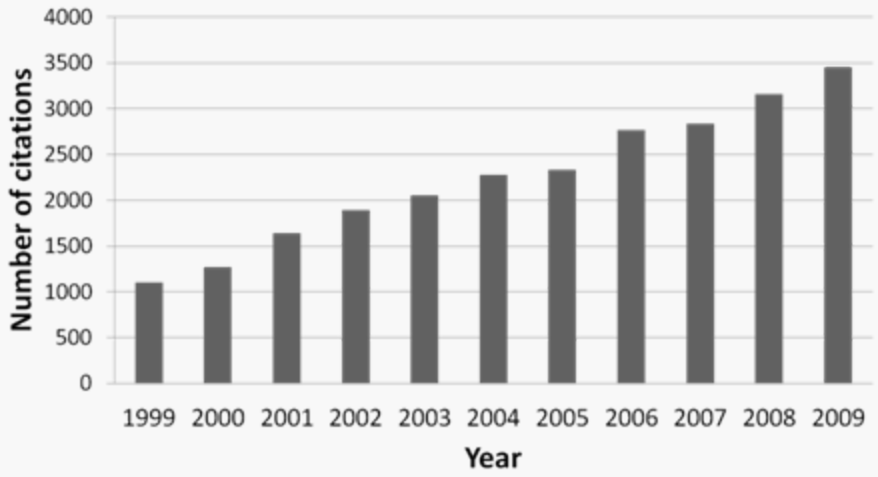
Bar-Ilan University



3rd Strasbourg Summer School on Chemoinformatics, Strasbourg, France, 25-29 June 2012

Presentation Outline

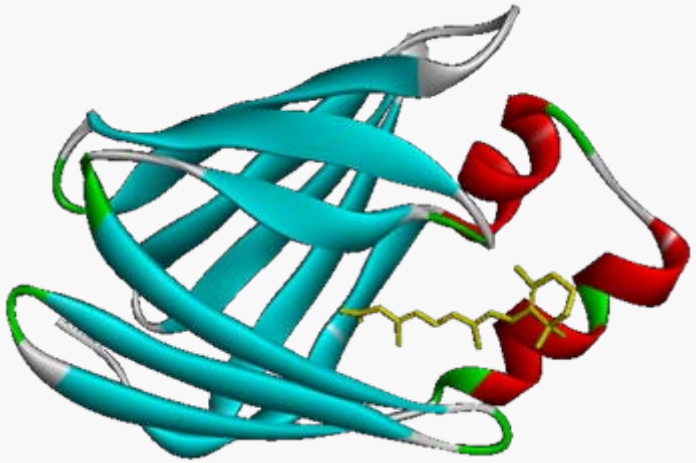
- Bioactive conformations: Definition and importance
- Experimental and computational identification of bioactive conformations
- Challenges in identifying and scoring bioactive conformations
- Bioactive conformational biasing
- Bioactive conformations of drugs: Preliminary analysis
- Conclusions and future directions



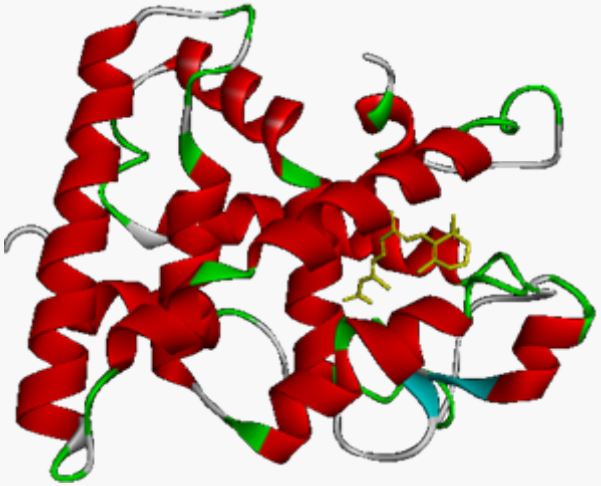
Bioactive Conformations

- The conformation adopted by a compound when bound to its bio-target
- The conformation responsible for the biological activity
- Bioactive conformations are target-dependent

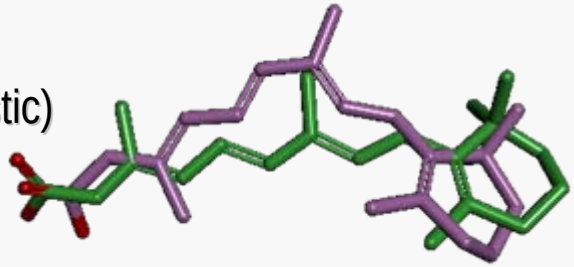
Retinoic acid binding protein



Nuclear receptor

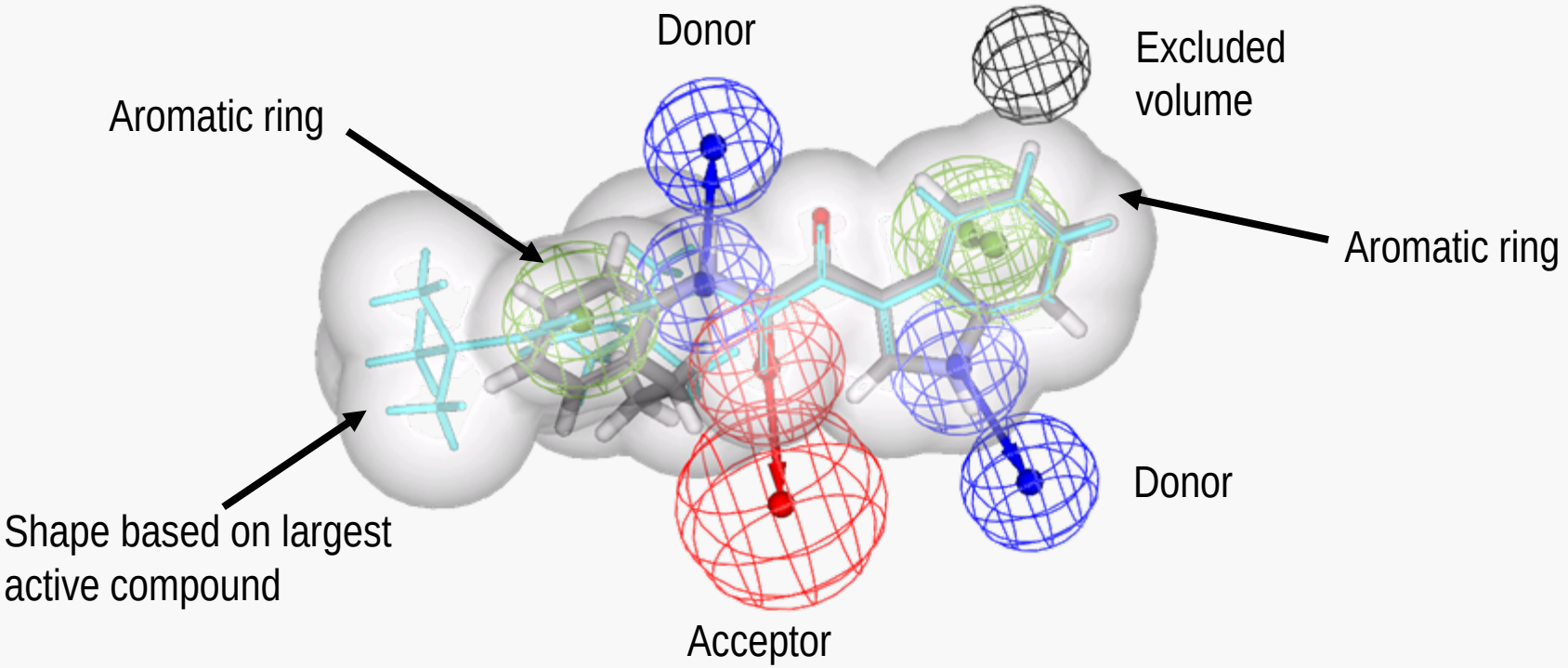


Aliretinoin (antineoplastic)
RMSD = 1.6Å



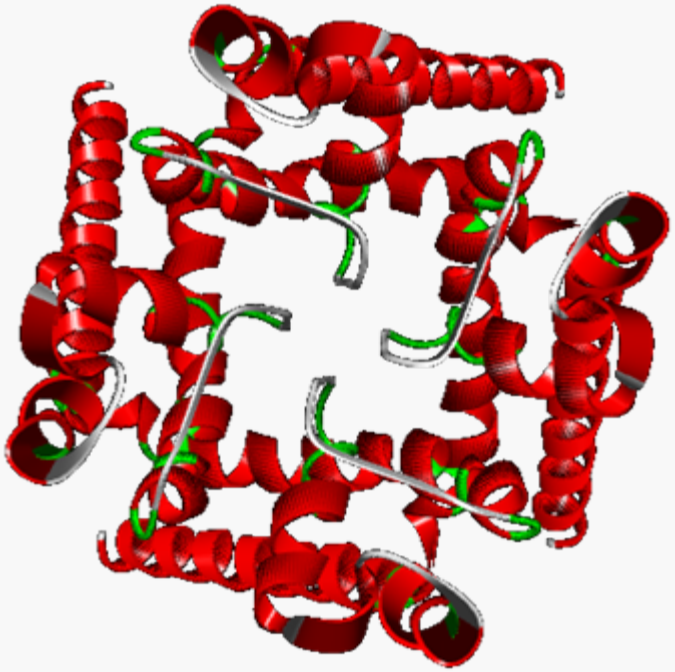
Ligand-Based Drug Design: Pharmacophore

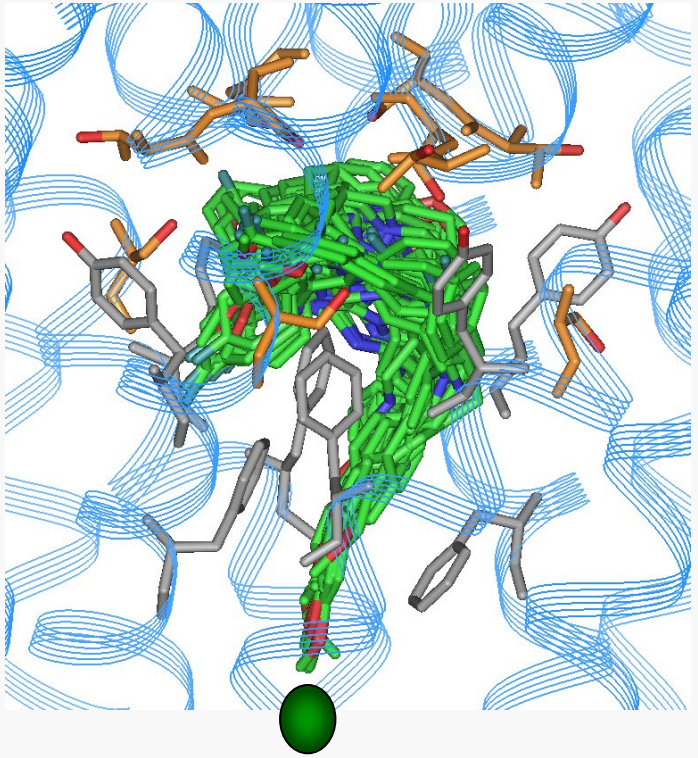
- Pharnacophore: A 3D arrangement of function groups which is responsible for the biological activity
  - Obtained by the superposition of active (and inactive) compounds
  - Assumption: Compounds represented by their bioactvie conformations



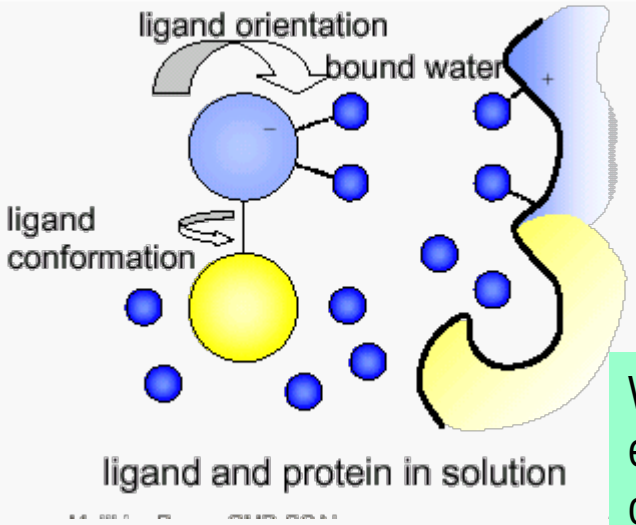
Target-Based Drug Design: Docking

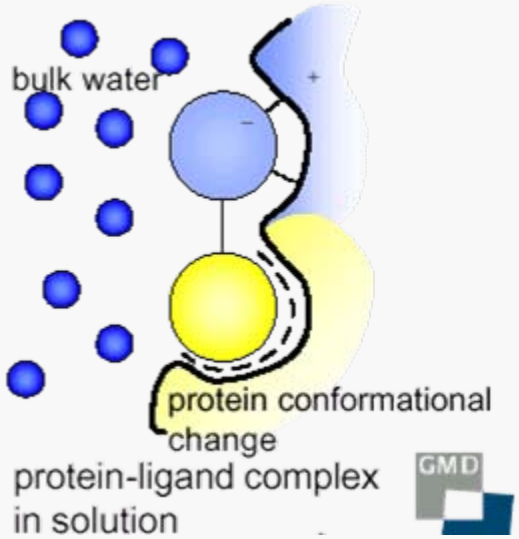
- Determine the most probable binding mode
- Approximate binding free energy
- Knowledge of bioactive conformation can eliminate erroneous binding modes





Target-Based Drug Design: Scoring





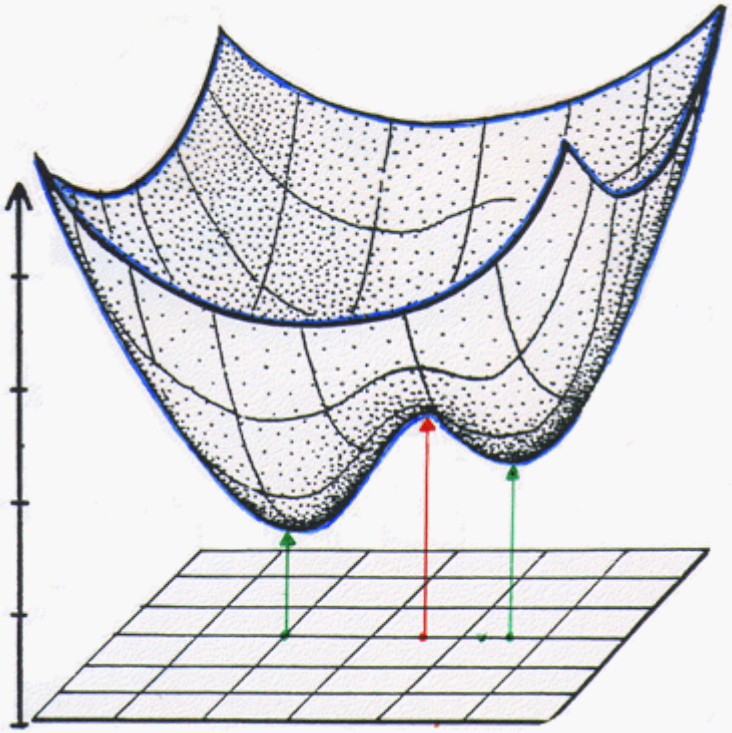
Contribution of Conformer Focusing to the Uncertainty in Predicting Free Energies for Protein-Ligand Binding (Tirado-Rives and Jorgensen, J. Med. Chem., 2006, 49, 5880-5884)

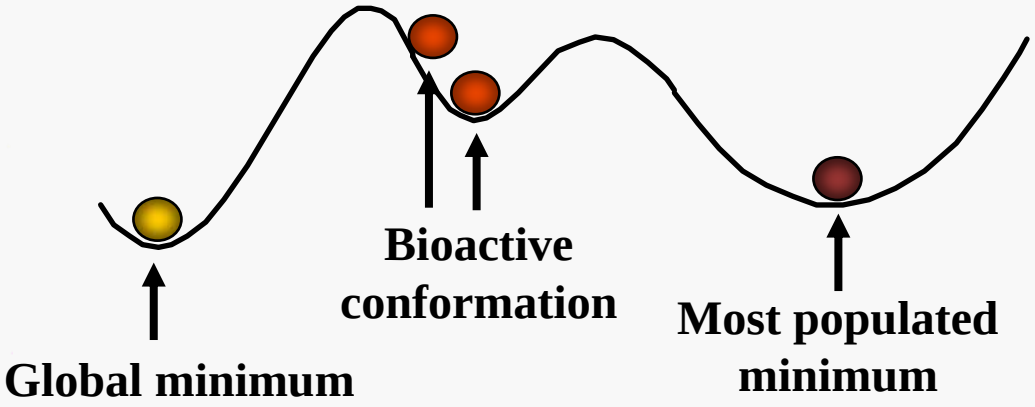
$$\Delta G_{cf} = (\varepsilon_k - \varepsilon_1) + \beta^{-1} \ln\left[\sum_i n_i \exp(-\beta(\varepsilon_i - \varepsilon_1))\right]$$

When a ligand binds to a protein, it is typically not in the lowest-energy conformation for the unbound ligand and there is also a loss of conformational degrees of freedom. The free-energy change for this "conformer focusing" is addressed here formally, and the associated errors with its estimation or neglect are considered in the context of scoring functions for protein-ligand docking and computation of absolute free energies of binding. Specific applications for inhibition of HIV-1 reverse transcriptase are reported. It is concluded that the uncertainties from this source alone are sufficient to preclude the viability of current docking methodology for rank-ordering of diverse compounds in high-throughput virtual screening.

Potential Energy Surface (PES)

- An *N* atoms ligand is defined by *3N* Cartesian coordinates or *3N-6* internal coordinates





$$\frac{N_i}{N_{total}} = \frac{\exp\left[-E_j / kT\right]}{\sum_i \exp\left[-E_i / kT\right]}$$

Experimental Sources for Bioactive Conformations

- Solid state (X-ray)
  - PDB (82,522 structure, 24,517 complexes (“has ligand” AND “300 < MW < 800”))
  - CSD (596,810 entries; January 1st 2012)
  - Potentially subjected to crystal packing forces
  - Represent a single conformer
- Solution (NMR)
  - Analysis complicated by multiple conformations
  - Only few studies

Methotrexate

Foloppe and Chen, Curr. Med. Chem. 2009, 16, 3381

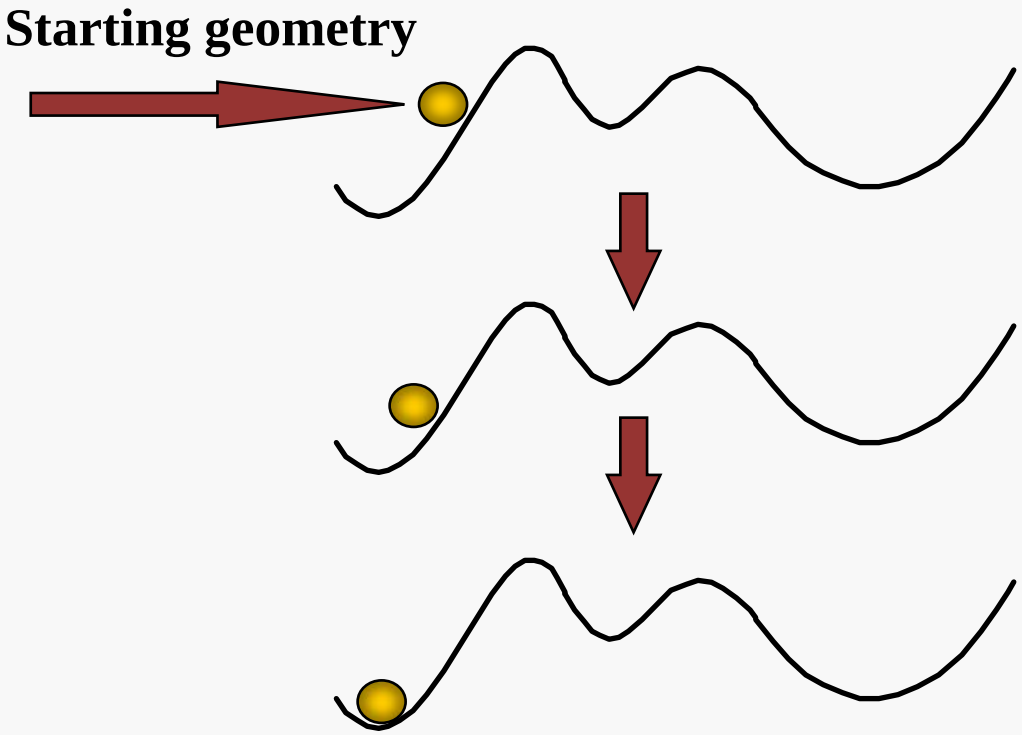
Computational Derivation of Bioactive Conformations

- Approximate the true PES
  - Force Field:

$$V(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$
$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) + \text{cross terms}$$

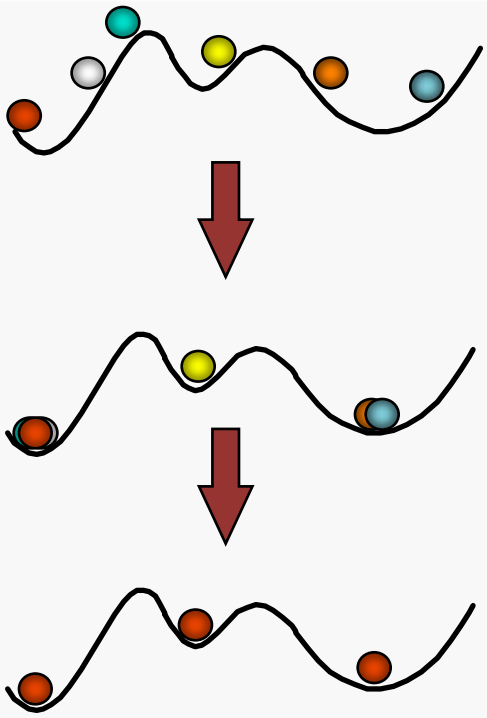
  - QM: $H\Psi = E\Psi$
  - QM/MM
- Sample the PES
  - Minimization
  - Conformational search
  - Molecular dynamics (MD)
  - Monte Carlo (MC)

Energy Minimization



- Depends on starting geometry
- Can only go down hill

Conformational Search



- Randomly or systematically generated starting geometries
- Energy minimization
- Duplicates elimination
- Representative structures for each potential minimum

- Gives coverage of potential surface.
- Combinatorial growth.
- Resulting ensemble reflects enthalpy only.

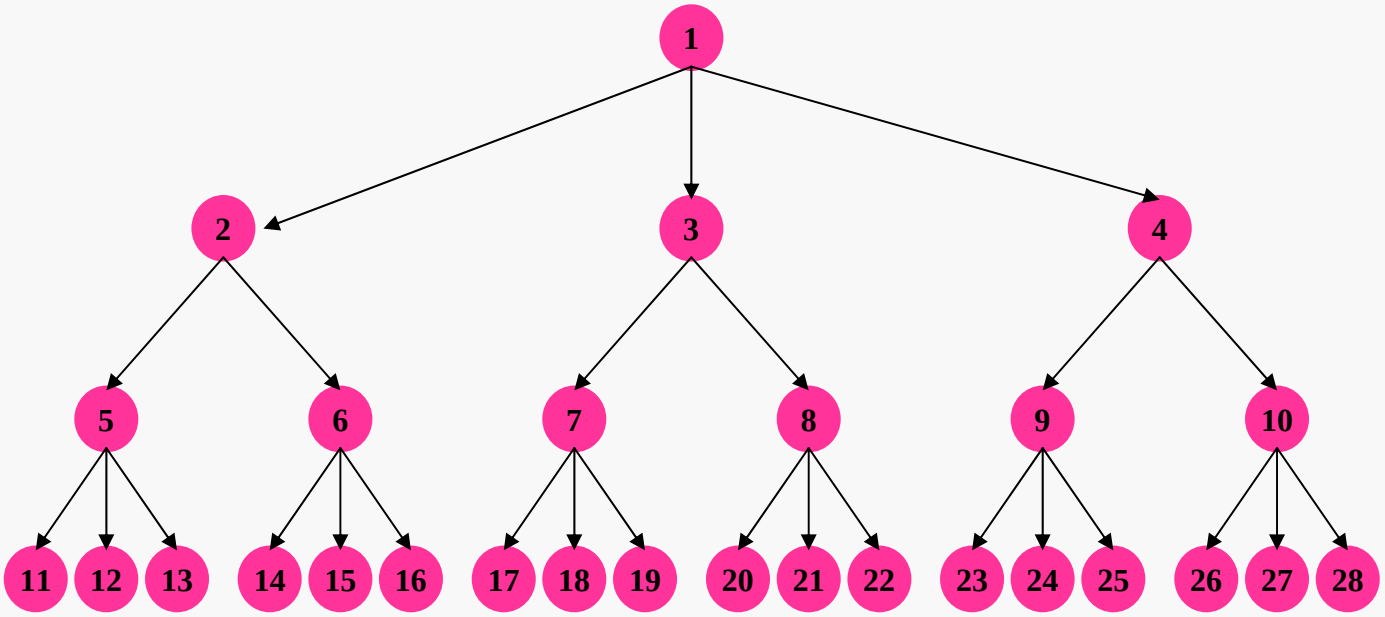
Systematic Search

- Test all combinations of all (torsional) DOF
- Screen each torsion with a pre-defined granularity
- Optionally minimize structures
- Computational cost is exponential with number of torsions

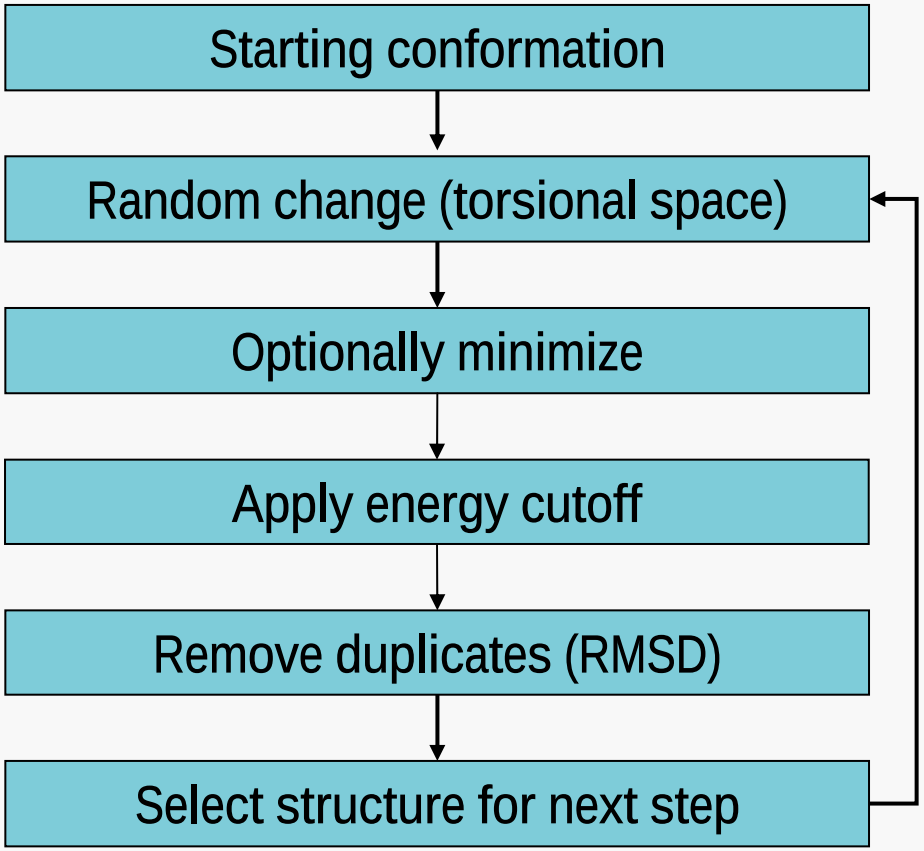
Torsion 1
(3 values)

Torsion 2
(2 values)

Torsion 3
(3 values)



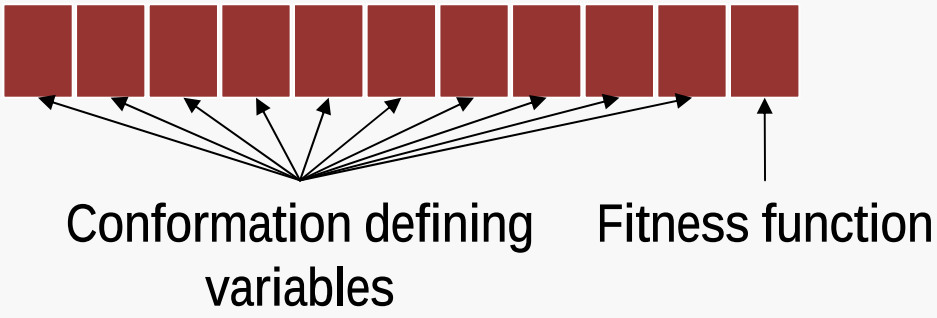
Stochastic Search

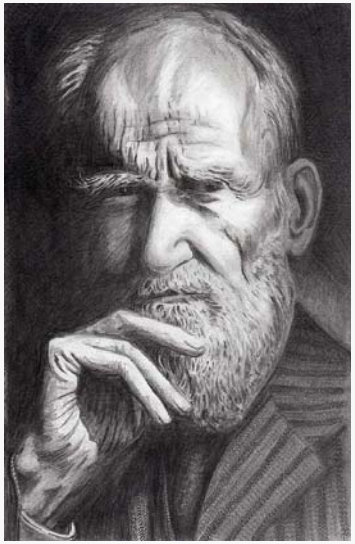


- Factors affecting performances
  - # cycles
  - Energy cutoff
  - RMSD threshold
  - Starting structure for next cycle

Genetic Algorithm

- Create a random populations of conformations (chromosomes)
- For each chromosome calculate a fitness value (conformational energy)
- Evolve population using genetic operators (selection of the fittest, mutations, cross-over)
- Optimize fitness function

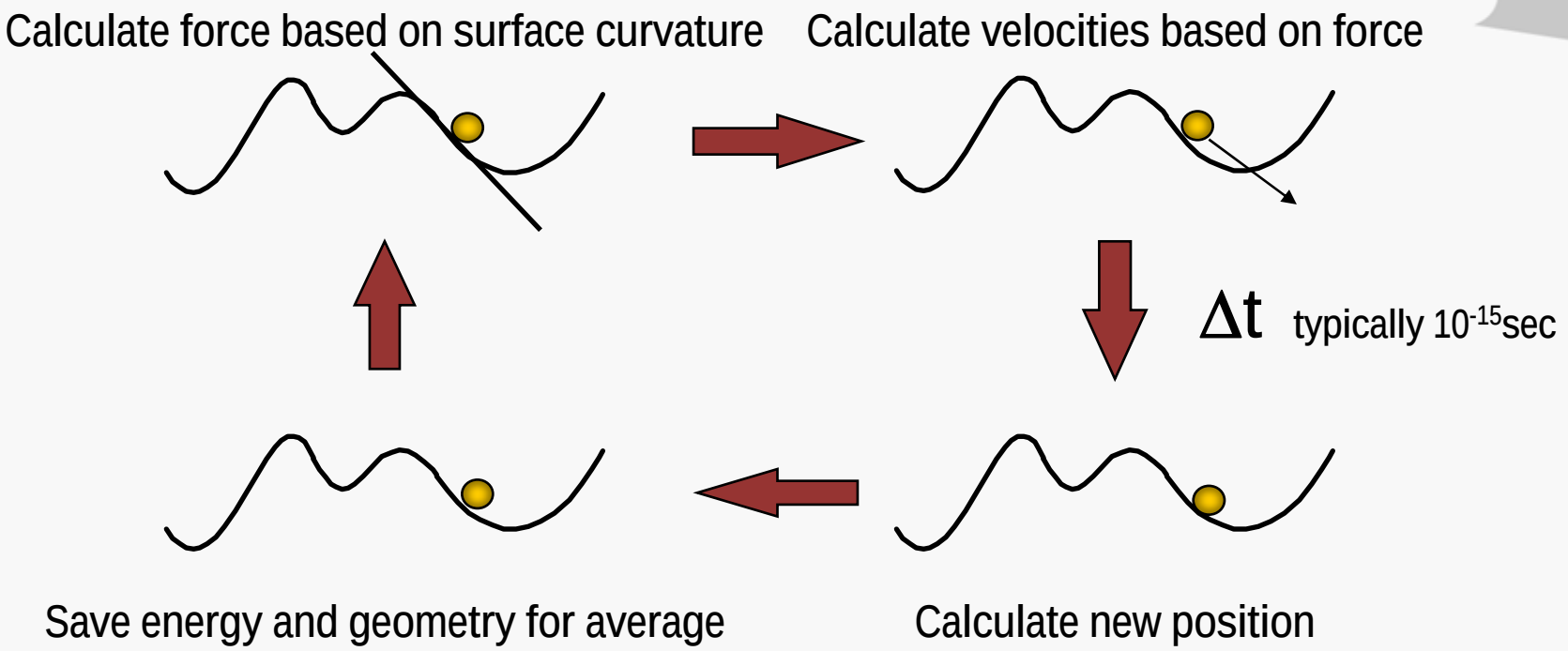




Additional Conformation Search Methods

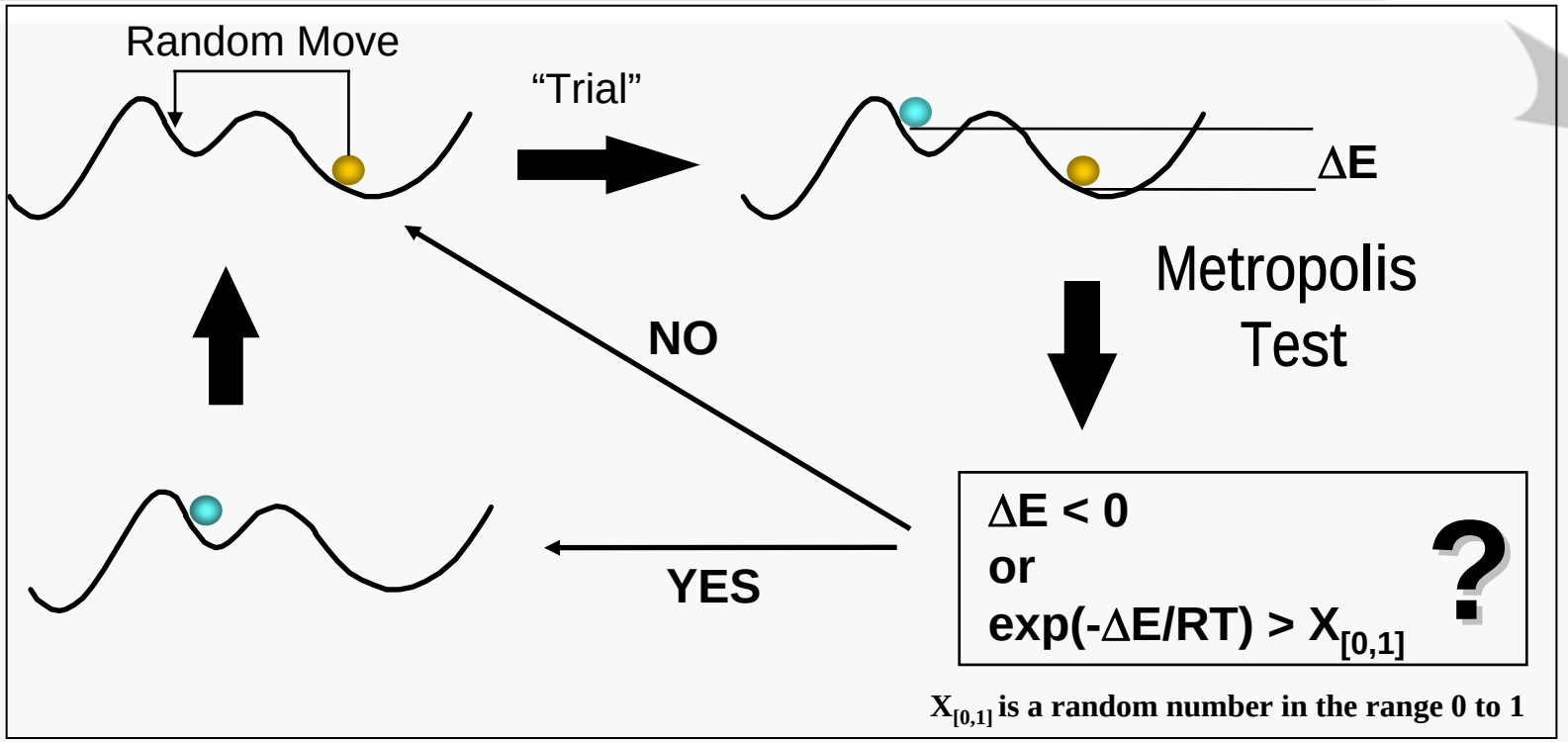
- Rule based (e.g., Omega)
  - Use pre-defined fragment conformations obtained, e.g., from CSD
- Tabu search (e.g., Catalyst's poling)
  - Avoid re-visiting already samples regiond of the PES
- Distance geometry
  - Used to derive structures from NMR data
- And many others…

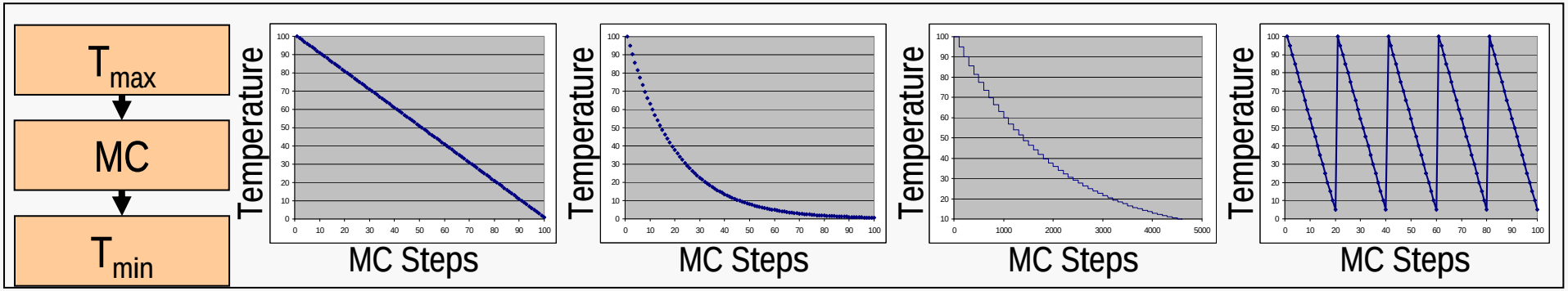
Molecular Dynamics



- Gives average quantities which reflect free energy
- Slow to cross barriers ~2-3 kcal/mol

Monte Carlo/Simulated Annealing (MC/SA)





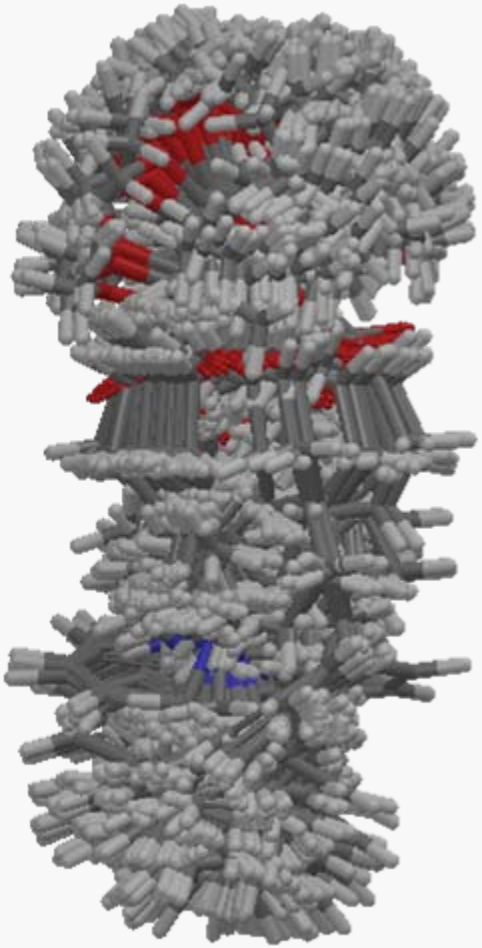
Performances of CS Methods

- Success determined in terms of having at least one structure close to the bioactive conformation
- All methods produce many conformations remote from the bioactive one, hence the need for focusing

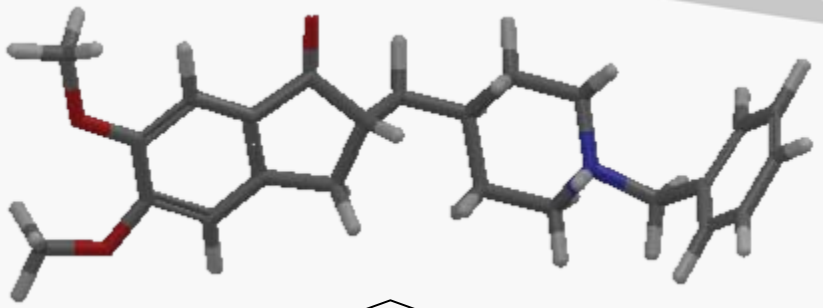
Software	Dataset	Performances
Balloon	311	90% with RMSD < 2A
CAESAR	918	60% with RMSD < 1A 90% with RMSD < 2A
Catalyst	32	47-50% with RMSD < 0.5A
	150	20% with RMSD < 0.5A 69% with RMSD < 1A
	193	70% with RMSD < 1A
	510	80% with RMSD < 1.5A 93% with RMSD < 2A
ConfGen	253	80% with RMSD < 1A
Confort	32	34% with RMSD < 0.5A
Cyndi	742	MECMB: 54% with RMSD < 1A FFMB: 37% with RMSD < 1A
Flo99	32	62-66% with RMSD < 0.5A
ICM	150	20% with RMSD < 0.5A 69% with RMSD < 1A
MacroModel: LMCS	32	69% with RMSD < 0.5A
MOE	256	95% with RMSD < 1.5A
Omega	32	41-50% with RMSD < 0.5A
	36	56-78% with RMSD < 0.5A
	150	27% with RMSD < 0.5A 69% with RMSD < 1A

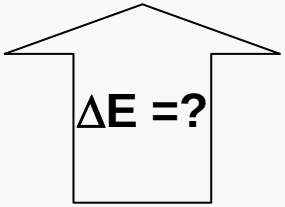
Challenges in the Field of Bioactive Conformations

Produce conformational ensemble

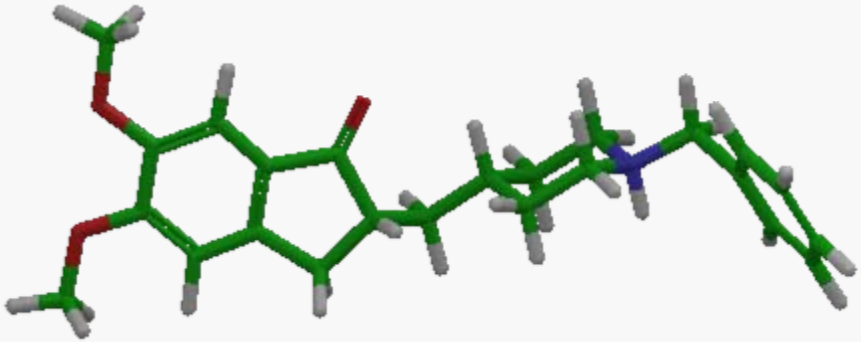


Identify bioactive conformation(s)





Score bioactive conformations



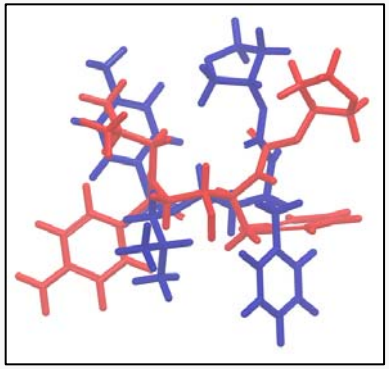
Challenges in Defining the Bioactive Structure

- Assume: crystal conformation represents the bioactive conformation
- Assumption questionable (but not enough solution phase data)

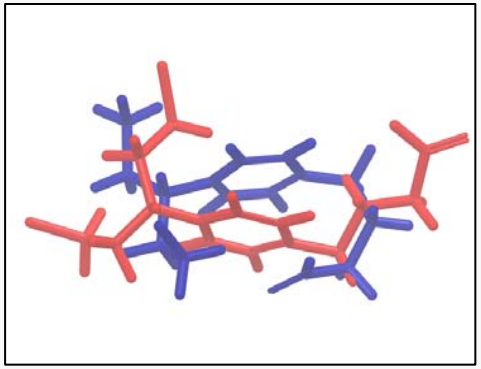
Identify Bioactive Conformations: Structure

- Assumption: Bioactive conformations more elongated than global energy minima

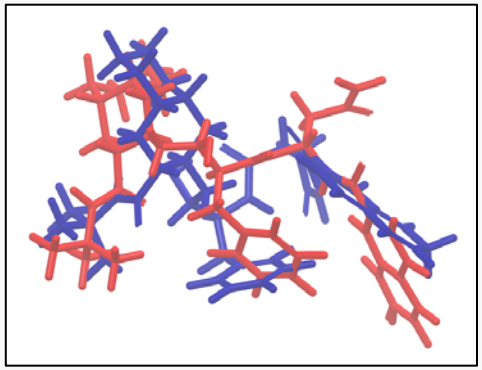
1HPV RMSD= 5.1Å



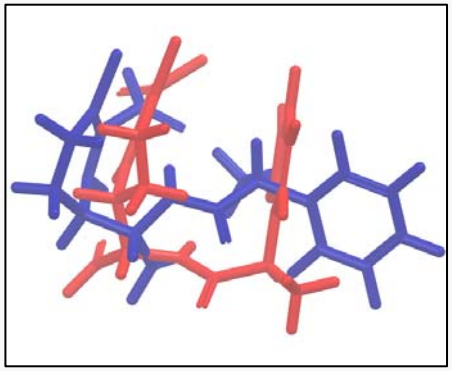
21GS RMSD= 2.8Å



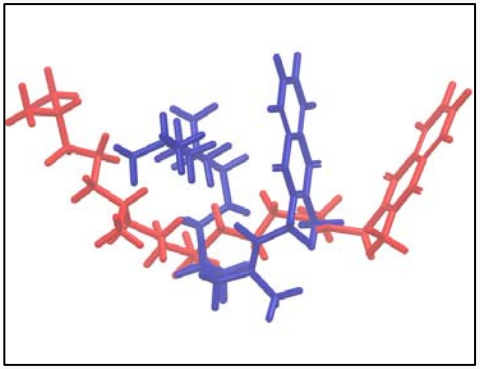
2NNP RMSD= 3.9Å



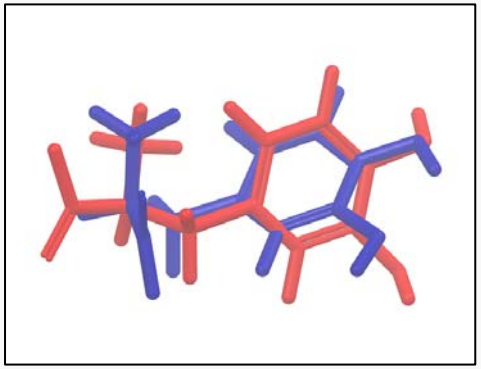
1FCN RMSD= 2.9Å



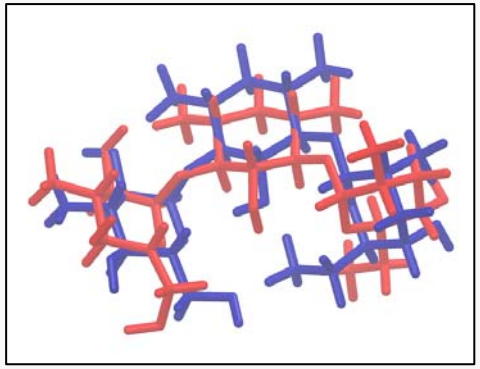
1JBO RMSD= 5.4Å



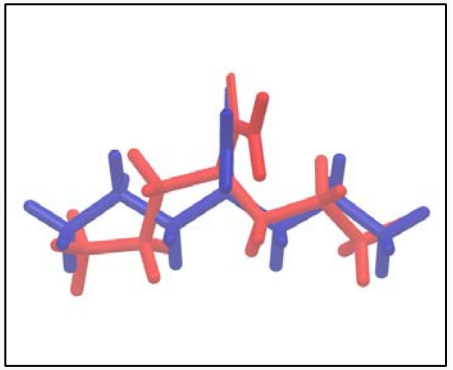
6PAH RMSD= 1.9Å



1M4D RMSD= 1.8Å

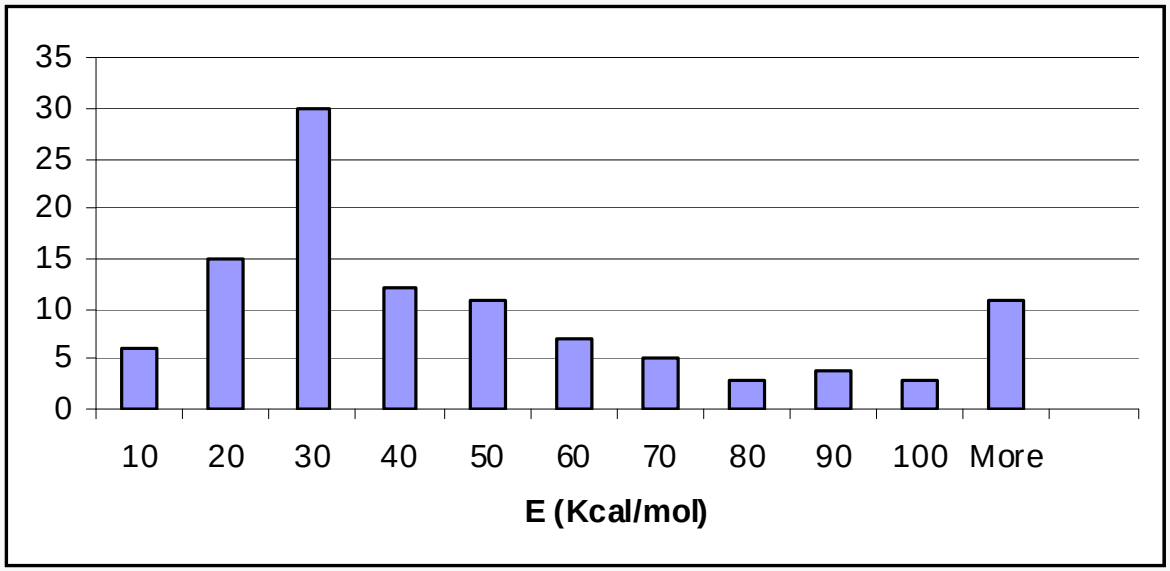


1NU3 RMSD= 1.4Å



Challenges in Determining the Bioactive Conformational Energy

$E_{bioactive}$-$E_{closest\_minima}$ (OPLS-AA; Kcal/mol; RMSD = 0.5±0.1)



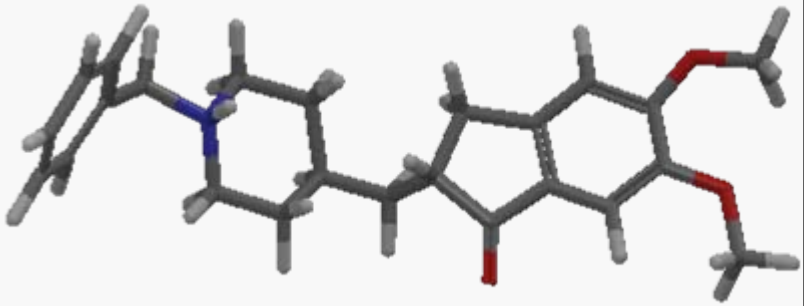
- Unconstrained minimization
- Protein-constrained minimization
- Flat-bottom constrained minimization
- B-Factor constrained minimization

$$E = 0 \qquad r \le \sigma$$
$$E = k(r - r_0)^2 \qquad r > \sigma$$
$$r = |r - r_0|$$

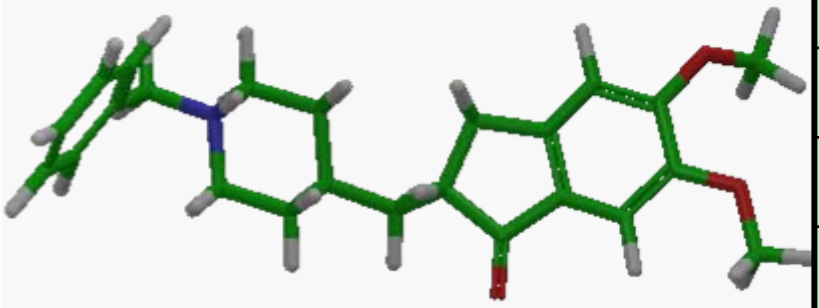
$$E = k(r - r_0)^2; k = 4\pi^2 k_B T / B$$

Minimization

E = 232 Kcal/mol



E = 34 Kcal/mol



PDB	Unconstrained	Protein constrained	B-factor constrained
1BZM	0.39	0.61	0.12
1CBX	0.56	0.76	0.07
1FKF	0.38	0.57	0.06
1HPV	0.74	0.51	0.12
1HVR	1.11	0.56	0.05
1ADD	0.51	0.85	0.06
1CPS	0.37	0.55	0.16
1PSO	0.88	0.58	0.11
1TLP	0.50	0.62	0.12
2GBP	0.17	0.29	0.00
Ave	0.56±0.28	0.59±0.15	0.09±0.05

Identify Bioactive Conformations: Energy

- Assumption: Bioactive conformations reside within well defined energy windows relative to global energy minima
- Reality: Not necessarily
- Some estimates are clearly unreasonable

- Discrepancies from
  - Inappropriate definition of bioactive conformations
  - Inappropriate force fields
  - Different data sets
  - Low resolution structures

Program	Force field	Energy cutoff (kcal/mol)
Catalyst	Modified CHARMM	20
Catalyst	Modified CHARMM	**10**, 20
Catalyst	Modified CHARMM	20
Catalyst	Modified CHARMM	7, 12, 15, **20**, 30, 40
MOE	MMFF94x	7, 12, 15, 20, 30, 40, > **15**
MacroModel	AMBER & MM3	12
MOE	MMFF94x	20
OMEGA	MMFF	3.3, 6.9, **50**
OMEGA	MMFF94s	3, **5**, 7
OMEGA	MMFF94s	5

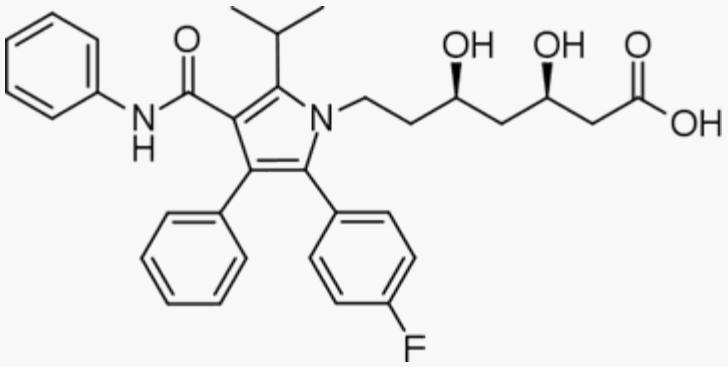
Bioactive Conformational Biasing: A New Method for Focusing Conformational Ensembles on Bioactive-Like Conformers

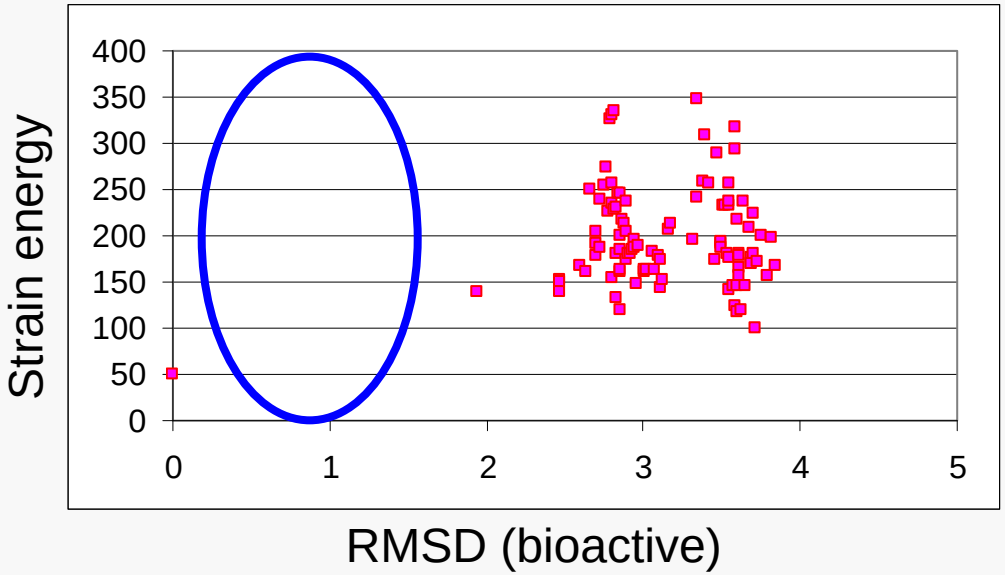
- Goal
  - Enrich conformational ensembles by bioactive-like conformations (RMSD < 1Å)
  - Retain a sufficiently large number of bioactive-like conformations
  - De-rich conformational ensembles by non-bioactive conformations (RMSD > 2.5Å)
- Dataset
  - 71 ligands (47 in training, 24 in test)
  - Ligand and protein diversity
  - High resolution proteins
- Conformational ensembles generated in MacroModel

Pre-Filtration

- Can we always identify bioactive conformations?

Atorvastatin

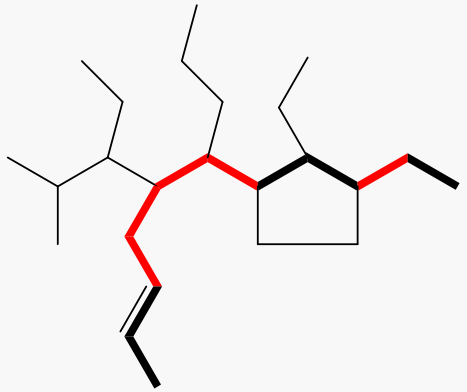




Pre-Filtration: M-PROB

- Can we always identify bioactive conformations? NO!

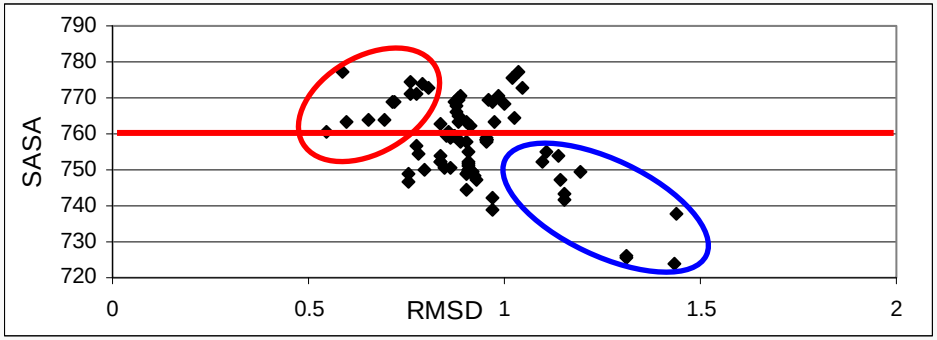
M-PROB: No. of rotatable bonds along maximal path

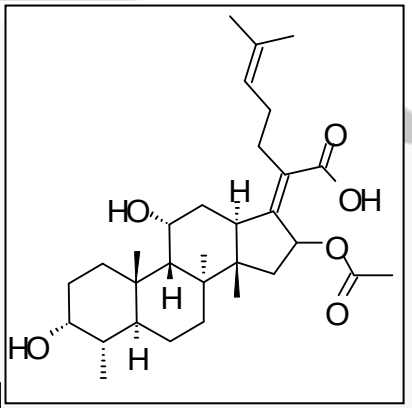


M-PROB	1-3	4-5	6-7	8-9	10-18
# ligands	16	10	11	5	5
# “good” ligands	16	7	9	0	3
# “bad” ligands	0	1	2	5	2

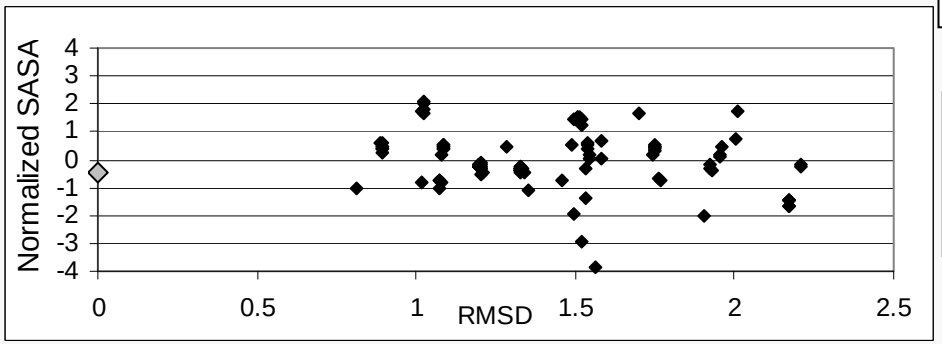
Single 3D Descriptors: SASA

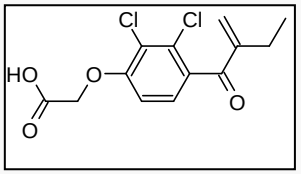
Fucidic Acid



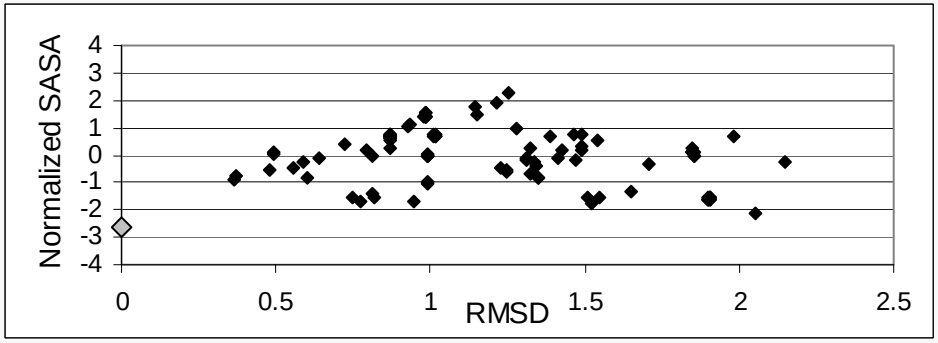


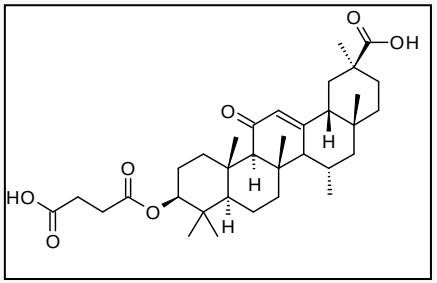
Etacrynic Acid



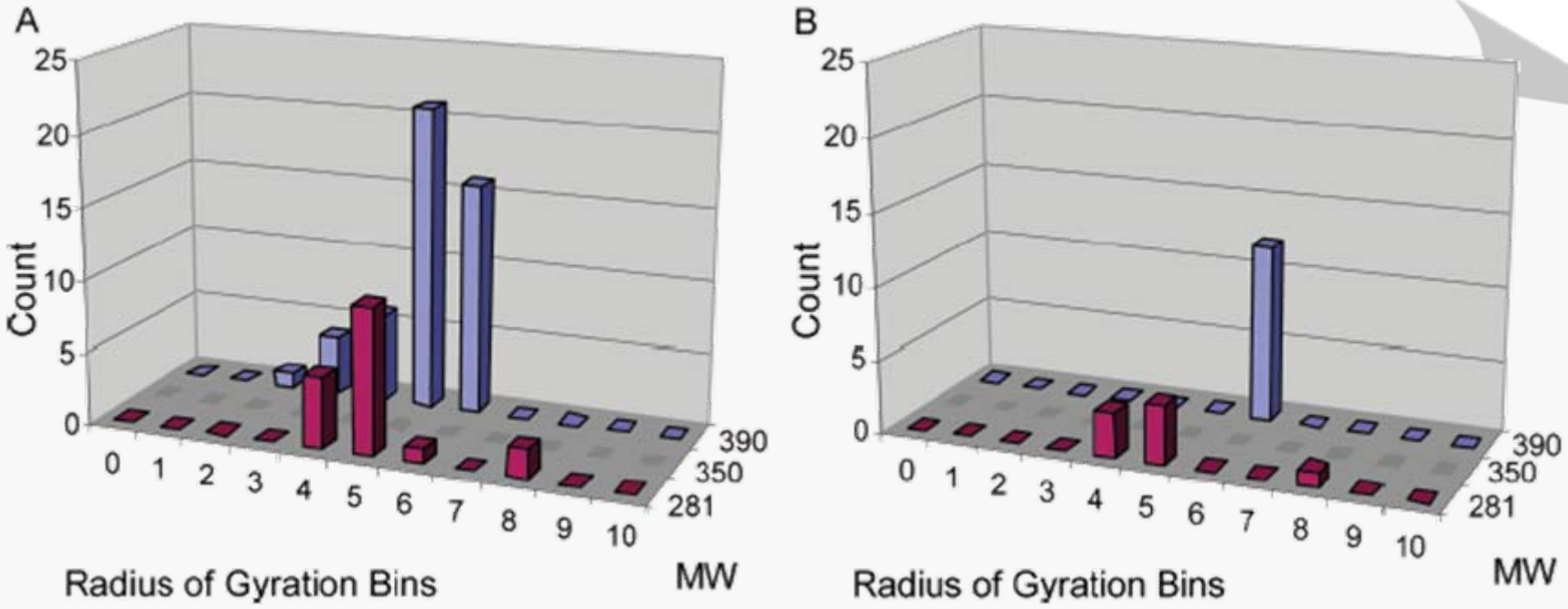


Cerbenoxolone





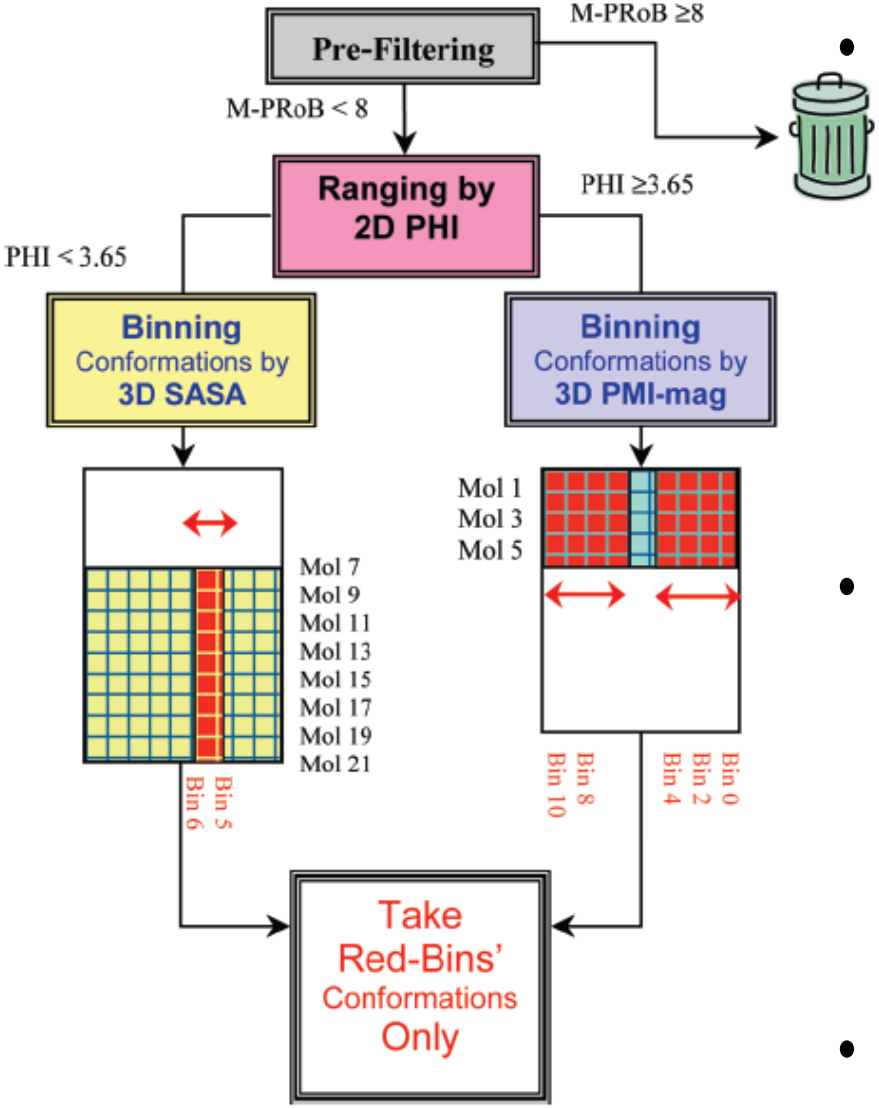
2D-3D Combination: MW & ROG



MW > 350 $\Rightarrow$ bin 6 (retrieve bioactive conformations of cyclothiazide; blue)

MW < 350 $\Rightarrow$ bins 4,5 (retrieve bioactive conformations of flufenamic acid; red)

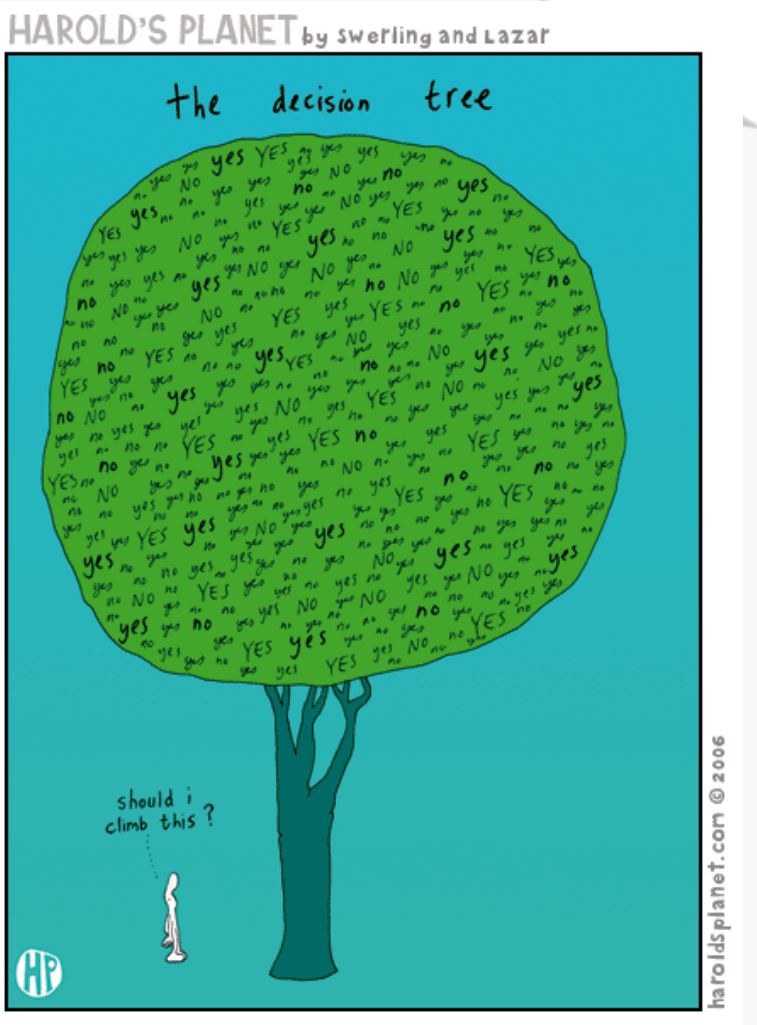
2D-3D-3D Models



- Pre-filtration
  - 5 compounds pre-filtered
  - 4 with no bioactive conformations
  - 1 with a single bioactive conformations
  - Sensitivity 80%
  - 2 compounds with no bioactive conformation not filtered
- Filtration
  - 36% of all conformations removed
  - 39% of "bad" conformations removed
  - 26% of "good" conformations removed
  - 74% compounds retained sufficiently large number of "good" conformations
- Overall success rate: 75%

Interim Conclusions & Future Work

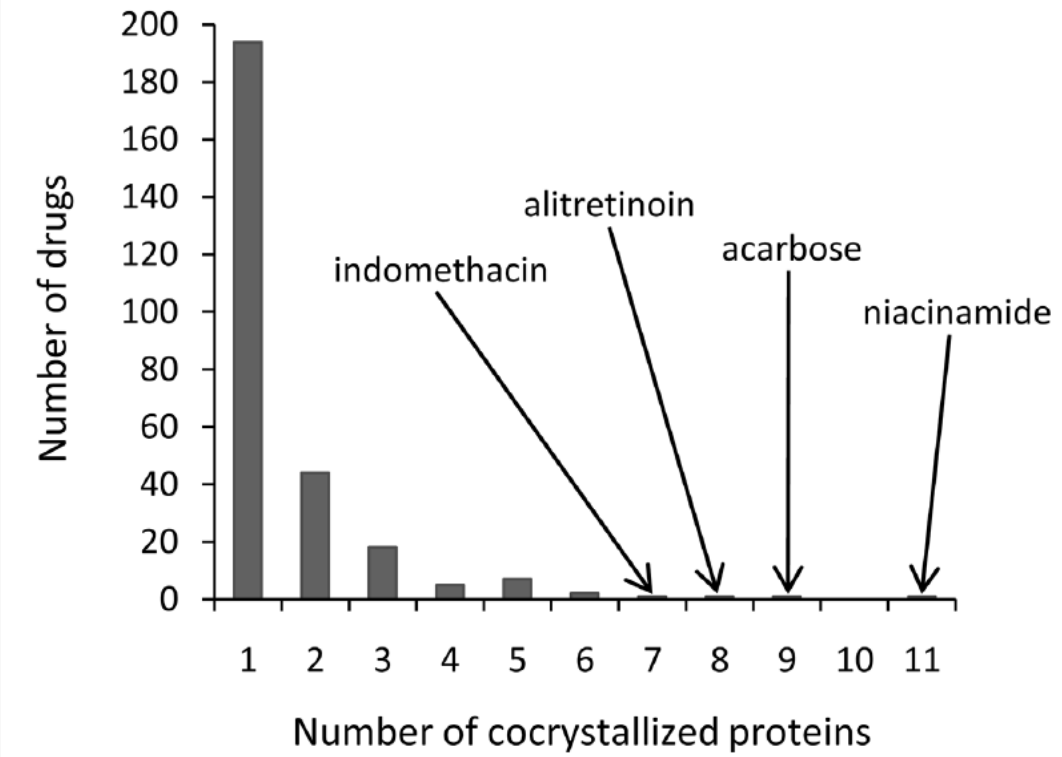
- Conformational ensembles could be focused on bioactive conformations using ligand characteristics
- A larger data set
- Incorporate target information
  - Bioactive conformations are target dependent
- A more flexible algorithm
  - Test more descriptors combinations



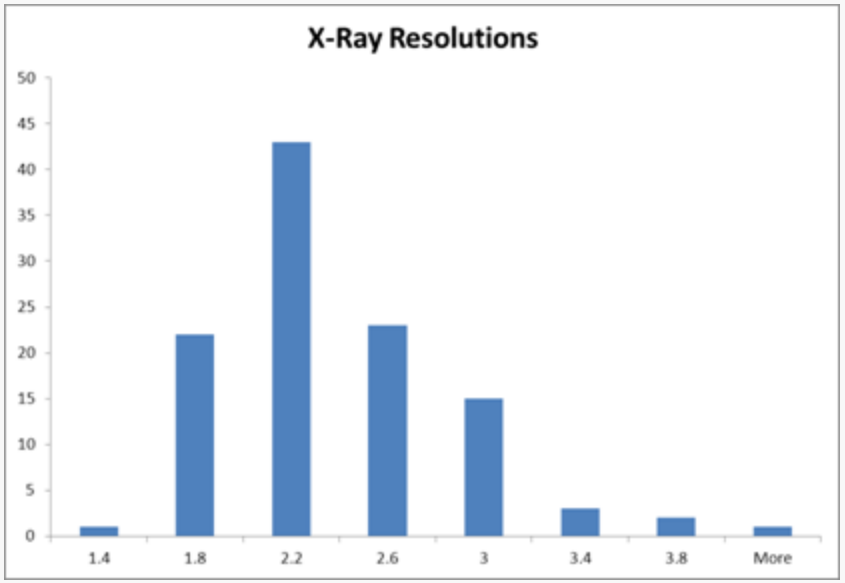
A Drug Binding Site Database

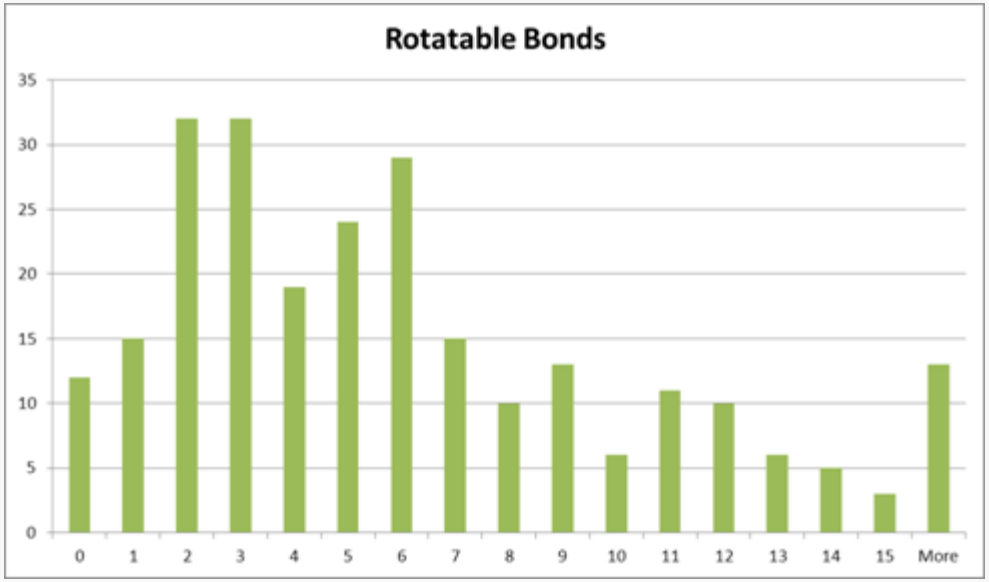
(Kinnings et al. PLoS Computational Biology, 2010, 6, 1)

- 274 approved drugs
- 962 drug binding sites
- 194 drugs co-crystallized with a single unique protein
- Multiple drugs crystallized with multiple proteins
  - Indomethacin (non-steroidal anti-inflamatory)(7)
  - Alitretinoin (antineoplastic)(8)
  - Acarbose (anti-diabetic)(9)
  - Niacinamide (vitamin)(11)

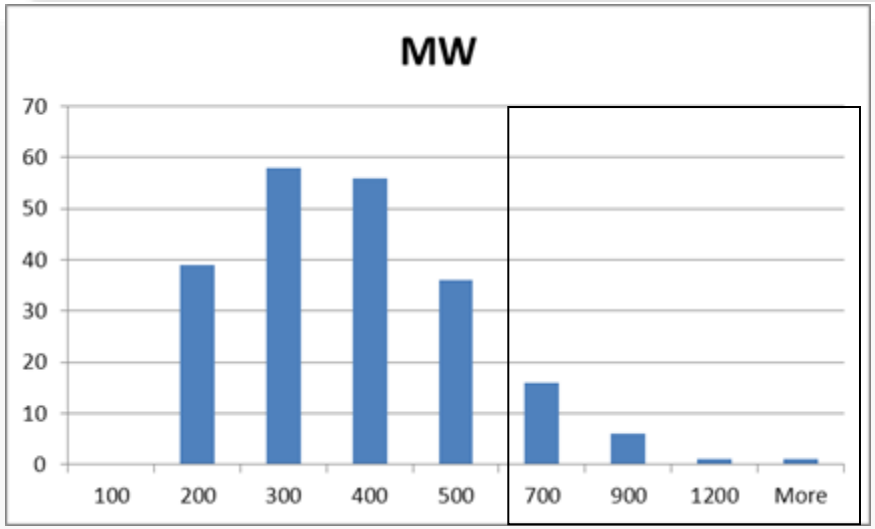


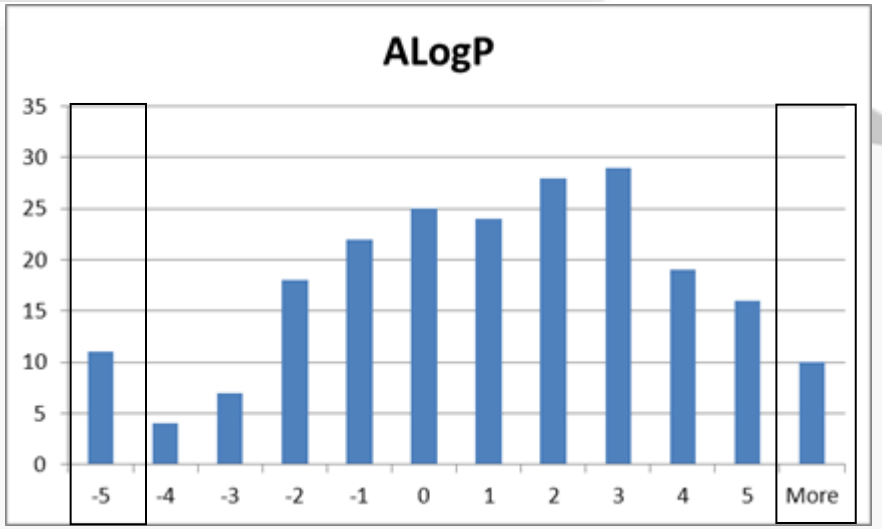
Database Characteristics

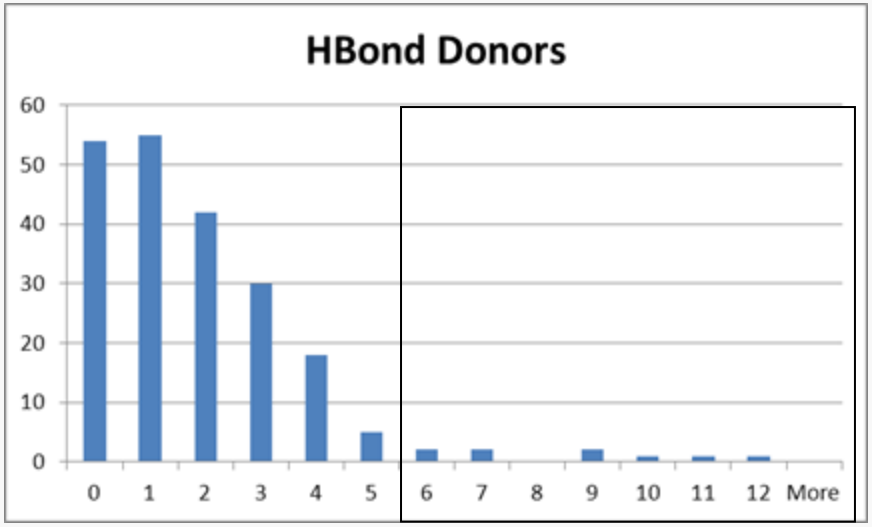


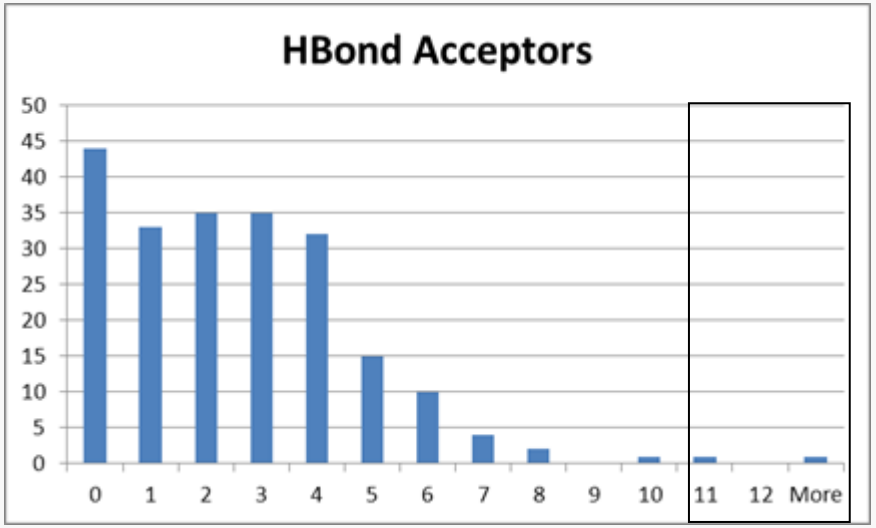


Database Characteristics: Lipinski's Rules

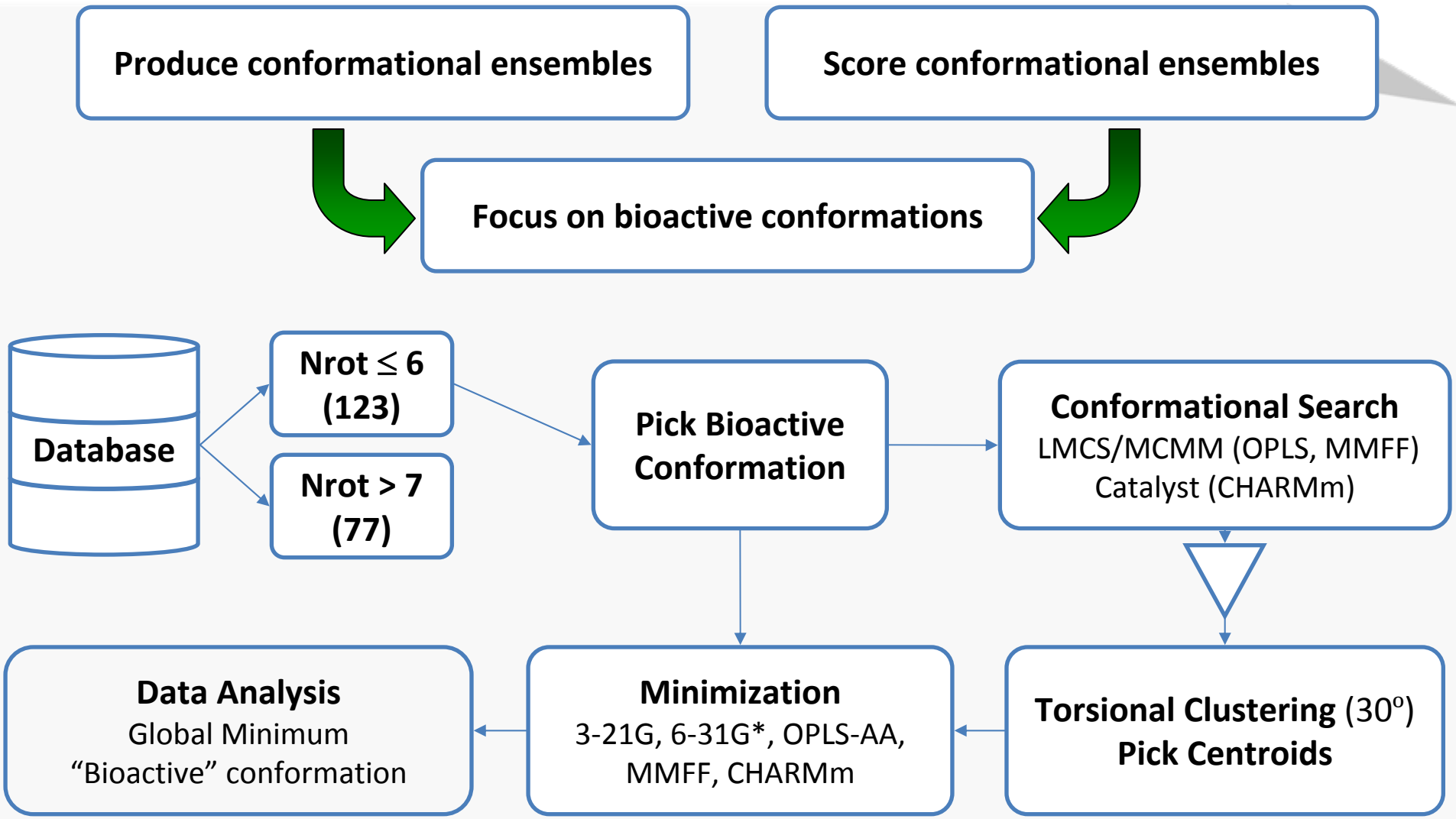




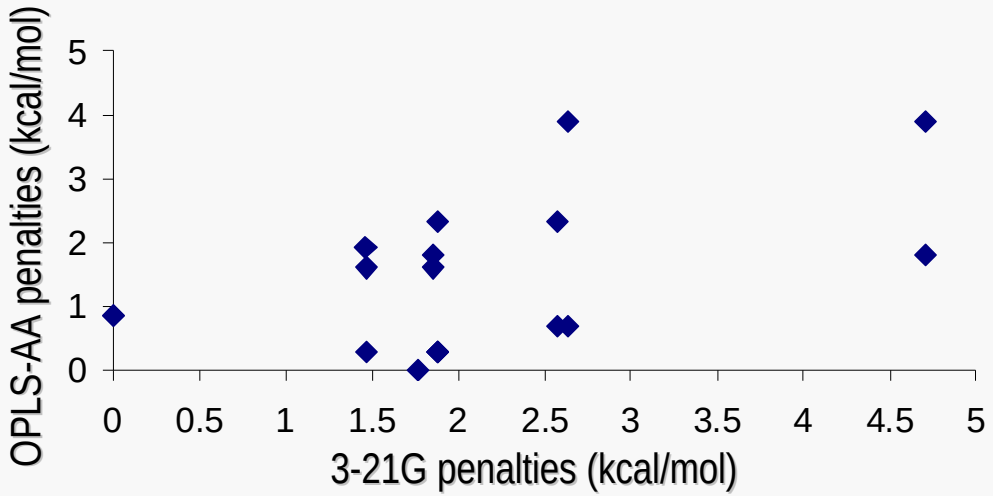


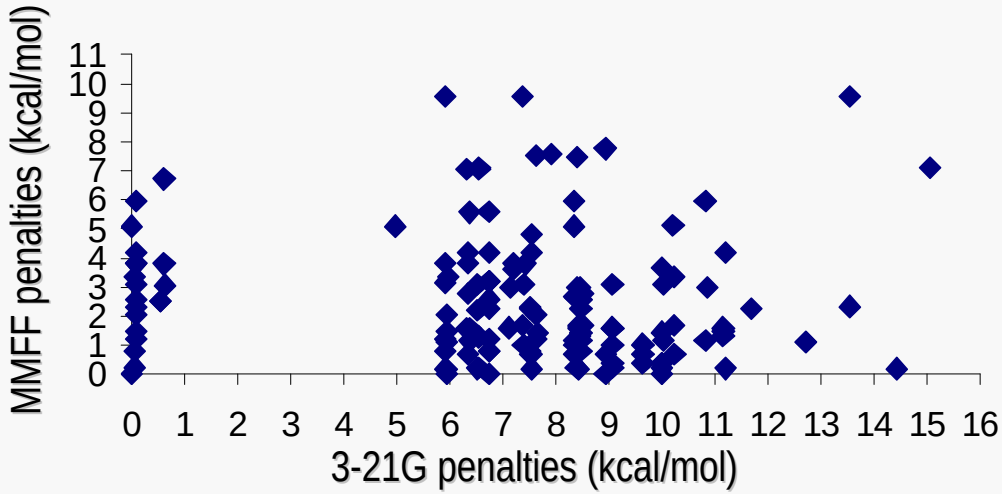


The Workflow

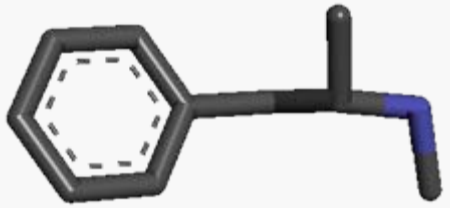


Did We Really Need to Work That Hard (~15K QM Calc.)?

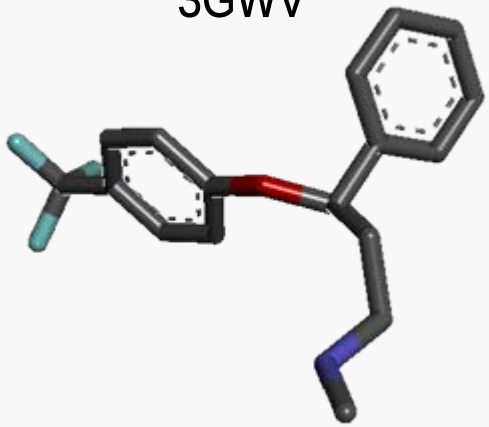




3GKZ



3GWV



Can CS Methods Generate Bioactive Conformations?

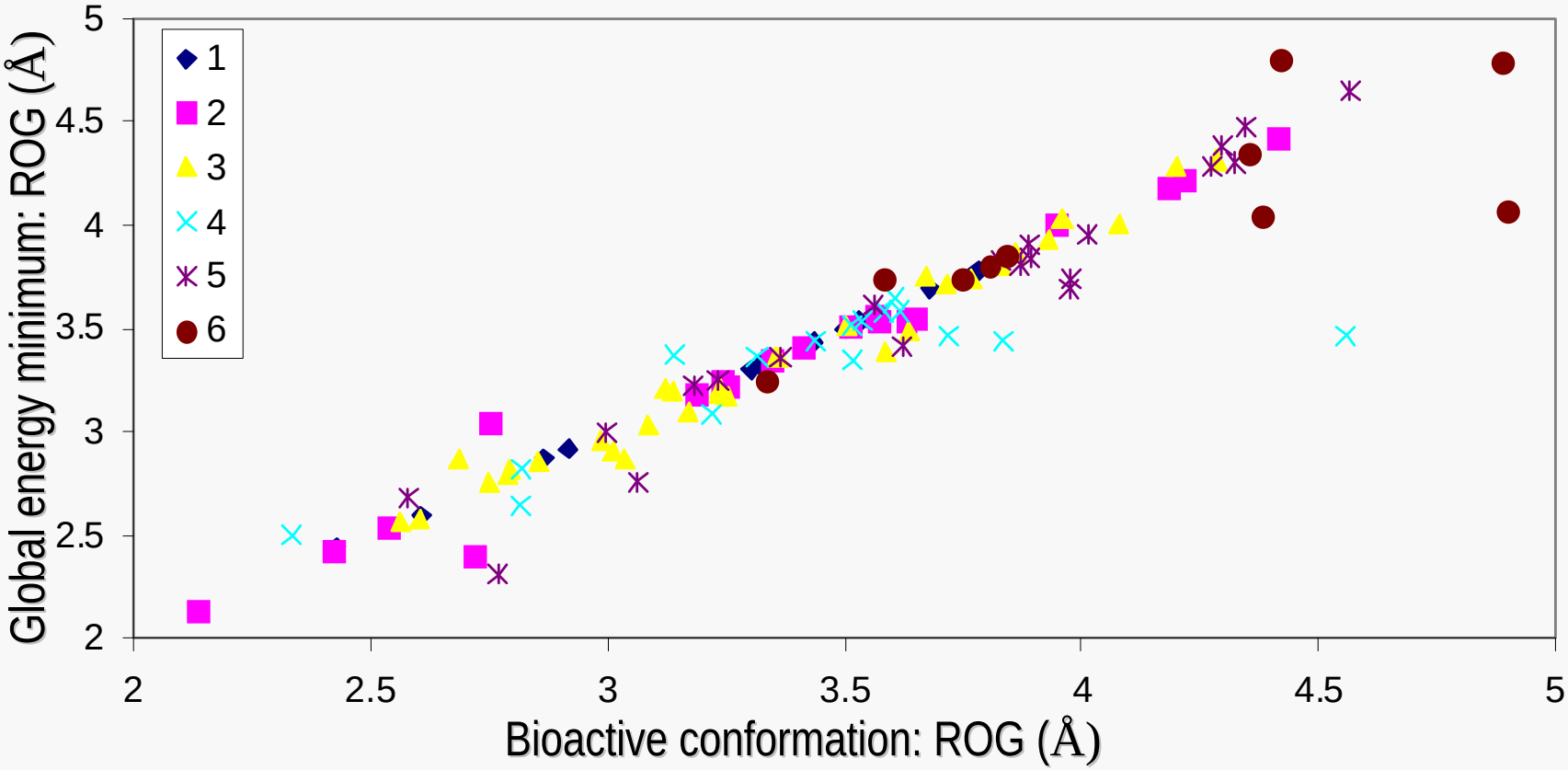
	# ligands	RMSD < 0.5Å	RMSD < 1.0Å	RMSD < 1.5Å	RMSD < 2.0Å	RMSD > 2.0Å
OPLS-AA	117	0.62/0.22	0.91/0.56	0.97/0.73	0.99/0.88	0.01/0.12
MMFF	119	0.63/0.20	0.92/0.54	0.98/0.79	1.00/0.93	0.00/0.07
CHARMm	120	0.78/0.20	0.97/0.52	1.00/0.74	1.00/0.91	0.00/0.09

	# ligands	RMSD < 0.5Å	RMSD < 1.0Å	RMSD < 1.5Å	RMSD < 2.0Å	RMSD > 2.0Å
OPLS-AA	110	0.75/0.28	0.98/0.67	1.00/0.81	1.00/0.93	0.00/0.07
MMFF	110	0.74/0.26	0.98/0.58	1.00/0.80	1.00/0.95	0.00/0.05
CHARMm	110	0.78/0.23	0.98/0.56	1.00/0.75	1.00/0.92	0.00/0.08
3-21G	110	0.73/0.21	0.98/0.51	1.00/0.75	1.00/0.94	0.00/0.06
6-31G*	88	0.83/0.32	0.99/0.66	1.00/0.83	1.00/0.97	0.00/0.03

- In general, our workflow can produce bioactive conformations slightly better than “standard” CS methods
- From within the methods tested in this work, OPLS-AA and 6-31G* perform the best

Identify Bioactive Conformation: The Structure

- Assumption: Bioactive conf. more elongated than global energy minima
- Reality: Not necessarily



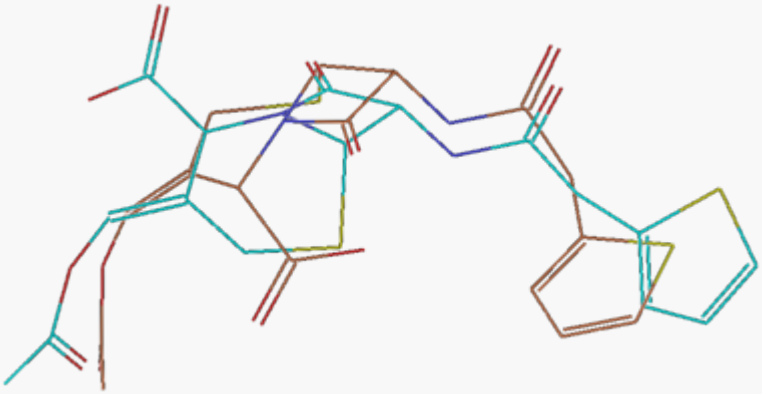
Hydrophobicity Dependent Ligand Unfolding

- Hydrophilic ligands tend to fold in their binding sites

RotBond = 5, 6	n	ClogP	$ROG_{bioactive}$ - $ROG_{global\ minimum}$
Hydrophilic ligands (ClogP < 0)	12	-1.9±1.1	-0.14±0.27
Hydrophobic ligands (ClogP > 0)	19	2.5±1.5	-0.02±0.17

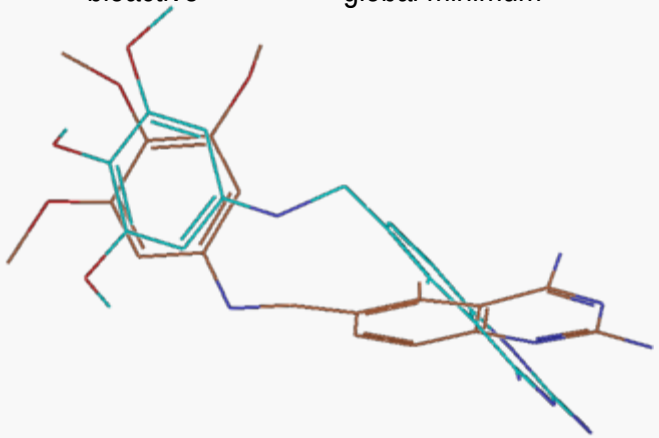
2ZQ9: ClogP = -1.13

$ROG_{bioactive} - ROG_{global\ minimum}$ = -0.83



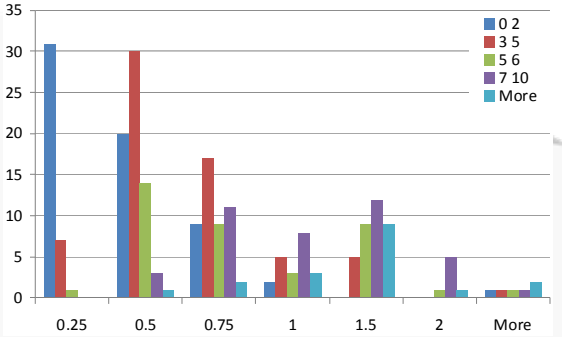
3CLB: ClogP = 1.15

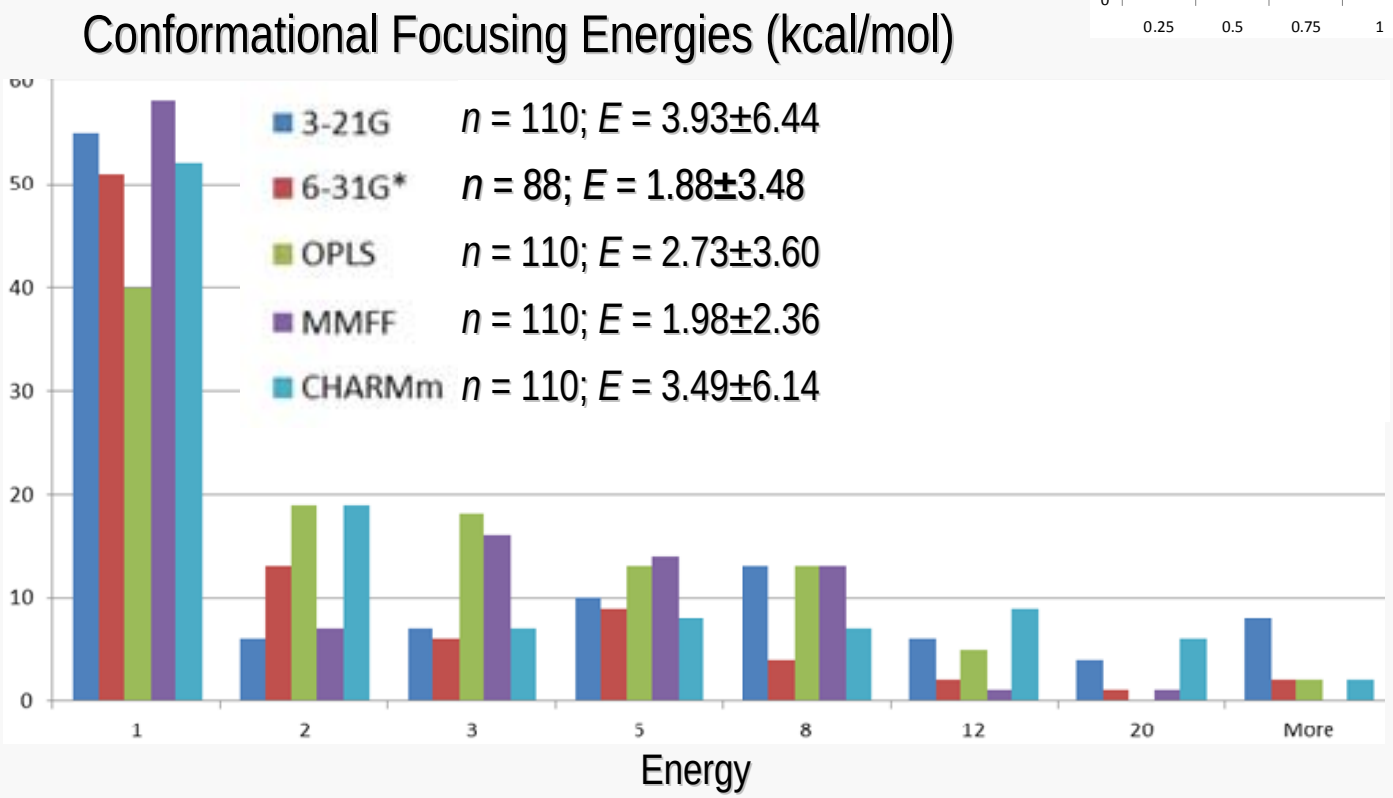
$ROG_{bioactive} - ROG_{global\ minimum}$ = 0.39



Force Field-Based Conformational Focusing Energies

- Conformational focusing energies calculated relative to the constrained-free minimized bioactive conformation





Comparison with Docking

- Glide docking with default parameters
- Success defined according to lowest energy structures

	# ligands	Average	SD	RMSD < 0.5Å	RMSD < 1.0Å	RMSD < 2Å
Docking	110	0.78	0.61	0.42	0.83	0.95
3-21G	110	1.05	0.63	0.21	0.51	0.75
6-31G*	88	0.84	0.57	0.32	0.66	0.97
OPLS-AA	110	0.92	0.62	0.28	0.67	0.93
MMFF	110	0.94	0.60	0.26	0.58	0.95
CHARMm	110	1.04	0.68	0.23	0.56	0.92

Conclusions I

- Bioactive conformations are important and interesting
- Can CS methods generate bioactive conformations?
  - For rigid ligands (# RotBonds $\leq$ 6) a bioactive conformation is likely to be found in the conformational ensemble, although not as the global minimum.
  - For more flexible ligands (# RotBonds $\geq$ 8) the probability of identifying bioactive conformations is lower
  - Our workflow performs better in these respect thans “simple” CS methods
  - Medium level QM calculations show promise
- Could bioactive conformations be identified based on their structures?
  - Probably but more work is needed
  - Bioactive conformations are not necessarily more elongated than global minima ones

Conclusions II

- Could bioactive conformations be identified based on their energies?
  - Our data support energy cutoffs in the order of 5-6 kcal/mol for force field calculations
  - QM data show a trend towards lower penalties as the size of the basis set increases
- How well do CS methods reproduce bioactive conformations compared with docking simulations?
  - Docking is better than CS but medium level QM is not far behind
  - What does this tell us about our scoring functions?
  - Should we re-visit rigid docking?

Acknowledgments

Lab members

- Hannah Avgy
- Dr. Efrat Noy
- Dr. Netaly Khazanov
- Dr. Neta Uritsky
- Yocheved Baime
- Tamar Tlashadze
- Shirin Kahremany
- Lubov Simhaev
- Abraham Yosipof
- Michael Zenin

Department of chemistry

- Dr. Pinchas Aped

Former Epix Colleagues

- Dr. Boaz Musafia