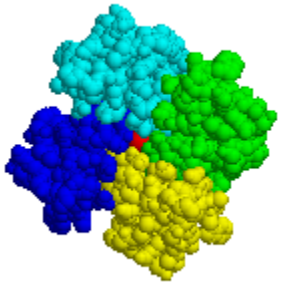


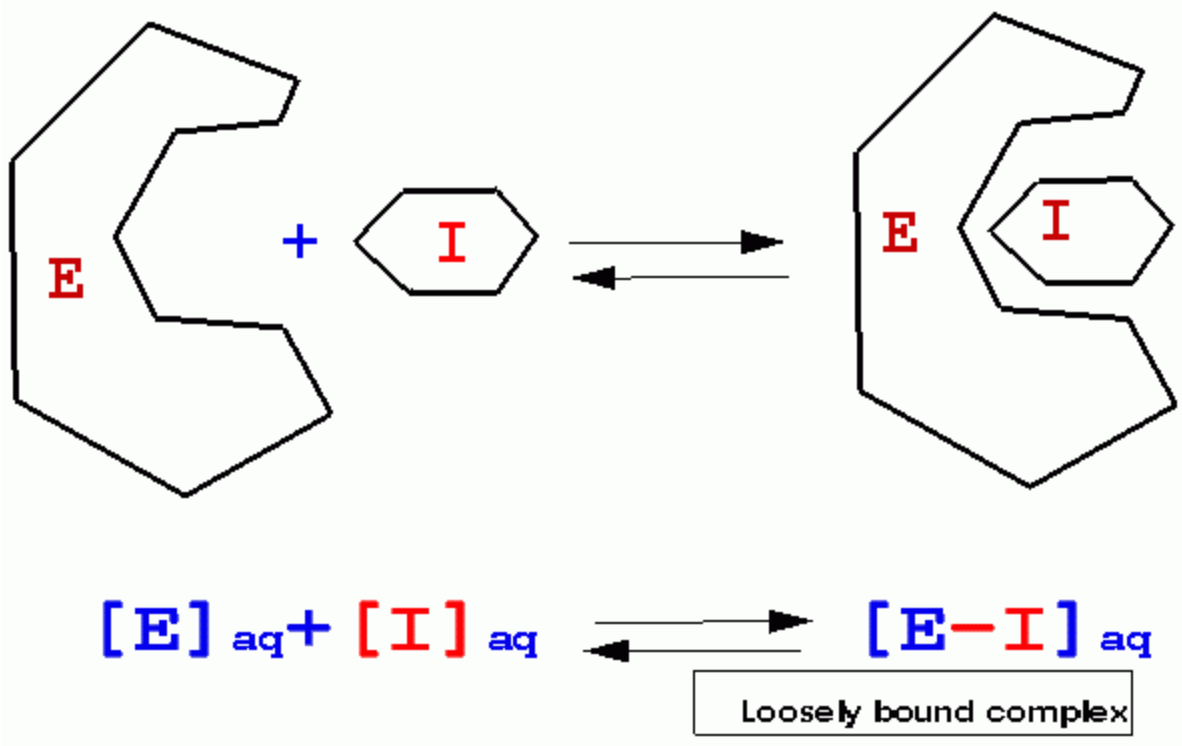
Modeling Protein-Ligand Binding Using AutoDock

S. Ravichandran
Advanced Biomedical Computing Center
National Cancer Institute
Frederick, MD 21702
Email: sravi@ncifcrf.gov
06/20/2006

I will not talk about

- Protein-Protein docking
- Manual Docking
- Incremental Constructs Method
- Virtual Docking
- Free Energy Methods
  - Docking methods will always discuss Free Energy
    - Stability is related to Free Energy

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick



Docking

- Docking (binding) gives the protein the ability to promote or inhibit certain reactions
- Importance:
  - Drug Designing
  - Designing bioactive compounds
  - …….

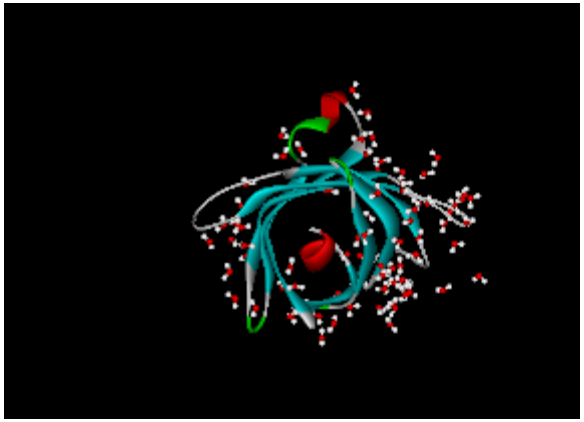
S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

What is Docking?

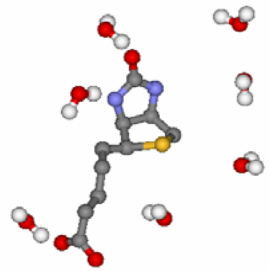
Structure of Target molecule

Posing: Orientation and Conformation

Conformational flexibility



+



ΔG_{bind} ⟷

$$\Delta G_{bind} = - RT\ln K_A$$

Docking (2 stages)

1) Structure modeling (Posing)

Pose Scoring

2) Ranking (Prediction of Activity)

Rank Scoring

More complicated than pose scoring

$$\text{Binding affinity} = K_A = K_i^{-1} = [E\text{--}I]/[E][I]$$

D. Kitchen et al, Nature Review, Vol 3, 935, 2004 (refs therein)
Kuntz et al several papers
Broojimans, N, Kuntz, I.D., Annu. Rev. Biophys. Biolmol Str, 32, 335, 2003

6/21/2006

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Figure Made with DSViewerPro

Binding Constants

- $R + L \leftrightarrow R\text{-}L$
  - At equilibrium
    - $k_2[R\text{-}L] = k_1[R][L]$
      - rate of dissociation, rate of association, k_1,k_2 (rate constants)
    - $k_2/k_1 = k_d = [R][L]/[R\text{-}L]$
      - corresponds to [L] at which binding site of the protein is 50% occupied
  - If we need a higher concentration of [R] & [L] for binding, then k_d is higher (binding is weaker)
    - μM(10^{-6}) than nM (10^{-9}) (small k_d, stronger binding)
  - sub-nanomolar binding is a rare occurance
    - Avidin (glycoprotein) binds k_d 10^{-15}M
      - binds to Biotin (Vitamin H) & prevents it from absorbing in the gastrointestinal tract

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Things you need

- Protein Molecule & Small molecule

Cambridge Crystallographic Data Centre

Any basic modeling software

- Software
  - AutoDock, Dock, Ludi, FlexX, Glide etc
- Computers
  - CPU, Disk-space, etc.

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

?s to ask before docking

- Critical analysis of the protein structures
  - Resolution, missing residues, High thermal factors
  - X-ray or NMR
    - If NMR, which model

Interferon-Gamma

Structure ID	Resolution (Å)	Release Date	Resolution (Å)	Source	Expression Host
1EKU	2.9000000953674316	2000-09-01	2.9	HOMO SAPIENS	ESCHERICHIA COLI
1FG9	2.9000000953674316	2000-08-11	2.9	HOMO SAPIENS	n/a
1FG9	2.9000000953674316	2000-08-11	2.9	HOMO SAPIENS	ESCHERICHIA COLI

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

?s to ask before docking

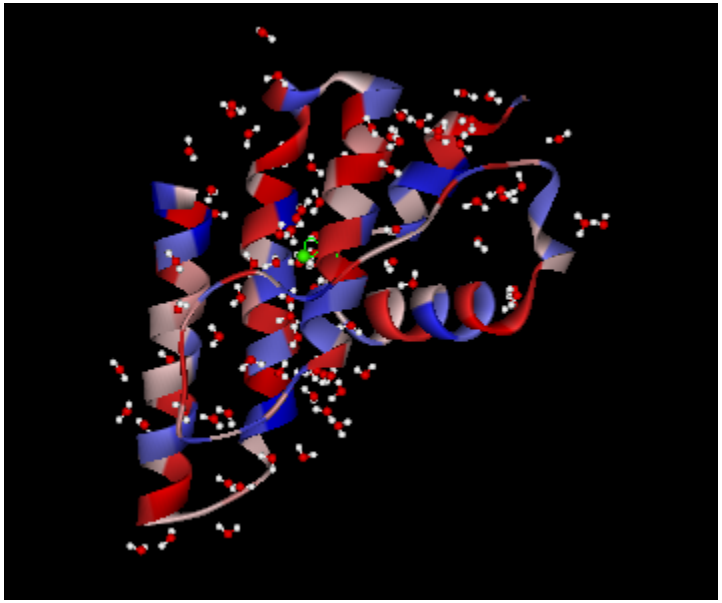
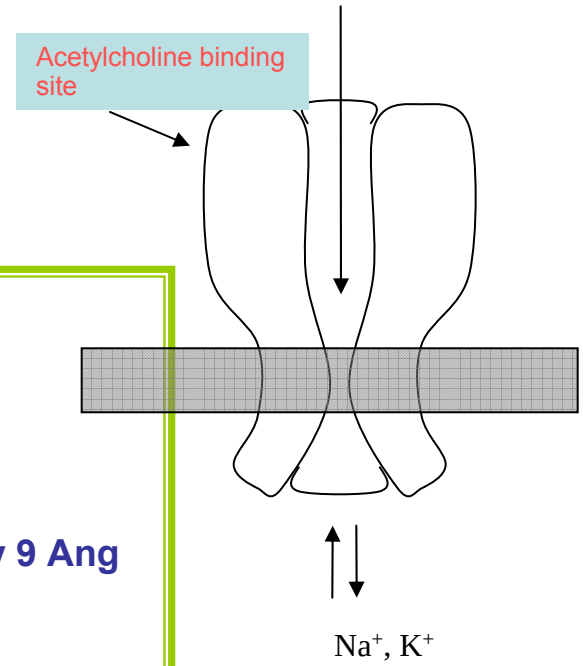


Figure Made with DSViewerPro

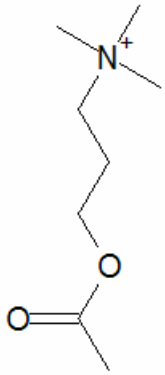
- What to do with bound ions, bound water molecules
  - delete them, keep them. Why?
- Crystal structure
  - Rigid nature, bound ions and water are held in the active site. Are they valid configurations?

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Binding domains in nAChRs



- At the extracellular hydrophilic portion binds
  - Ach
  - Agonist
    - Conformational change (opening of channel) (millisecond time scale) –Electron microscopy 9 Ang Res.
- Hydrophobic (transmembrane region)
  - Non-Competitive inhibition
- Three states: Open, Closed, desensitized
- NCI binding
  - Marketed drugs are also NCIs
    - Antidepressents nefazoden, sertraline act as NCIs for nAChRs in the brain
    - Side effects: Impairment of Cardiovascular function during Ketamine anesthesia due to NCI-Ketamine binding



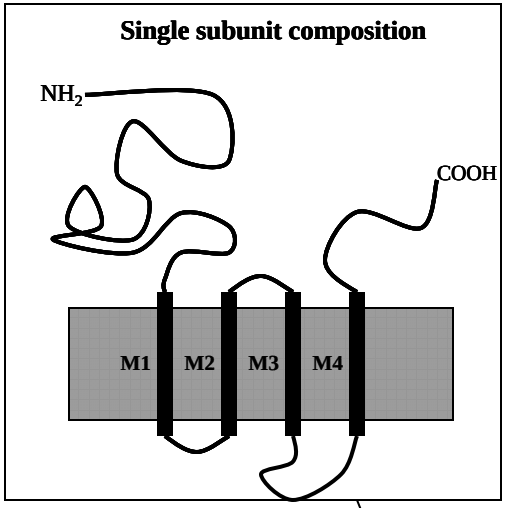
S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

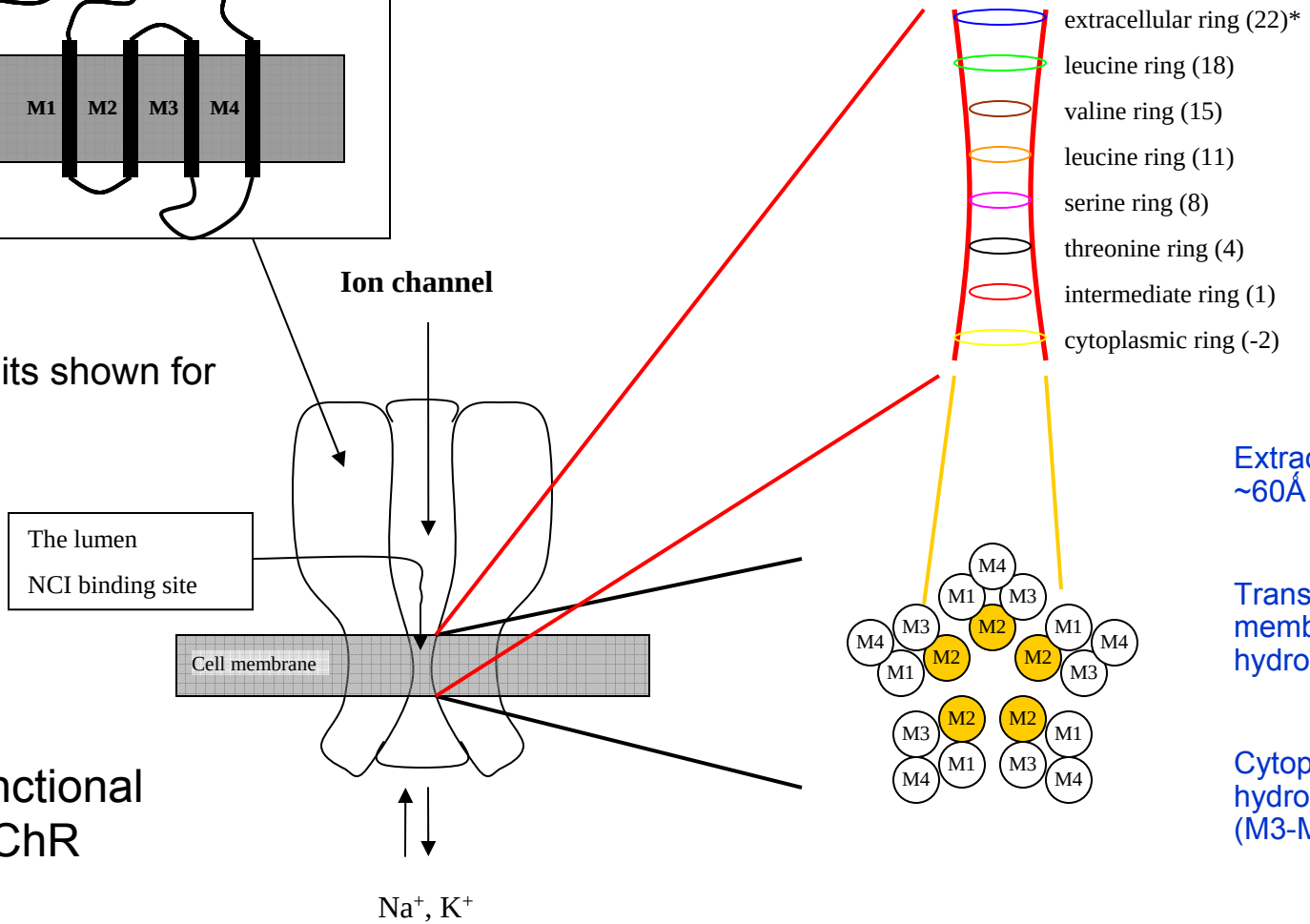
Arias, Brain Research Review, 25 (1997) 133-191

K.Jozwiak, S. Ravichandran, J.R. Collins and I.W. Wainer, *J. Med. Chem*, Jul 29, 47:16 4008-4021, **2004**
D.A. Doyle, Science, 280:69 (1998); Dutzler et al, Nature, 415, 287 (20002)

3D structure of nAChRs lumen region is not available.

Membrane proteins, Too large for NMR





Rings are important for proper functioning

Singlepoint mutation-> diseases

3 subunits shown for clarity

Extracellular ~60Å

Trans-membrane hydrophob ~40Å

Cytoplasmic hydrophilic (M3-M4) ~15Å

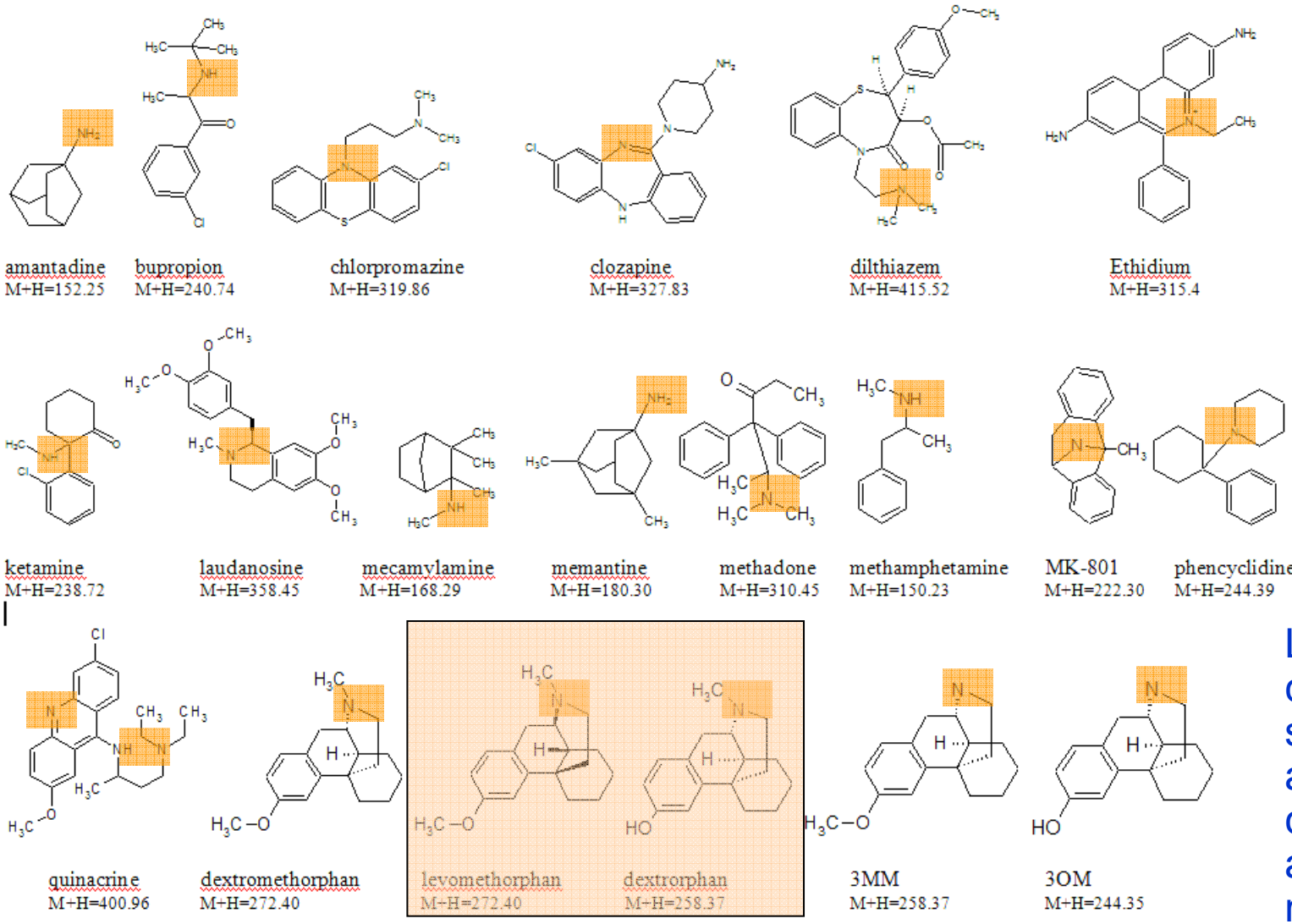
Functional nAChR

6/21/2006

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

NCI ligands

K.Jozwiak, S. Ravichandran, J.R. Collins and I.W. Wainer, *J. Med. Chem*, Jul 29, 47:16 4008-4021, **2004**



Ligands (neutral & charged) built using SYBYL. Charges were added using semi-empirical procedure (AM1/Mulliken)

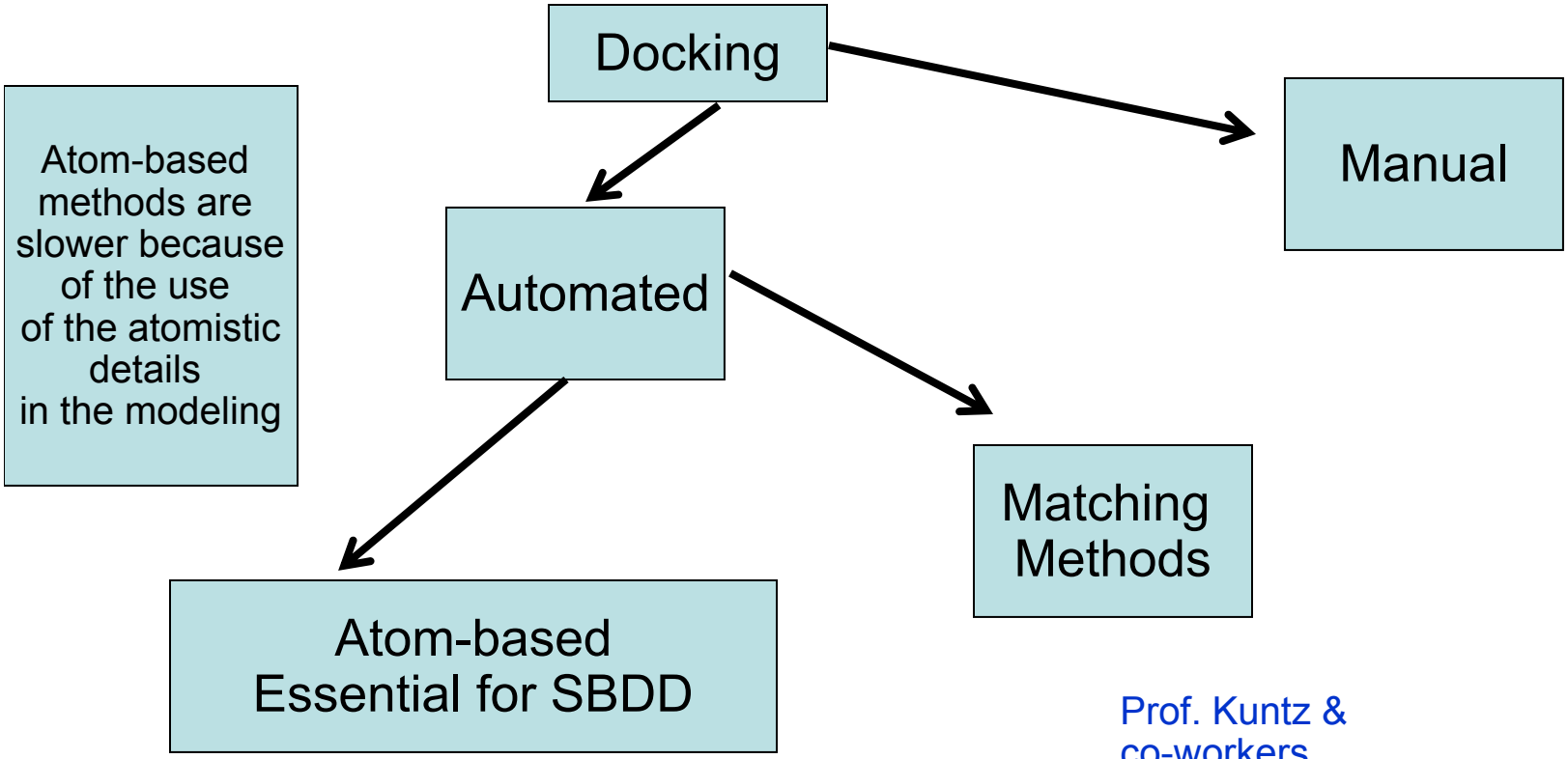
Lumen is a cation selector-NCIs are +vely charged with a amine moiety

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

- Problem
  - To find molecular binding sites (hot-spots) by computer
- Why is it difficult?
  - 2 Rigid molecules
    - 6 degrees of freedom
      - Orientational(Quaternion:3), translational(x,y,z:3)
  - If one of them is flexible
    - » 6 + # of torsions (n)
  - Drug molecule
    - flexible (14 degrees of freedom)
      - 1028 variations

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Different approaches in Docking



Atom-based methods are slower because of the use of the atomistic details in the modeling

AutoDock and others, Scripps Research Institute, California

Prof. Kuntz & co-workers, Univ of California San Francisco

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Docking Space

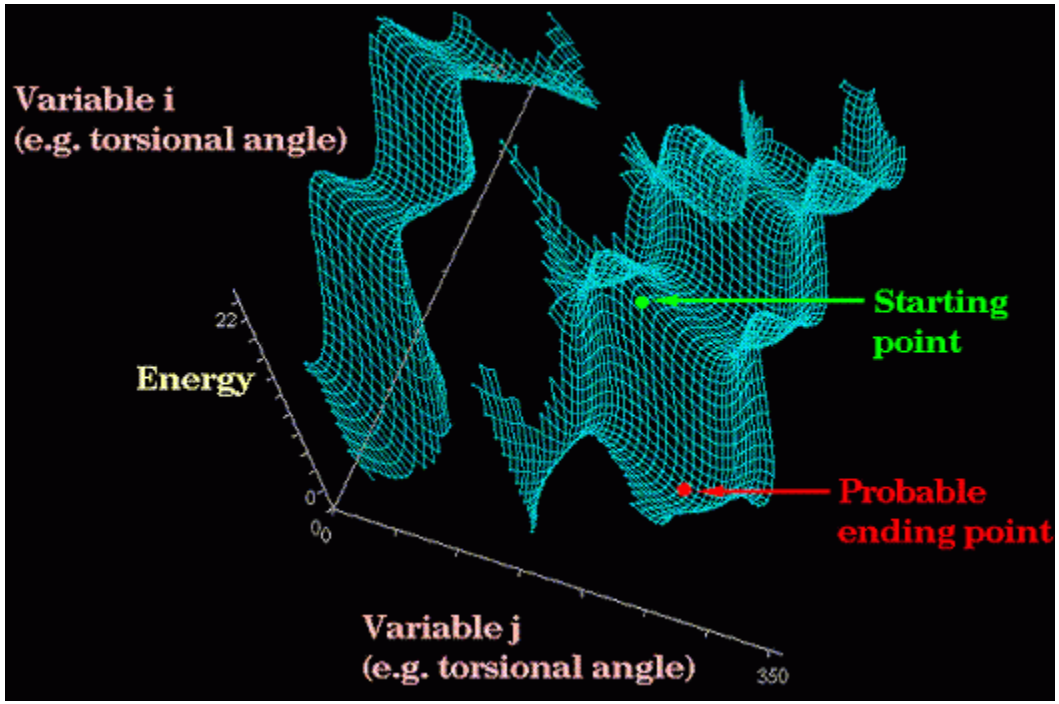
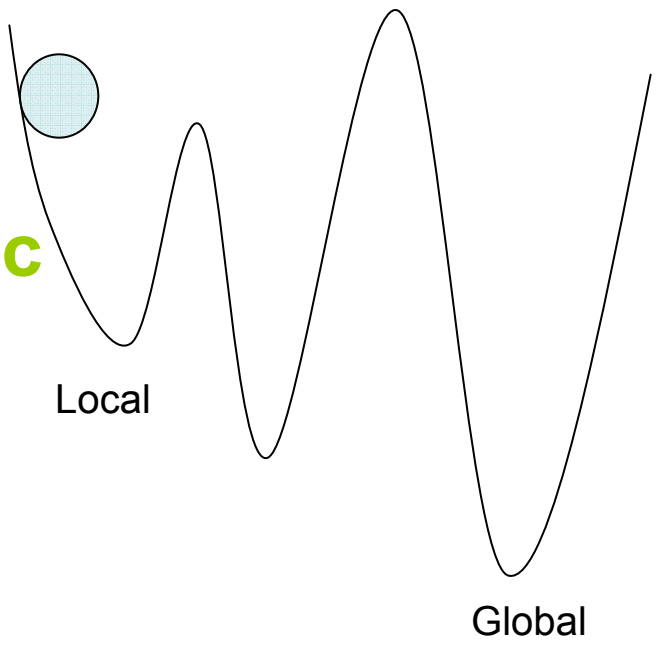


Image taken from NIH Molecular Modeling page

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Search Methods in Docking

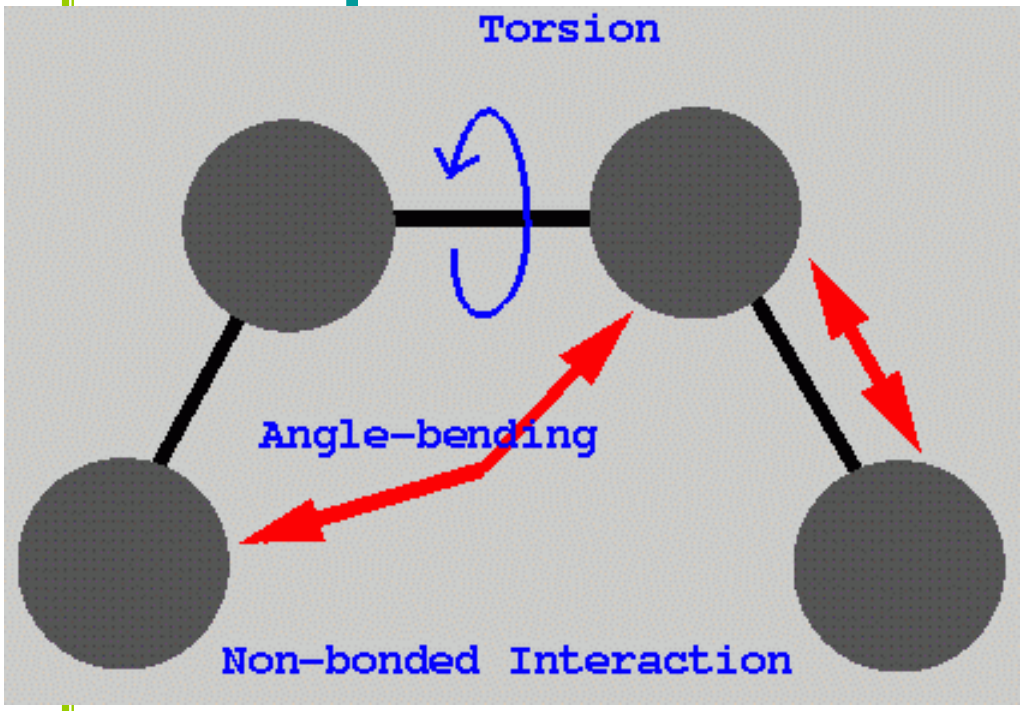
- **Systematic Search**
  - **Exhaustive, Expensive**
- **Stochastic**
  - **Random, non-deterministic**
  - **Simulated Annealing**
  - **Evolutionary Algorithm**
    - **Genetic Algorithms**
      - **GALS**
    - **Others (Tabu etc.)**



S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

ForceField

- **ForceField is an analytical Functional form for the independent energy terms and parameters**



VDW

$$E = \sum_{i,j} (A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6}) +$$

$$\sum_{i,j} (C_{ij}/r_{ij}^{12} - D_{ij}/r_{ij}^{10}) +$$

H-bond

$$\sum_{i,j} (q_i q_j/[\varepsilon(r_{ij}) r_{ij}]$$

Electrostatic

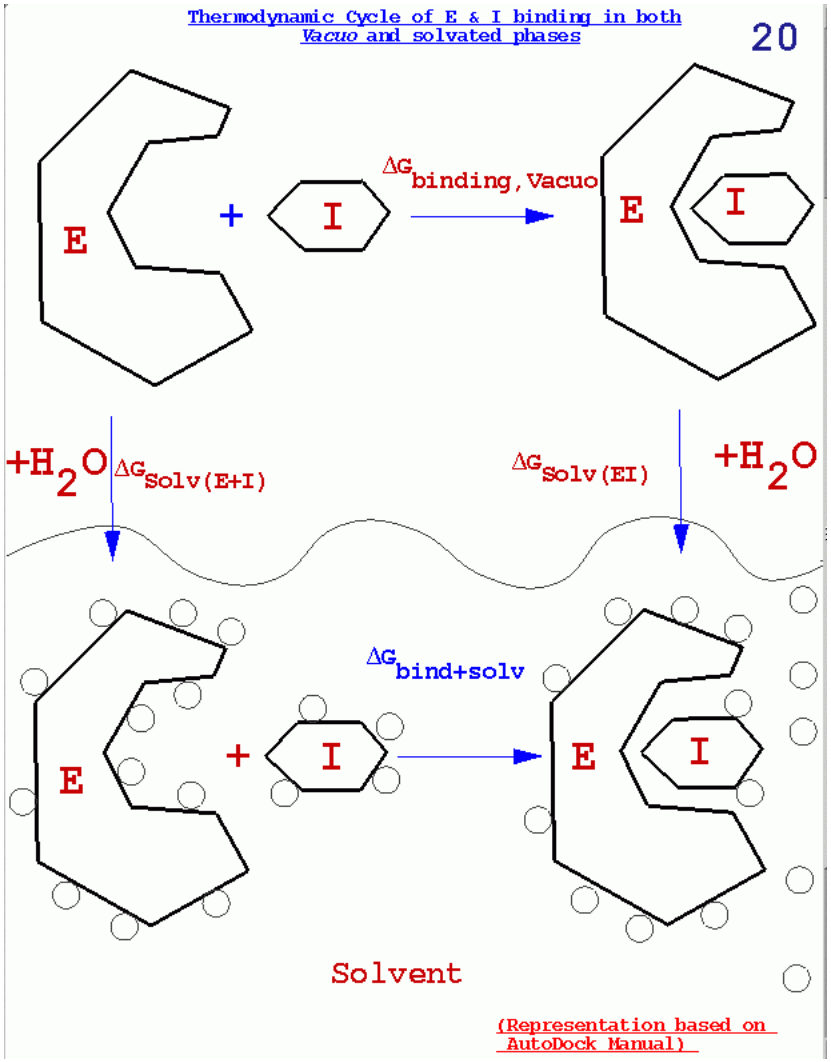
Force → $\frac{-\partial E}{\partial r_i}$

Scoring

- **Identifying the good from the bad**
  - **Several approaches**
    - **Force-Field, Empirical, Knowledge-based**
    - $\Delta G = -RT \ln K_{eq}$
    - $= -RT \ln (K_{on}/K_{off})$
    - $= \Delta H - T\Delta S$
  - **Empiricial relationship between molecular structure and binding free energy**
  - $\Delta G = K_{vdw}*V_{vdw} + K_{hb}*V_{hb} + K_{ele}*V_{ele} + K_{tor}*V_{tor} + K_{sol} * V_{sol}$
    - **Coefficients are empirically determined using linear regression analysis from a set of protein-ligand complexes with known binding constants**

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Scoring Function & Binding Constants



$$\Delta G = \Delta H - T\,\Delta S$$
$$= -RT \ln(k_{on}/k_{off})$$
$$= -RT \ln k_{eq}$$

Empirical relationship between molecular structure and binding free energy

$$\Delta G = K_{vdw}*V_{vdw} + K_{hb}*V_{hb} + K_{ele}*V_{coul} + K_{tor}*V_{tor} + K_{sol}*V_{sol}$$

Protein–ligand complex	PDB code	Log(K_i)[a]
Concanavalin A / α-methyl-D-mannopyranoside	4cna	2.00
Carboxypeptidase A / glycyl-L-tyrosine	3cpa	3.88
Carboxypeptidase A / phosphonate ZAA=P=(O)F	6cpa	11.52
Cytochrome P-450$_{cam}$ / camphor	2cpp	6.07
Dihydrofolate reductase / methotrexate	4dfr	9.70
α-Thrombin / benzamidine	1dwb	2.92
Endothiapepsin / H-256	zer6	7.22
ε-Thrombin / MQPA	1etr	7.40
ε-Thrombin / NAPAP	1ets	8.52
ε-Thrombin / 4-TAPAP	1ett	6.19
FK506-binding protein (FKBP) / immunosuppressant FK506	1fkf	9.70
D-Galactose / D-glucose binding protein / galactose	2gbp	7.60
Hemagglutinin / sialic acid	4hmg	2.55
HIV-1 Protease / A78791	1hvj	10.46
HIV-1 Protease / MVT101	4hvp	6.15
HIV-1 Protease / acylpepstatine	5hvp	5.96
HIV-1 Protease / XK263	1hvr	9.51
Fatty-acid-binding protein / C_{15}COOH	2ifb	5.43
Myoglobin (ferric) / imidazole	1mbi	1.88
McPC603 / phosphocholine	2mcp	5.23
β-Trypsin / benzamidine	3ptb	4.74
Retinol-binding protein / retinol	1rbp	6.72
Thermolysin / Leu-hydroxylamine	4tln	3.72
Thermolysin / phosphoramidon	1tlp	7.55
Thermolysin / *n*-(1-carboxy-3-phenylpropyl)-Leu-Trp	1tmn	7.30
Thermolysin / Cbz-Phe-*p*-Leu-Ala (ZFpLA)	4tmn	10.19
Thermolysin / Cbz-Gly-*p*-Leu-Leu (ZGpLL)	5tmn	8.04
Purine nucleoside phosphorylase (PNP) / guanine	1ulb	5.30
Xylose isomerase / CB3717	2xis	5.82
Triose phosphate isomerase (TIM) / 2-phosphoglycolic acid (PGA)	2ypi	4.82

[a] Adapted from Böhm.[54]

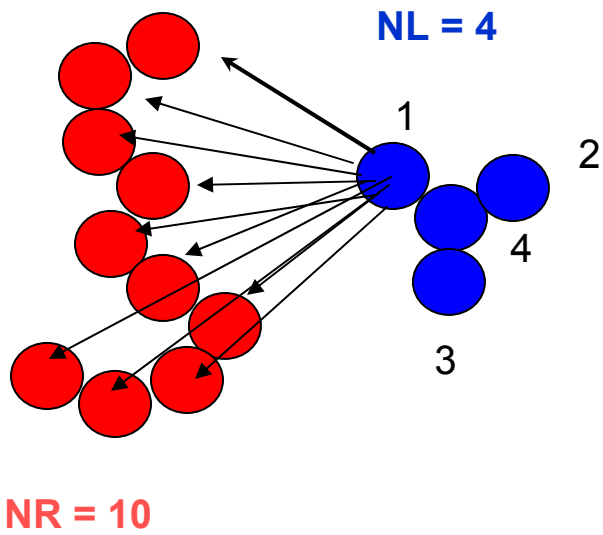
Morris et al. J. Comp. Chem 19(14) 1639-1662 (1998)

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Flexibility

- Ligand
  - Systematic search
    - (6 rotatable bonds with 30 degree increments result in 2,985,984 conformations)
  - Random search
    - Simulated annealing, Genetic Algorithm
  - Simulation Method
    - MD
- Protein
  - Molecular Dynamics
  - Ensemble of protein structures-Docking

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick



Why Grids?

Energy Calculations

11 12 13 14 15 16 17 18 19 110
21 22 23 24 25 26 27 28 29 210
31 32 33 34 35 36 37 38 39 310
41 42 43 44 45 46 47 48 49 410

Total terms = NR * NL

= 40 per simulation step

Biotin has 19 atoms. Receptors are of the range 1000 atoms or more

1000 * 19 = 19000 Terms to deal with each step !!

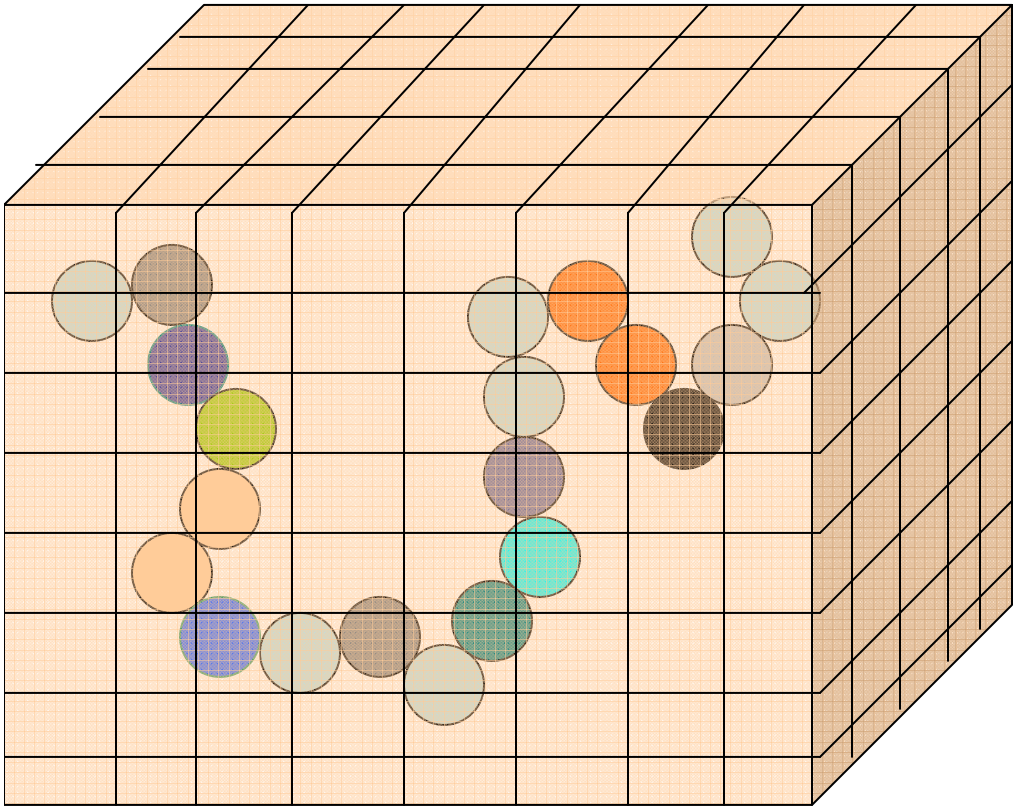
Programs like AutoDock usually do a Million Energy evaluation / simulation

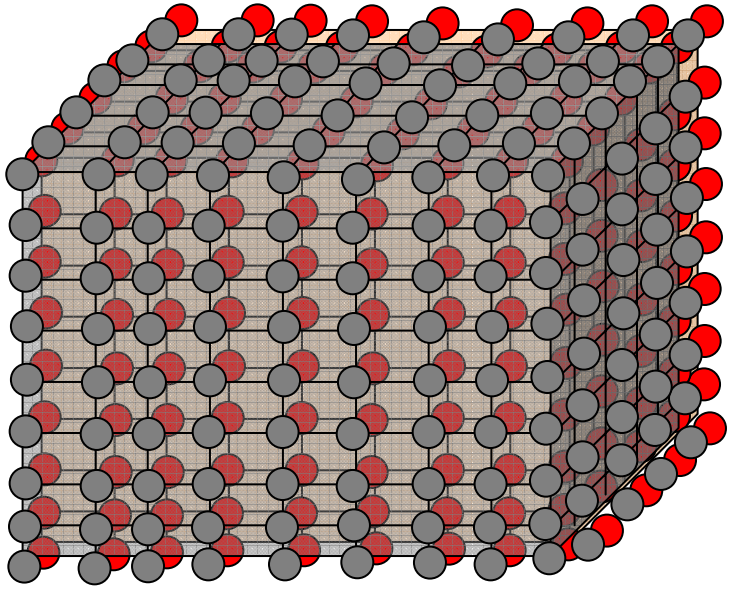
19000 * Million !!

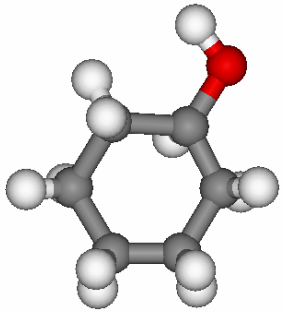
How to avoid this computational heaviness without losing molecular information?

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Grid Maps







H

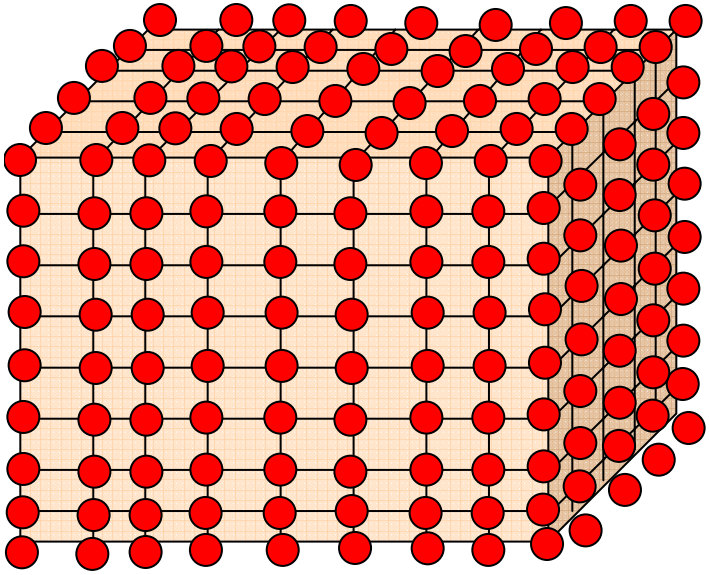
C

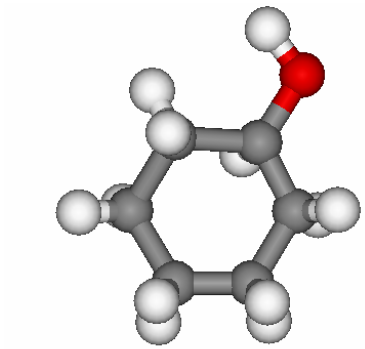
O

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

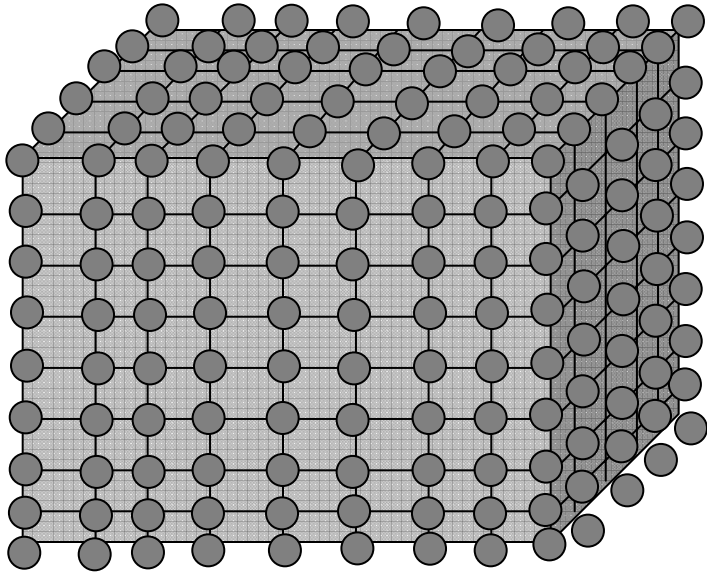
Grid Maps

Oxygen Map

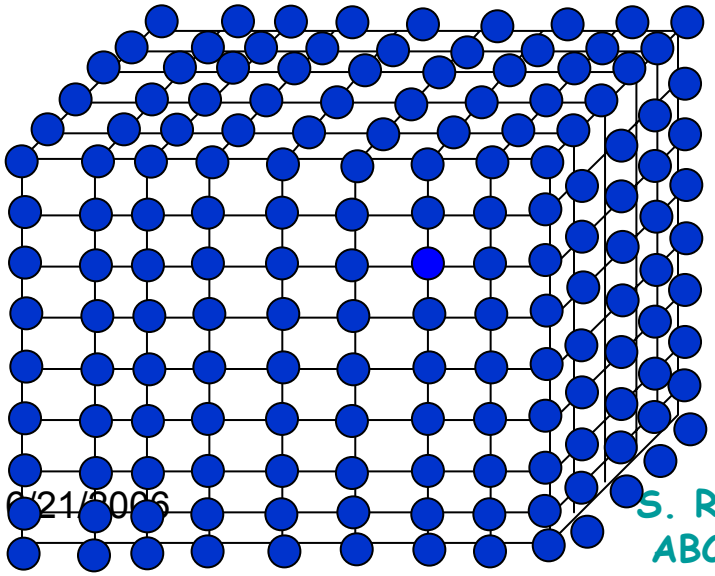




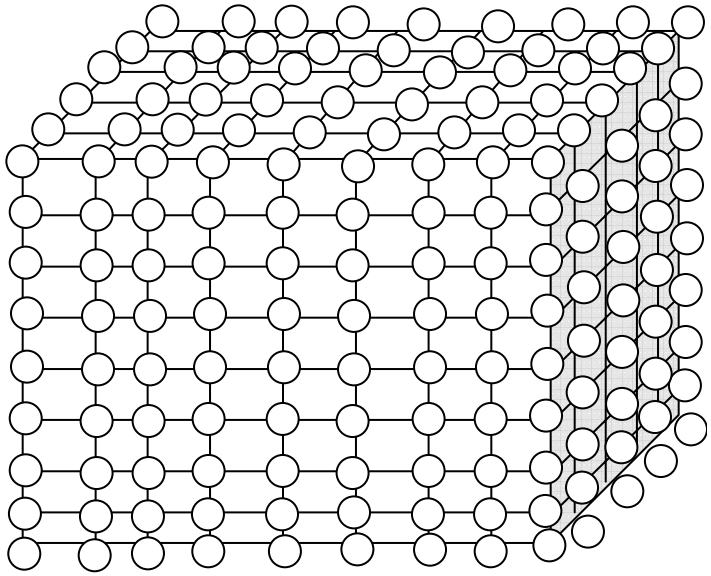
Carbon Map



Electron Map



Polar H Map



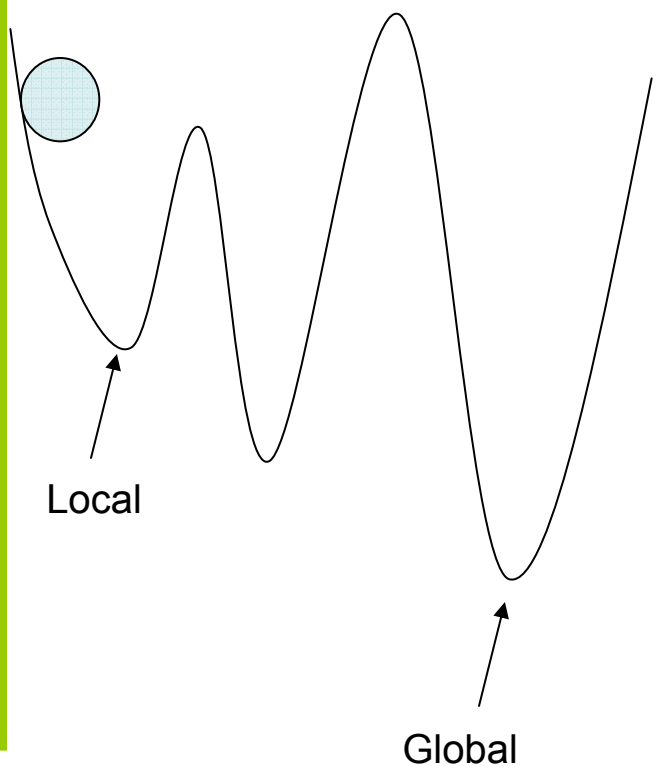
S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Methods of Docking

- Simulated Annealing Genetic Algorithms

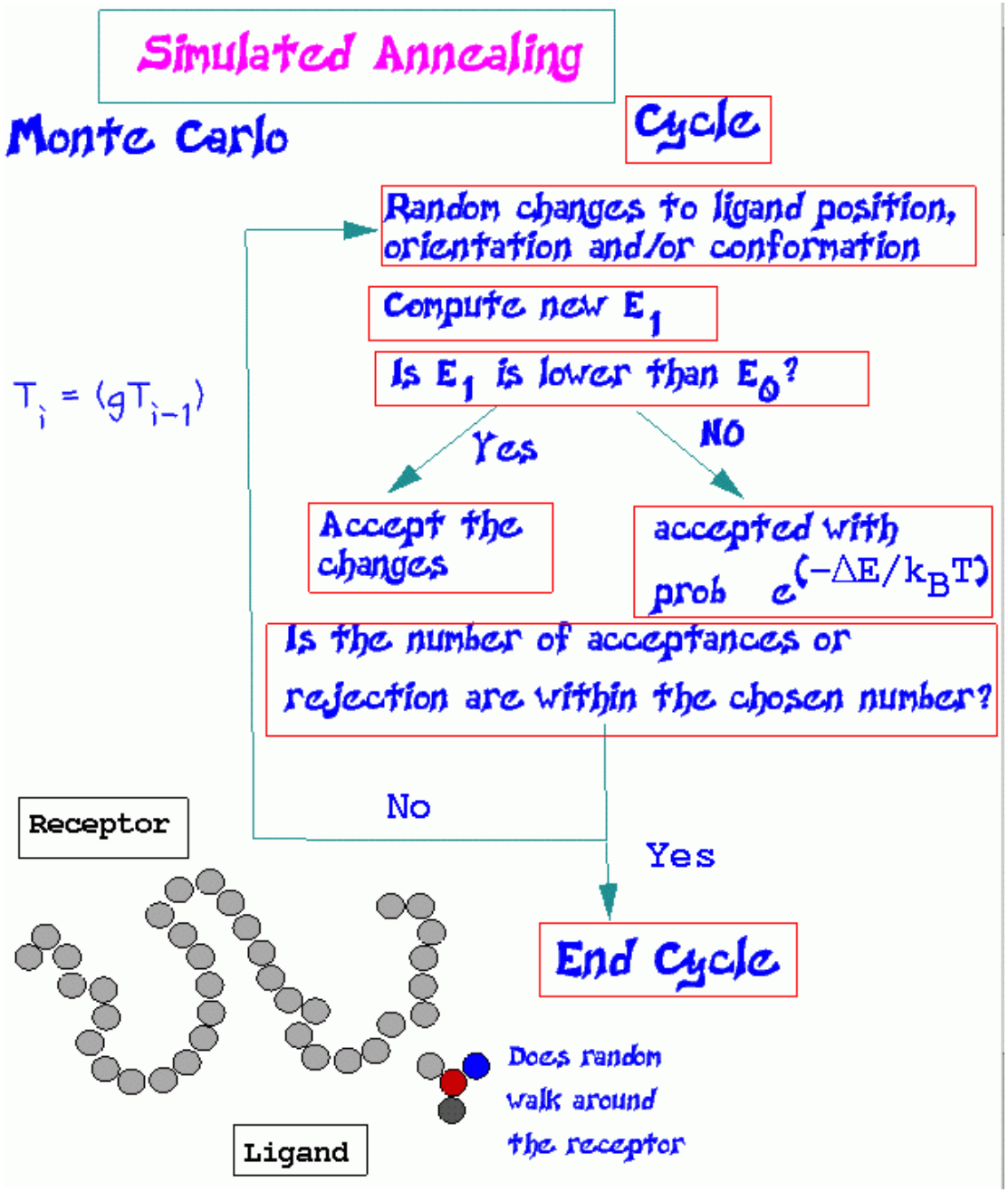
- Simulated Annealing
  - Annealing is a process in which the temperature of the substance is reduced (slowly) until the material crystallizes in a single crystal (usually correspond to global minimum free-energy)

  - Simulated Annealing: Computational method of mimicking Annealing
    - Global and Local Search
      - Global at high T & Local at low T



S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Simulated Annealing



S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Genetic Algorithms

Evolutionary Computing

Genetic Algorithm

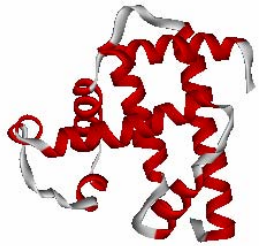
Based on Darwin theory of evolution.

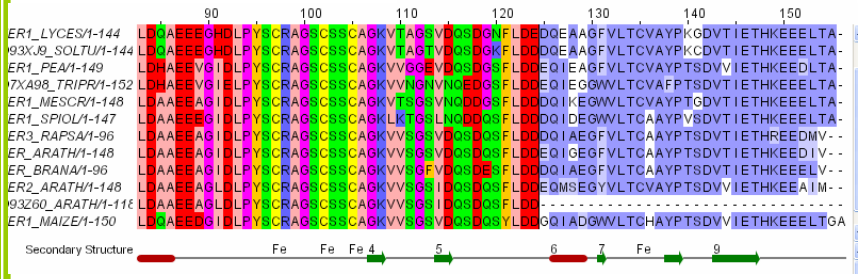
Concept: Problems are solved by evolutionary process

- **Living Organism**
  - **Made up of Cells**
    - **Has the same Set of Chromosomes (DNA)**
  - **Genome: Set of all chromosomes**
    - **Chromosomes consist of Genes**
  - **Genotype: Particular set of genes**
    - **Each gene encode a protein**
    - **Each gene encode a trait (eye color etc.)**
    - **Each gene has a position in chromosome (locus)**
  - **Reproduction**
    - **Cross-over, Mutation**

I.Rechenberg 1960)

John Holland (1975), Genetic Programming John Koza (1992)





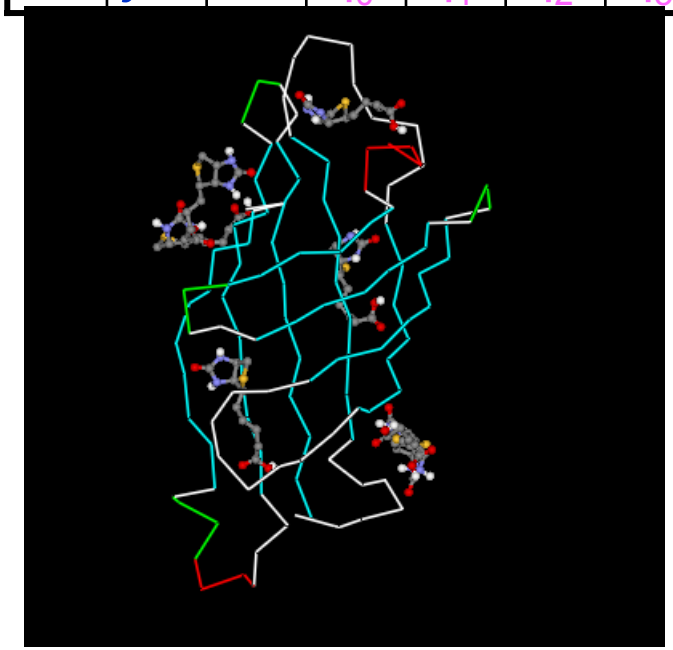
Seq Alignment Jalview 2.07

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Traditional Chromosomes are bit strings

Chromosome formed such that it has some info. About the solution it tries to solve

x	y	z	q_0	q_1	q_2	q_3	τ_1	τ_2	τ_3	τ_4	τ_5



Real Valued Genes

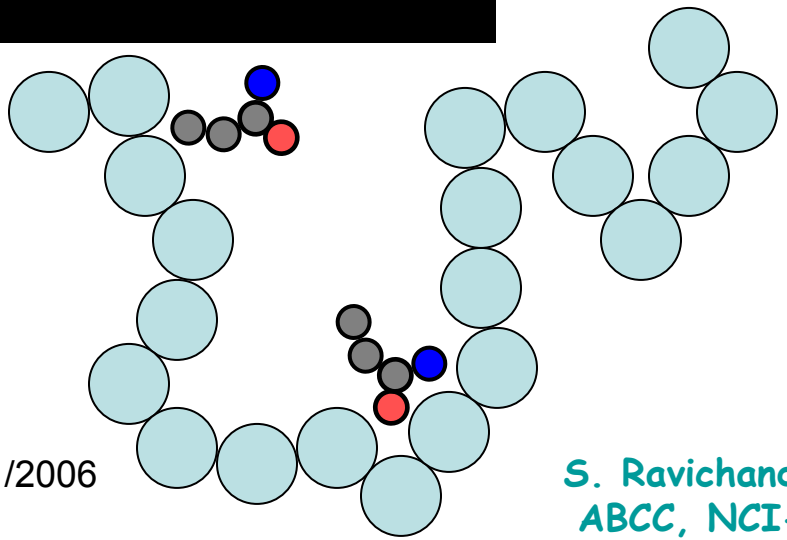


Figure Made with DSView erPro

6/21/2006

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

What are Chromosomes?

Set of Solutions Populations or Chromosomes

How?

What are Selection, Elitism, CrossOver Mutation?

Evaluate their fitness

Create new population Elitism, Selection Crossover, Mutation

Is Termination Condition reached?

NO

YES

STOP

Parents are perturbed to form children

GA Operators

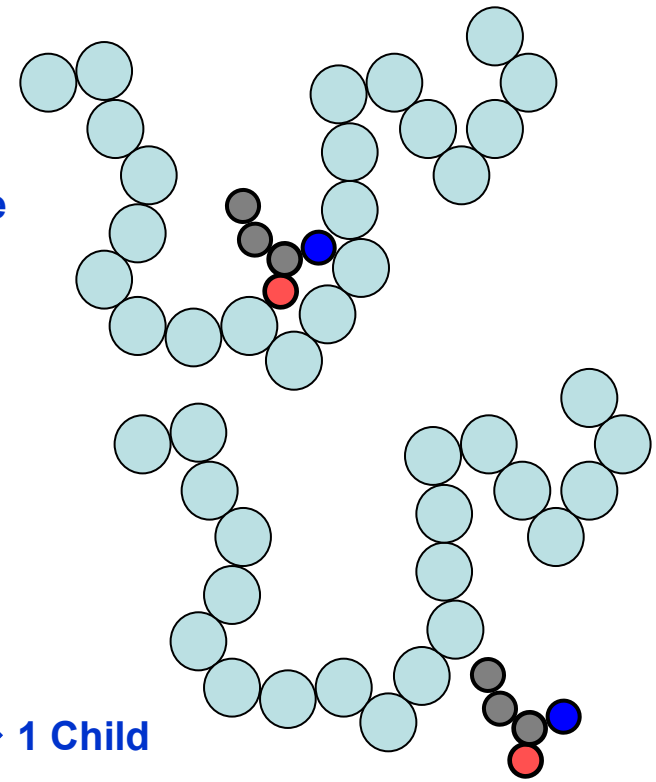
- **Mapping and Fitness**
  - Translates Genotypes to Phenotypes to calculate Energy
  - Which individuals reproduce (Fitness)
- **Elitism**
  - Survival of the best individual to the next generation
- **Mutation**
  - Random changes using Cauchy distribution

Genotype

Gene1			Gene2				Gene3				
x	y	z	q_0	q_1	q_2	q_3	τ_1	τ_2	τ_3	τ_4	τ_5

CHROMOSOME

Phenotype



1 Parent → 1 Child

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

GA Operators: Two Point Cross-Over

Parent 1

Position	Orientation	bonds Rotatable

Parent 2

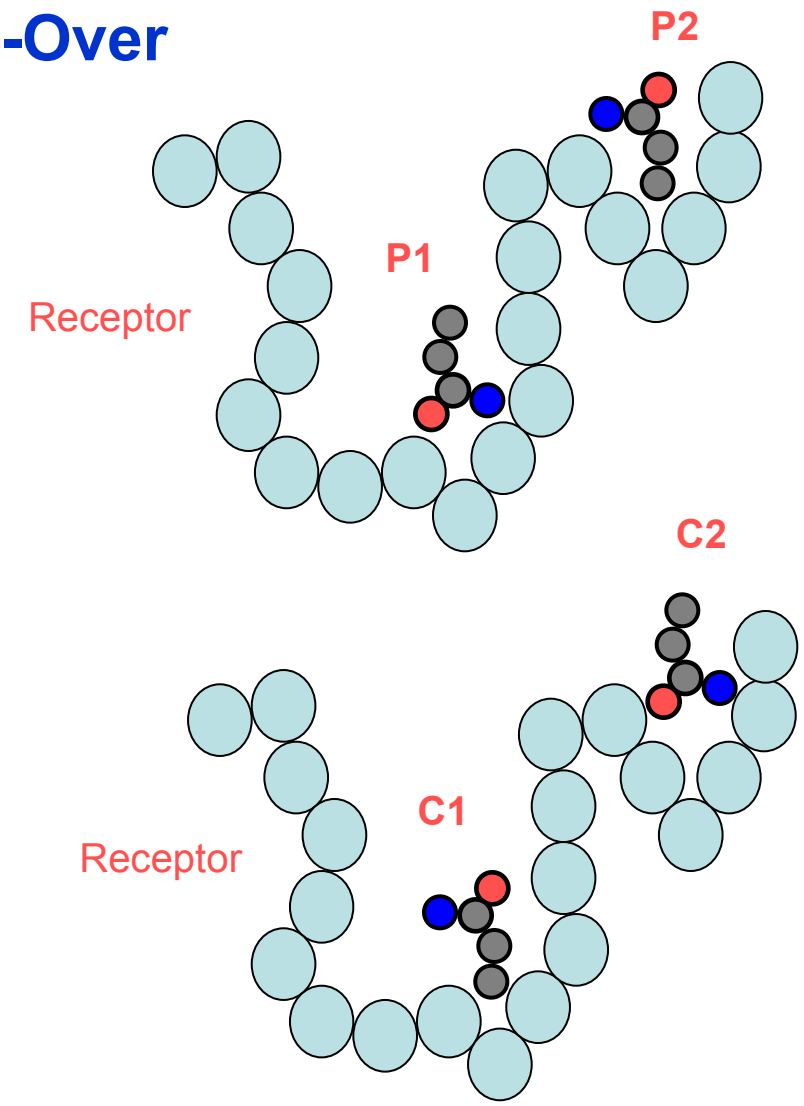
Position	Orientation	Rotatable bonds

Children 1

Position	Orientation	Rotatable bonds

Children 2

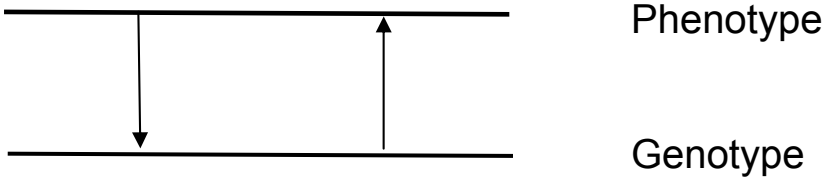
Position	Orientation	Rotatable bonds



S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

AutoDock-Lamarckian GA

- ***Environmental adaptations of an individual's phenotypic characteristics acquired during lifetime can become heritable traits (genotype)***

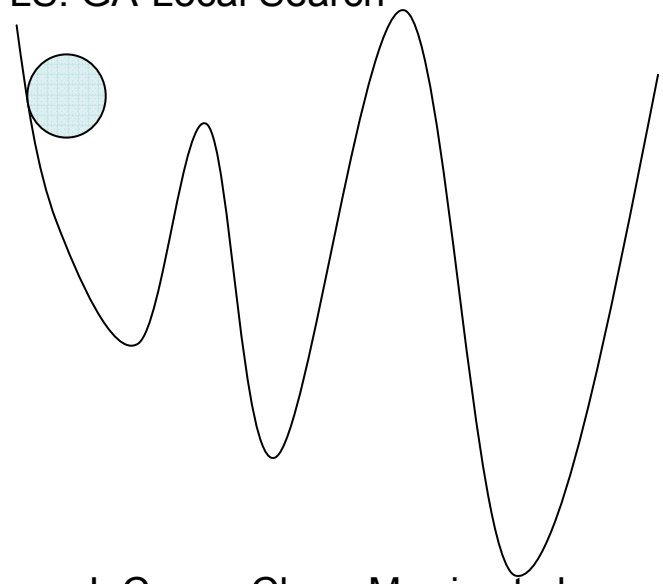


Mostly GA mimic Darwinian evolution and apply Mendelian Genetics (one way transformation from Genotype to Phenotype). In those cases (due to inverse function (P -> G) it is possible to finish GA by local search

Solis & Wets LS operator searches thro G space rather than P space (no need for inversion)

GA-LS: GA-Local Search

Normally Mutation plays the role of LS



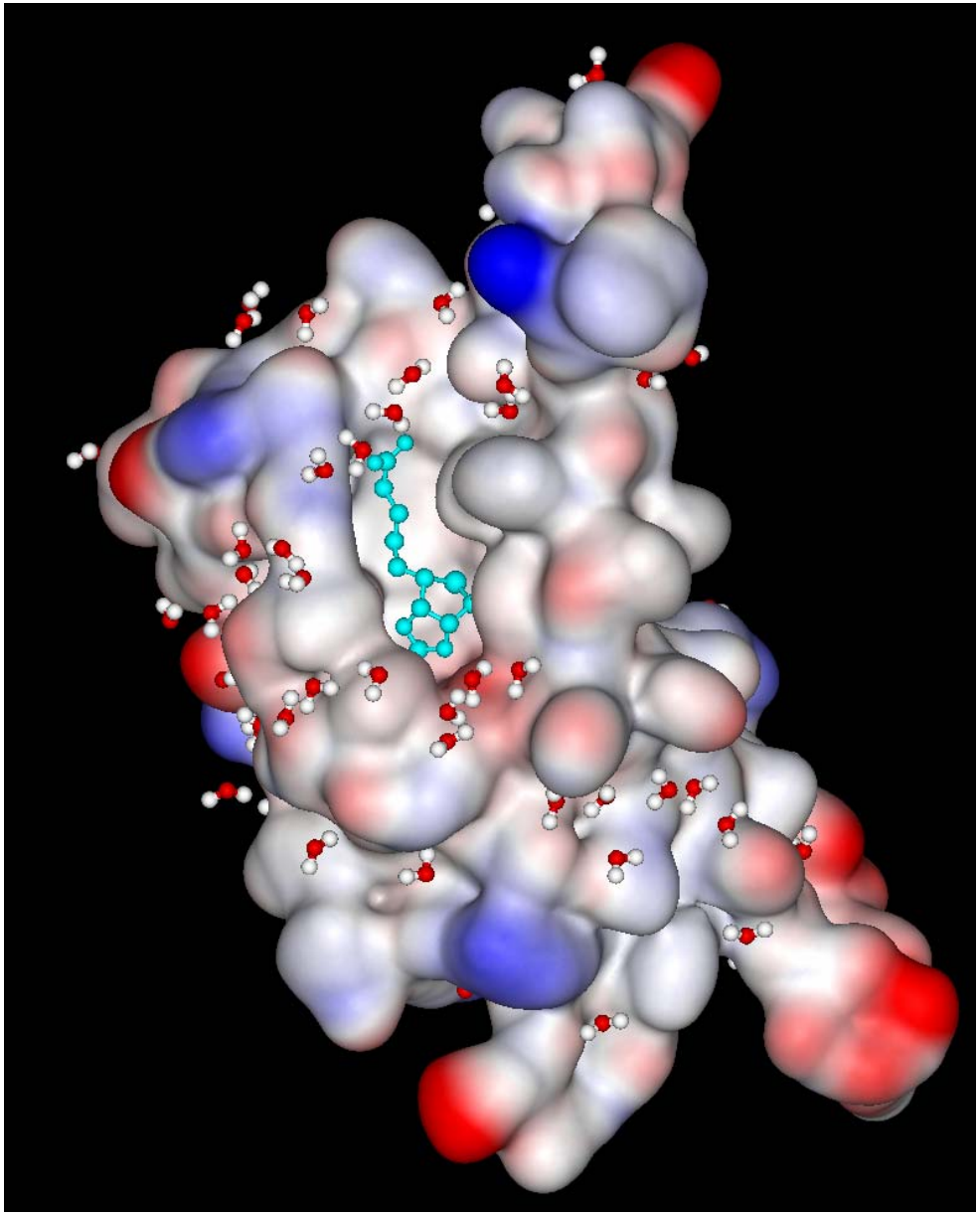
J. Comp. Chem Morris et al, 1998, 19(14), 1639

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Energy

- Docked Energy = Intermolecular energy + Intramolecular energy
  - Intermolecular energy is computed using grid maps which include desolvation term
- Free Energy (predicted) = Intermolecular energy + Torsional Free energy

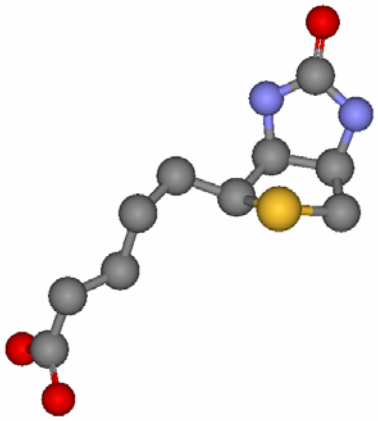
S. Ravichandran, Ph.D.
ABCC, NCI-Frederick



Test System

1STP
Streptavidin/Biotin Complex

Weber et al 1992



S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

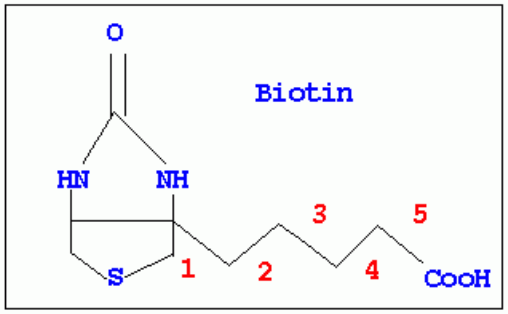
Figure Made with DSViewerPro

Test Case Docking Parameters

Test Case

Streptavidin/Biotin (1stp) (2.6 Ang Resol)

Weber et al 1989



Autogrid:

Number of points:

116, 104, 124 (x,y,z),

117 X 104 X 124 = 1508832

Grid Spacing = 0.375 Ang.

Map types C,H,N,S,O & e

CPU Time: 7.53s AutoGrid

Octane SGI workstation

Docking: LGA-LS, Population Size = 50;

Elitism = 1, Cross-over rate = 0.8

rate of mutation = 0.02, GA_num_evals 250000

Lowest Docked Energy Mean	Mean Docked Energy	Number in Cluster	Ref. RMSD	Est Free Ene. of Binding
-10.78	-10.46	6	0.39	9.03

CPU Time: 3m 25.13s

Octane SGI workstation

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

Sample Clustering Histogram

SAMPLE CLUSTERING HISTOGRAM

Clus -ter Rank	Lowest Docked Energy	Run	Mean Docked Energy	Num in Clus	Histogram 5 10 15 20 25 30 35
1	-10.27	17	-10.21	2	##
2	-10.19	20	-10.12	3	###
3	-10.14	3	-10.14	1	#
4	-10.05	4	-10.05	1	#
5	-9.05	7	-9.05	1	#
6	-8.37	13	-8.37	1	#
7	-8.18	1	-8.18	1	#
8	-8.07	9	-8.07	1	#
9	-7.80	12	-7.80	1	#
10	-7.47	10	-7.47	1	#
11	-7.41	6	-7.41	1	#
12	-7.40	11	-7.40	1	#
13	-7.24	18	-7.24	1	#
14	-7.15	15	-7.15	1	#
15	-7.07	5	-7.07	1	#
16	-6.90	2	-6.90	1	#
17	-6.82	8	-6.82	1	#

Estimated Free Energy of Binding = - 8.11kcal/mol
Observed = -18.27kcal/mol

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Bound Conformation Violet --
Lowest Energy Conformation Red

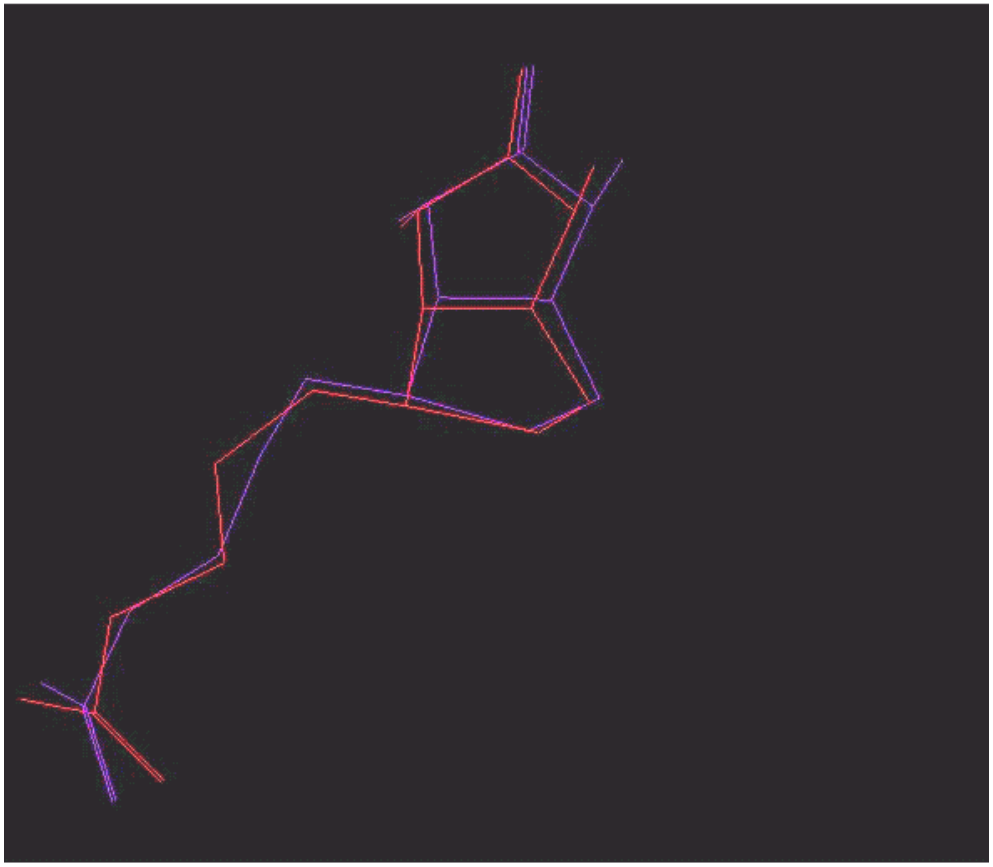


Figure Made with Sybyl 6.8

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

Issues to watch-out in GA

- Premature convergence
- Too long for convergence
  - How many runs?
  - How many energy evaluations?
- Choice of parameters
  - Cross-over rate: 80-90%
  - Mutation rate: 0.5-1%
  - Population size: 100-200
    - depends on the problem

Reference:

Csaba Hetenyi, David van der Spoel

FEBS Letters, 580 (2006) 1447-1450

Protein Science, 11, 1729-1737

http://cs.felk.cvut.cz/~xobitko/ga/

Limitations

- Simulated Annealing
  - More than 8 torsions
- LGA
  - 20-30 torsions
- AutoDock
  - No Structure available
    - Homology Modeling
  - Poor quality structures
    - Resolution numbers, incomplete models
  - Ligand too flexible
    - Too many torsions
  - Induced Fitting (protein flexibility)
    - Multiple dockings starting from different conformer models
      - NMR

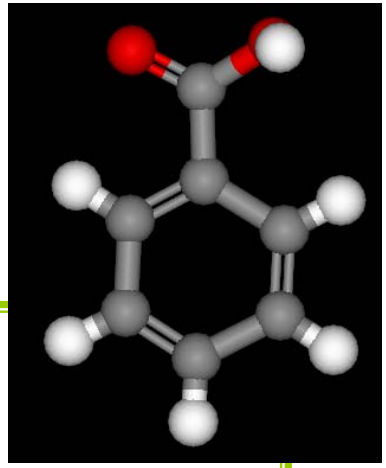
Morris et al, J. Comp. Chem, Vol 19, 1639-1662 (1998)

AutoDock Run Setup Using AutoDockTools

- Biomolecule (protein/DNA/RNA)
  - PDB
    - Add essential H, Fix charges, Kollman Charges (AMBER), Save the molecule as MOL2 file
  - Small molecules
    - Partial Charges
      - Gasteiger, MOPAC, Gaussian
    - Add all Hs
      - Save as MOL2

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

AutoTors



- Read the Ligand molecule
  - Assign the root
  - Setup the tree and branches
  - Assign the rotatable bonds (torsions)
  - Rename Aromatic Carbons as 'A'
    - aliphatic 'C'
  - Merge non-polar Hs

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Grid Parameter File

- Read the macromolecule
  - PDBQ, PBDQS
- Set map types
  - based on the ligand file
- Set up the GRID box
- Write GPF (Grid Parameter File)

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

Docking Parameter File

- Choose the macromolecule
- Choose the ligand
- Choose the docking algorithm
  - GA-LS, GA, LS, SA
- Docking Run parameters
- Write DPF file

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

AutoDock Runs

- Start AutoGrid
- Start AutoDock
- Use AutoDockTools or other software to analyze the results
  - DLG (Docking Log File)
- weblinks
  - http://w3.to/autodock (AutoDock)
    - A.J. Olson and co-workers
  - http://www.scripps.edu/~sanner/python/adt/
    - AutoDockTools: Dr. Michel Sanner and co-workers

S. Ravichandran, Ph.D.
ABCC, NCI-Frederick

THANK YOU

S. Ravichandran

http://ncisgi.ncifcrf.gov/~ravichas

Advanced Biomedical Computing Center

National Cancer Institute

Frederick, MD 21701