

Molecular Mechanics & Dynamics

Introduction to Structural Bioinformatics

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (1)

Outline

- o Applications of Computed Structure
- o Molecular Mechanics
 - o Equations
 - o Minimization
 - o Limitations
- o Molecular Dynamics
 - o Algorithms
 - o Applications
 - o Free energy simulations
- o MD Examples:
 - o MD simulation of coiled coils
 - o Relative binding constants for HIV protease inhibitors

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (2)

Why Compute Structure?

- o HAVE: information about composition
WANT: two (connectivity) or three dimensional structure
- o HAVE: structure and physico-chemical properties
WANT: rationalization of their relationship
- o HAVE: information about the relationship
WANT: predict new structures/properties

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (3)

Central Postulate

Structure \equiv Energy

- o Any geometrical system has an associated Energy determined by its coordinates
- o A fundamental description includes a multidimensional potential energy surface
- o Minima on this surface correspond to stable configurations

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (4)

Computed Structure

- o General Problem: Locate the energy minima starting from arbitrarily chosen configuration
- o Need: mathematical description of structure/energy relationship
- o Structure: Cartesian or internal coordinates

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (5)

Outline

- o Applications of Computed Structure ✓
- o Molecular Mechanics
 - o Equations
 - o Minimization
 - o Limitations
- o Molecular Dynamics
 - o Algorithms
 - o Applications
 - o Free energy simulations
- o MD Examples:
 - o MD simulation of coiled coils
 - o Relative binding constants for HIV protease inhibitors

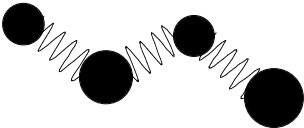
Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (6)

Molecular Mechanics

- o Definition: A mathematical model to predict accurate structures and energies of molecules
- o Basic Concept: optimize the geometry of a molecule (computed bonds and bond angles) to adopt "natural" meanings and values

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (7)

MM View of Molecules



Molecules: masses joined by springs

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (8)

MM: Simple Chemical View

- o Molecules consist of atoms
- o Atoms are bonded to each other according to the rules of valence (bonded terms)
- o Molecules are flexible (torsional terms)
- o Non-bonded atoms interact with each other (non-bonded terms)

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (9)

MM Formula

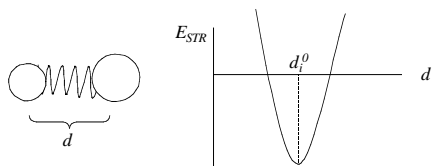
$$E = \sum E_{STR} + \sum E_{BEND} + \sum E_{OOP} + \sum E_{TOR} + \sum E_{VDW} + \sum E_{ELE}$$

STR = stretching; BEND = bending; OOP = out of plane bending
 TOR = torsional; VDW = van der Waals; ELE = electrostatic

The MM energy expression presents the relationship between Chemical structure and its energy.

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (10)

Bond Stretching Energy

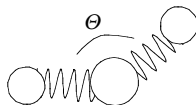


$$E_{STR} = \sum_{i=1}^n \frac{1}{2} k^{d_i} (d_i - d_i^0)^2$$

d_i = length of i^{th} bond
 d_i^0 = equilibrium length of i^{th} bond
 k^{d_i} = bond stretching force constant

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (11)

Angle Bending Energy



$$E_{BEND} = \sum_{i=1}^n \frac{1}{2} k^{\theta_i} (\theta_i - \theta_i^0)^2$$

θ_i = angle between two adjacent bonds
 θ_i^0 = equilibrium value for the i^{th} angle
 k^{θ_i} = angle bending force constant

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (12)

Out of Plane Bending Energy

$$E_{OOP} = \sum_{i=1}^n \frac{1}{2} k_i^{OOP} d_i^2$$

d_i = distance between center atom and plane
 k^{OOP} = out of plane bending constant

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (13)

Torsional Energy

$$E_{TOR} = \frac{V_1}{2}(1 + \cos \omega) + \frac{V_2}{2}(1 - 2 \cos \omega) + \frac{V_3}{2}(1 + 3 \cos \omega)$$

V_1, V_2, V_3 = torsional barriers
 α = torsion angle

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (14)

van der Waals Energy

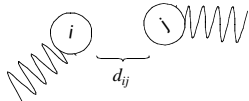
$$E_{VDW} = \sum_{i=1}^n \sum_{j=i+1}^n \epsilon \left[\left(\frac{d_{ij}}{R_i + R_j} \right)^{-12} - 2 \left(\frac{d_{ij}}{R_i + R_j} \right)^{-6} \right]$$

ϵ = van der Waals constant
 d_{ij} = distance between atoms i and j
 R_i, R_j = van der Waals radius of atoms i and j

$$E_{VDW} = \epsilon \left[\left(\frac{d_0}{d} \right)^{12} - 2 \left(\frac{d_0}{d} \right)^6 \right]$$

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (15)

Electrostatic Energy



$$E_{ELE} = \sum_{i=1}^n \sum_{j>i} \frac{Q_i Q_j}{4\pi\epsilon d_{ij}}$$

ϵ = dielectric constant
 d_{ij} = distance between atoms i and j
 Q_i, Q_j = net atomic charge at atoms i and j

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (16)

How to find a stable structure

- o Potential energy surface = multidimensional surface describing the energy of a molecule in terms of the nuclear positions
- o Algorithm:
 - o deform (twist, bend, stretch, pull...)
 - o calculate energy changes
 - o mathematically optimize (steepest descent, conjugate gradient...)
- o Result: nearest (local) energy minimum

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (17)

How to find a global minimum

- o Multi minimum problem: minimization only proceeds to lower energy and thus only finds minimum closest to starting point
- o Need to start from many independent (random) locations
- o Algorithms include Monte-Carlo simulations e.g.

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (18)

MM Optimization

$$E = \sum E_{STR} + \sum E_{BEND} + \sum E_{OOP} + \sum E_{TOR} + \sum E_{VDW} + \sum E_{ELE}$$

- o Structure and Energy are equivalent
- o Optimize Geometry & minimize Energy (don't mix them up!!)
- o Task: find stable conformers starting from arbitrary geometry
- o Molecular properties associated with ground state: heat of formation, vibrational spectra, shape (somewhat)

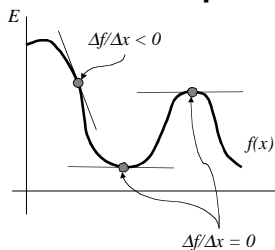
Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (19)

MM Optimization [2]

- o Iterative geometry optimization always aims at lowering potential energy
- o Minimum energy depends on starting geometry
- o No known GENERAL method to find global minimum; only local minima closest to starting geometry are found

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (20)

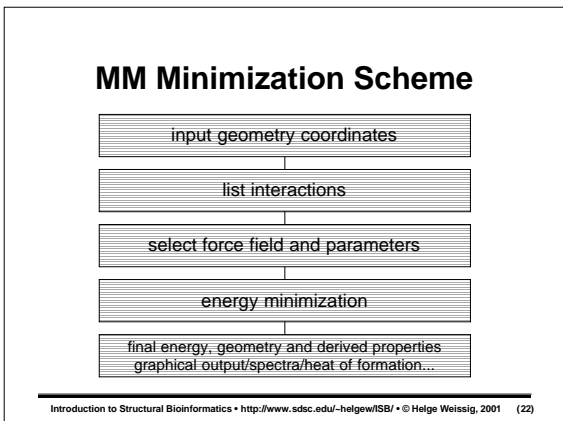
MM Optimization [3]

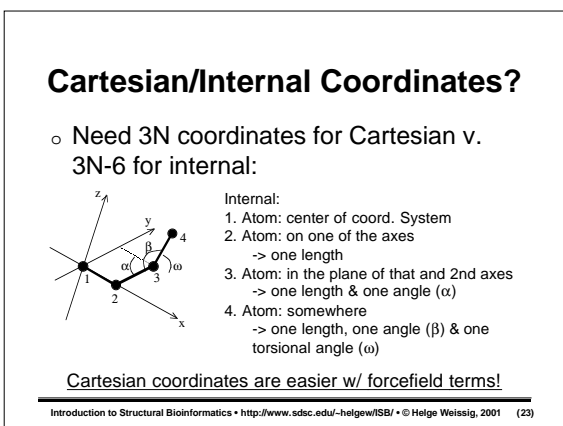


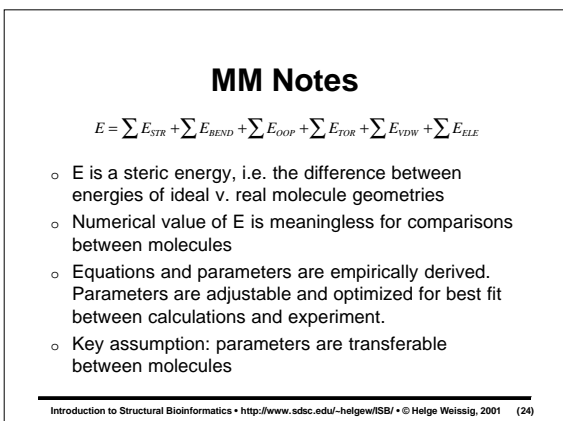
Numerically, finding energy minima involves identification of point(s) on the potential energy surface where the first derivatives are zero

This is true for local maxima as well!!!

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (21)







MM Limitations

- o MM will not:
 - o Provide parameters
 - o Calculate "natural" MM values (bonds, angles, charges ...)
 - o Handle unstable structures (e.g. transition states)
 - o Simulate structure modifications (chemical reactions)
 - o Account for conformation dependent charges
 - o Generally account for effects of re-distribution of electrons around nuclei.

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (25)

MM Cans & CANNOTs

<ul style="list-style-type: none"> o Describe energy as a function of coordinates (i.e. enthalpy) o Allow atomic motion (minimization) but always downhill on the PE surface o Describe conformations and conformers 	<ul style="list-style-type: none"> o Describe energy as a function of time and temperature (i.e. entropy and free energy) o Allow atoms move uphill the PE surface o Describe conformational states (more realistic!) o Include important parameters of "molecular life": temperature, atomic mass, time
---	--

Related to molecular motion!!

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (26)

Outline

- o Applications of Computed Structure ✓
- o Molecular Mechanics ✓
 - o Equations
 - o Minimization
 - o Limitations
- o Molecular Dynamics
 - o Algorithms
 - o Applications
 - o Free energy simulations
- o MD Examples:
 - o MD simulation of coiled coils
 - o Relative binding constants for HIV protease inhibitors

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (27)

Molecular Dynamics

- o Definition: computational technique that studies the motions of atoms and molecules that occur due to their interactions
- o Calculates fluctuations of atomic positions as a function of time
- o Obtains macroscopic properties of a system from microscopic interactions
- o "Snapshot": set of atomic coordinates at a certain moment during the simulation
- o "Trajectory": set of consecutive snapshots

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (28)

Motions in "real life"

- o Kinetic properties of liquids and solids (also cellular membranes!)
- o Reaction mechanisms
- o Solubilities
- o Thermodynamic properties
- o Binding
- o Folding of proteins, polymers, nucleic acids
- o MD: extension of classical MM modeling by application of Newtonian laws

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (29)

Protein Dynamics

t [s]

10⁻¹⁴ } Relative vibration of bonded atoms

10⁻¹² } Rotation of side chains at protein surface

10⁻⁹ } Torsional vibration of buried groups

10⁻⁶ } Relative motion of different globular regions

10⁻³ } Rotation of medium-sized side chains in protein interior

1 } Local denaturation

10

McCammeon & Harvey, 1987: Dynamics of Proteins and Nucleic Acids

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (30)

MD Success Applications

- o Structure: position, distance & orientation
-> NMR & X-ray data refinement!!!!
- o Mobility: B-factors, occupancy
- o Dynamics: vibrational frequencies, relaxation rates, diffusion
- o Thermodynamics: density, free energy, viscosity, conductance

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (31)

Basic Equations

- o New position of atom i: $r_i(t + \Delta t) = r_i(t) + v_i \Delta t$
- o Velocity of atom i: $v_i(t) = v_i(t - \Delta t) + a_i \Delta t$
- o Acceleration of atom i: $a_i = \frac{F_i}{m_i}$
- o Force at atom i:

$$F_i = \partial / \partial r_i E(r_1 \dots r_n)$$

$$E(r_1 \dots r_n) = \text{MM potential energy}$$

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (32)

Temperature in MD

- o Kinetic energy: $U_{kin} = \sum \frac{1}{2} m_i v_i^2 = \frac{3}{2} nkT$
 n = number of atoms
 k = Boltzmann constant
 T = absolute temperature
- o Temperature is maintained by scaling of velocities at each MD step
- o Typically 300K

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (33)

Temperature in MD [2]

- o Simulated annealing runs at higher T (e.g. 1000K) to enhance conformational sampling
- o Periodic cooling is used to settle into local minimum energy conformation
- o Most effective way for conformational sampling of large biopolymer systems

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (34)

MD Ensembles

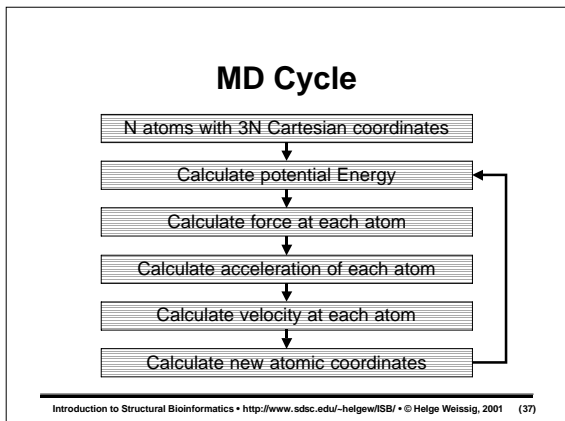
- o Constant Energy and volume (NVE)
- o Constant temperature and volume (NVT)
- o Constant temperature and pressure (NPT)

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (35)

MD Programs

- o AMBER: Weiner & Kollman, J. Comp. Chem. 2: 287 (1981)
- o CHARMM : Brooks et al., J. Comp. Chem. 4: 187 (1983)
- o GROMOS: van Gunsteren et al., PNAS 80: 4315 (1983)
- o DISCOVER: Hagler et al., J. Am. Chem. Soc. 96: 5319 (1974)
- o CEDAR: Carson & Hermans, in "Molecular Dynamics and Protein Structure", ed. J. Hermans (1985)

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (36)



MD Time Steps

- o Rule of thumb: use 1/10th of highest frequency motion (i.e. bond vibration, 10^{-14} s)
- o Resulting time steps:
 - $\Delta t = 10^{-15}$ sec = 1 fs = 0.001 ps
- o Fixed bond lengths (SHAKE method) allows increase to 2 fs

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (38)

Special Considerations

- o Long range interactions
 - o ~ n^2 non-bonded pairs
 - o List must be updated frequently (every 20-40 time steps)
 - o Truncate long range interaction with application of cutoff distance
 - o Alternatively: only allow atoms in active site to move, but still take immobile atoms' interactions into calculation!

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (39)

Special Considerations [2]

- o Solvent effects
 - o Continuum solvent model: assume $\epsilon = \epsilon_{i,j}$

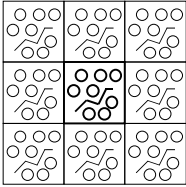
$$E_{ELE} = \sum_{i=1}^n \sum_{j>i}^n \frac{Q_i Q_j}{4\pi\epsilon d_{ij}} > E_{ELE} = \sum_{i=1}^n \sum_{j>i}^n \frac{Q_i Q_j}{4\pi d_{ij}^2}$$

- o Explicit solvent model: use specific, solvent dependent value for ϵ

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (40)

Special Considerations [3]

- o Solvent Effects (cont.)
 - o Periodic Boundary Conditions (PBC):
 - o Size of box is determined by dimensions of the solute molecule, D, and the cutoff distance R_{cutoff} : $S = D + 2R_{cutoff}$
 - o PBC allows for constant pressure if S is adjusted at each MD time step



Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (41)

Typical MD Protocol

- o Describe the system
- o Describe MD features
- o Equilibrate the system
- o Run simulation
- o Analysis

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (42)

Typical MD Protocol [2]

- o Describe the system
 - o Define atomic positions
 - o Define system topology (connectivity)
 - o Define potential function (parameter file)
 - o Add missing atoms (H, hetero atoms...)
 - o Optimize geometry
 - o Add solvent, counter ions ...

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (43)

Typical MD Protocol [3]

- o Describe MD features
 - o Time step
 - o # of time steps (length of simulation)
 - o Frequency of non-bonded list update
 - o Conditions (T, V/P)
 - o Initial velocities
 - o Special considerations (PBC...)

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (44)

Typical MD Protocol [4]

- o Equilibrate the system
 - o Equilibrate solvent with solute
 - o Define # of heating and equilibration steps
 - o Monitor system properties (PE, KE, T, P) during heating and equilibration for convergence

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (45)

Typical MD Protocol [5]

- o Run simulation
 - o Specify save frequency for coordinates and velocities
 - o Monitor evolution of KE, PE, T, P
 - o Monitor desired parameters (RMSD, important distances...)

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (46)

Typical MD Protocol [6]

- o Analysis
 - o Structure stability
 - o Important molecular motions
 - o Average structure
 - o thermodynamics

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (47)

MD limitations

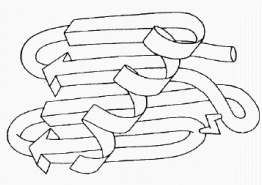
- o Very small time steps (1-2 fs)
- o Total simulation time 300 - 500 ps
- o Limited exploration of conformational space
- o Accumulation of numerical errors during simulation run

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (48)

MD limitations [2]

- o In general: MD modeling is good at predicting unrealistic structures

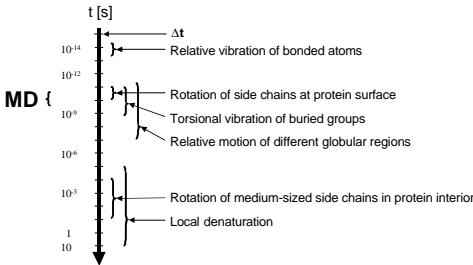
Incredulase



Richardson & Richardson: "Some design principles: Betabellin" in "Protein Engineering" (Oxender & Fox eds., 1987)

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (49)

MD in Protein Dynamics



Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (50)

Outline

- o Applications of Computed Structure ✓
- o Molecular Mechanics ✓
 - o Equations
 - o Minimization
 - o Limitations
- o Molecular Dynamics ✓
 - o Algorithms
 - o Applications
 - o Free energy simulations
- o MD Examples:
 - o MD simulation of coiled coils
 - o Relative binding constants for HIV protease inhibitors

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (51)

Coiled Coil Challenge

- o Design a dynamically stable, shorter coiled coil
- o Three-heptad, 22 residue coiled coil (P44)
- o Disulfide bridges stabilize N- and C-termini
- o MD simulations are run on crystallographic GCN4 structure and *de novo* designed structure
- o Role of disulfide bridges is analyzed by MD simulations after "cleavage" (deletion of connectivity records)

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (55)

Coiled Coil References

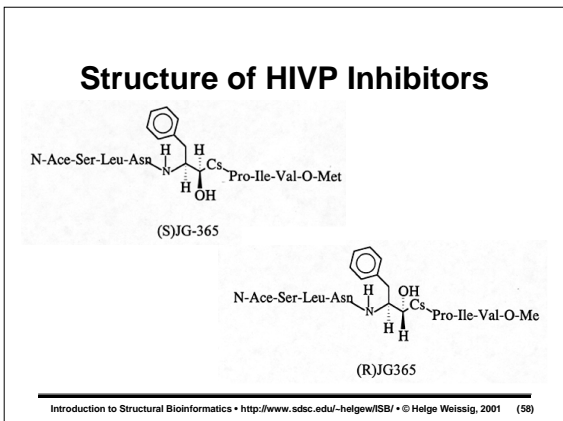
- o Theory:
 - o Crick, *Acta Cryst.* **6**: 685 (1953)
 - o Pauling & Corey, *Nature* **171**: 59 (1953)
- o Experiments:
 - o Zzta (O'Shea et al.)
- o Computer Simulations:
 - o Tropsha et al., *PNAS* **88**: 9488 (1991)
 - o Krystek et al., *Int. J. Peptide Protein Res.* **38**: 229 (1991)
 - o Nilges & Brunger, *Protein Eng.* **4**: 649 (1991)
 - o Treutlein et al., *Biochemistry* **31**: 12726 (1992)
 - o Rozzelle et al. *Protein Sci.* **3**: 345 (1994)

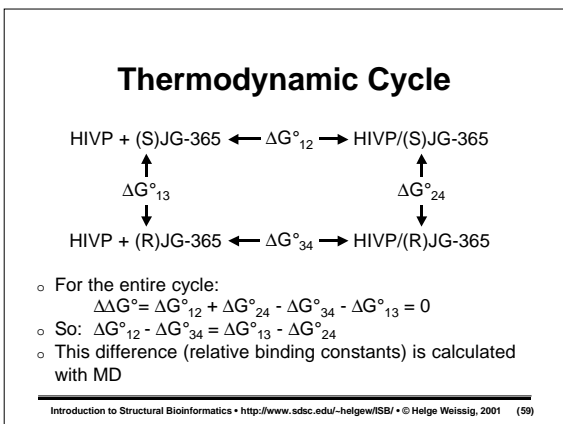
Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (56)

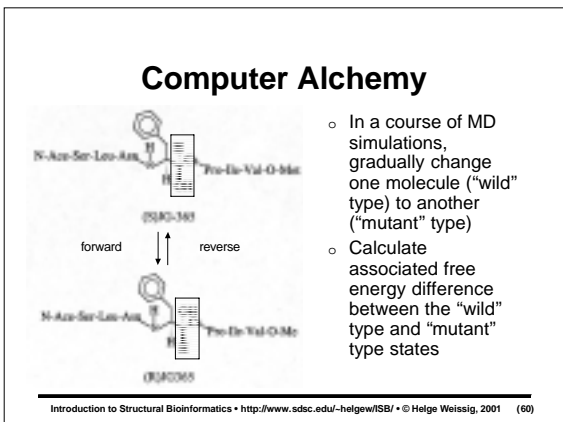
HIV Protease Inhibitors

- o Importance:
 - o Biological and pharmaceutical significance
 - o Abundance of structural information for both free enzyme and enzyme/inhibitor complexes
- o Task: calculation of relative binding constants

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (57)







Computer Alchemy [2]

- o To calculate free energy, the potential energy of the system is made dependent on an artificial parameter λ
- o λ is the relative contribution of one molecule and goes from 1 to 0 during the MD simulation:

$$E = E_{common} + \lambda E_{molecule\ 1} + (1 - \lambda) E_{molecule\ 2}$$

- o Thus:

$$\Delta G_{mol1,mol2}^{\circ} = \int_0^1 \frac{\partial E}{\partial \lambda} d\lambda \approx \sum_i \frac{\partial E}{\partial \lambda_i} \Delta \lambda$$

- o In practice $\Delta \lambda = 1/\Theta$ where Θ is the total simulation time

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (61)

MD Results

Theory

- o Tropsha & Hermans (1992): -2.9 Kcal/mol
- o Ferguson et al. (1991): -2.8 Kcal/mol

Experiment

- o Rich et al. (1991): -2.6 Kcal/mol

Introduction to Structural Bioinformatics • <http://www.sdsc.edu/~helgew/ISB/> • © Helge Weissig, 2001 (62)
