Molecular simulation and structure prediction using CHARMM, Amber and the MMTSB Tool Set

Introduction

Charles L. Brooks III
MMTSB/CTBP
2009 Summer Workshop
CTBP and MMTSB: Who are we?

• The CTBP is the Center for Theoretical Biological Physics
  – Funded by the NSF as a Physics Frontiers Center
  – Partnership between UCSD, University of Michigan and Salk, lead by UCSD
    • The CTBP encompasses a broad spectrum of research and training activities at the forefront of the biology-physics interface.
  – Principal scientists include
    • José Onuchic, Herbie Levine, Henry Abarbanel, Charles Brooks, Mike Holst, Terry Hwa, David Kleinfeld, Andy McCammon, Wouter Rappel, Terry Sejnowski, Wei Wang, Peter Wolynes
Welcome to the Center for Theoretical Biological Physics (CTBP), an NSF-sponsored Center, founded on the campus of the University of California, San Diego as a partnership between UCSD, The Scripps Research Institute, The Salk Institute for Biological Studies and the San Diego Supercomputer Center. The CTBP encompasses a broad spectrum of research and training activities at the forefront of the biology-physics interface. Positions at both the graduate and postgraduate level are available; in addition, the center has an active visitor program for senior scientists.

Molecular Simulation and Structure Prediction using CHARMM and the MMTSB Tool Set

CTBP Summer School Conference
July 31 - August 4, 2006, UCSD

http://ctbp.ucsd.edu
CTBP and MMTSB: Who are we?

- The MMTSB is the Center for Multi-scale Modeling Tools for Structural Biology
  - Funded by the NIH as a National Research Resource Center
  - Partnership between University of Michigan, Rutgers University, Scripps and Georgia Tech, lead by University of Michigan
    - The MMTSB aims to develop new tools and theoretical models to aid molecular and structural biologists in interpreting their biological data.
  - Principal scientists include
    - Charles Brooks, David Case, Jack Johnson, Vijay Reddy, Jeff Skolnick
CTBP and MMTSB: Who are we?

NIH Research Resource Center for the Development of Multiscale Modeling Tools for Structural Biology

News

- Upcoming workshop to feature newly released NMFF software for cryoEM structure refinement.
- MMTSB collaborative project featured as Journal of Molecular Biology cover.
- VIPERdb featured in recent issue of Science.

Research Areas

- Protein Modeling
- Nucleic Acid Modeling
- Virus Structures
- EM Data Fitting
- Collaborations

People

- Investigators
- Collaborators
- Advisory Committee

Workshops

- Future Workshops
- Past Workshops

Software

- General Modeling:
  - MMTSB Tool Set
  - CHARMM / Amber
- Multiscale NA Modeling:
  - Yamnp Tools
  - YUP

Web Services

- Virus Structures:
  - VIRus Particle ExploreR
  - New VIPERdb
- Protein Modeling:
  - CASP4 structures
  - Structure evaluation
  - Structure refinement
  - Loop prediction
  - Ab initio prediction
  - Utility Functions
  - Go Model Builder
  - CHARMM SBMD
  - GB/PB Comparison
- EM-Maps:
  - emotion

http://www.mmtsb.org
CTBP and MMTSB: Who are we?

• Activities
  – Fundamental research across a broad spectrum
  – Software and methods development and distribution
    • MMTSB distributes multiple software packages as well as hosts a variety of web services and databases
  – Training and research workshops and educational outreach
    • Both centers have extensive workshop programs
  – Visitors
    • Both centers host visitors and collaborators for short and longer term (sabbatical) visits
Overview of MMTSB activities

• Research
  – Virus assembly, maturation and structural analysis
  – Structure prediction and protein folding*
  – Homology modeling*
  – Protein, RNA and DNA modeling
  – Large-scale motions in biology
    • Functional displacements in the ribosome
    • Molecular motions from cryo-EM maps
    • Fitting atomic structures into EM densities
Center for the Development of Multi-scale Modeling Tools in Structural Biology (MMTSB)

• Tools and resources
  – Virus Particle Explorer (ViPER) web-base of virus structures and assemblies
    • http://viperdb.scripps.edu
  – MMTSB computational structural biology toolset*
  – CHARMM, Amber, Situs, nab and YAMMP resource pages
  – NMFF - software package for flexibly fitting atomic structures into electron density maps from cryo-EM and tomography

http://www.mmtsb.org
Virus Particle Explorer (VIPERdb)

As featured in Science magazine's NetWatch! Click here to read more.

Data & Analysis  Utilities & Tools
Contact Us  Site Map
Search the Site  Need Help?

2466 visits since 2/11/05
235 virus entries in VIPERdb

http://viperdb.scripps.edu
Your instructors and mentors: Who are we?

- Charles Brooks
- David Case
- Michael Feig
- Jianhan Chen
- Jana K. Shen
- Ross Walker
  - All biophysicists involved in MMTSB and CHARMM/AMBER or MMTSB development
  - Jennifer Knight, Sishi Tang
    - Biophysicists working in Brooks/Case groups as postdoctoral collaborators
What is CHARMM?

• CHARMM is a software package for molecular simulation and analysis of proteins, nucleic acids, lipids, carbohydrates
  – Originated in the group of Martin Karplus at Harvard University circa 1975
  – Currently distributed in more than 1000 laboratories
  – Under continual development by more than 50 developers worldwide
  – CHARMM website and forum provide a venue to explore documentation, discuss results and get advice from advanced CHARMM users and developers

http://www.charmm.org
What is Amber?

- Amber is a suite of programs for molecular simulation and analysis of proteins, nucleic acids, lipids, carbohydrates
  - Originated in the group of Peter Kollman at UCSF circa 1980
  - Currently distributed in 100’s of laboratories
  - Under continual development by more than 10 developers worldwide
  - Amber website and reflector provide a venue to explore documentation, discuss results and get advice from advanced Amber users and developers

http://www.amber.org
What is the MMTSB Tool Set?

• The MMTSB Tool Set is a collection of Perl-based scripts and modules that provide natural user interfaces to CHARMM, Amber, TASSSER, MODELLER, NAMD and other molecular modeling packages
  – Currently downloaded more than 8000 times
  – Under development by in a number of laboratories
  – User forum as part of CHARMM forums

http://www.mmtsb.org
Molecular Mechanics and Modeling
Molecular Mechanics and Modeling - Why

**Experimental Techniques**
- X-ray, NMR (dynamics and structure)
- Light/X-ray/neutron scattering (dynamics and structure)
- Imaging/Cryo-EM (dynamics and structure)
- Calorimetry, pKas, thermodynamics, physical measurements

**Theoretical Methods**
- Development of mathematical models $U(r_1, r_2, ..., r_n)$
- Development of methods to explore models
- Molecular Mechanics
- Exploration of model phenomenology and properties
- Understanding biomolecular structure, dynamics and function

Development of new theories and models to rationalize and predict experimental observations
Overview and Objectives

• What is the basis of molecular mechanics?
  – Mathematical foundations: potential energy functions, energy minimization, molecular dynamics, implicit solvent, boundary conditions

• What are some uses of molecular simulations & modeling?
  – Conformational searching with MD and minimization
  – Exploration of biopolymer fluctuations and dynamics
  – MD as an ensemble sampler

• Free energy simulations
  – Energy minimization as an estimator of binding free energies
  – Application of FEP to protein stability
  – Approximate association free energy of molecular assemblies
  – Approximate pK_a calculations using continuum models
Basic elements of molecular modeling and molecular models
Mathematical Models - Force Fields

- MM force field is a compromise between speed and accuracy.
- Force field is the mathematical basis for expressing structure-energy relationships in biopolymers.
- Common form (CHARMM, Amber, etc.):

  \[
  U(\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N) = \sum_{\text{bonds}, i} \frac{1}{2} k_i^b \cdot (r_i - r_{i0})^2 + \sum_{\text{angles}, i} \frac{1}{2} k_i^\theta \cdot (\theta_i - \theta_{i0})^2 \\
  + \sum_{\text{torsions}, i} k_i^\phi \cdot [1 + \cos(n_i \phi_i - \delta_i)] \\
  + \frac{1}{2} \sum_{\text{nonbond pairs, (i, j)}} \left\{ \epsilon_{\text{min}}^{ij} \left[ \left( \frac{r_{ij}}{r_{ij0}} \right)^{12} - 2 \left( \frac{r_{ij}}{r_{ij0}} \right)^6 \right] + \frac{q_i q_j}{\varepsilon r_{ij}} \right\}
  \]

- Energy terms - bonds

From spectroscopy, IR, etc.
Mathematical Models - Force Fields

• Angles

\[ u_{\text{angle}} = \frac{1}{2} k_i \theta \cdot (\theta_i - \theta_i^0)^2 \]

From spectroscopy, IR, etc.

• Dihedrals

\[ u_{\text{dihedral}} = k_i^\phi \cdot [1 + \cos(3\phi_i)] + k_i^\phi \cdot [1 - \cos(\phi_i - \pi)] \]

From spectroscopy, IR, NMR, empirical, QM
Mathematical Models - Force Fields

- Nonbonded - Lennard-Jones
  \[ u_{L-J} = \varepsilon_{\text{min}}^{ij} \left[ \left( \frac{r_{\text{min}}^{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{\text{min}}^{ij}}{r_{ij}} \right)^{6} \right] \]

- Nonbonded - electrostatics
  \[ u_{\text{Coulomb}} = \frac{q_i q_j}{\varepsilon r_{ij}} \]

Non-bonded interactions derived from quantum chemistry, thermodynamics, empirical schemes
Assessing and deriving energy functions

- Quantum chemistry provides means of deriving non-bonded energy functions
  - $\phi/\psi$ map for alanine dipeptide from QC calculations

| Table 8. Interaction Energies (kcal/mol) of Two $\beta$-Sheet Conformation Alanine Dipeptides$^a$ |
|---------------------------------|---|
| MM3($\varepsilon = 1.5$)        | - 7.23 |
| HF/ccc-pVTZ(- f) (CP corrected) | - 8.15 |
| MMFF($\varepsilon = 2.0$)       | - 8.04 |
| HF/6-31G** (CP corrected)       | - 9.35 |
| HF/ccc-pVTZ(- f) (non-CP corrected) | - 9.42 |
| MM3($\varepsilon = 1.5$)        | - 9.69 |
| AMBER 3                         | - 9.78 |
| MM3($\varepsilon = 1.0$)        | - 10.23 |
| LMP2/ccc-pVTZ(- f) (HF CP corrected) | - 10.73 |
| CVFF                            | - 10.77 |
| MMFF($\varepsilon = 1.5$)       | - 11.01 |
| AMBER$^a$                       | - 11.07 |
| HF/6-31G** (non-CP corrected)   | - 11.68 |
| LMP2/ccc-pVTZ(- f) (HF non-CP corrected) | - 12.00 |
| CCFF                            | - 12.14 |
| AMBER$^b$($\varepsilon = 1.0$)  | - 12.98 |
| MSICCHARm                        | - 12.99 |
| MM2$^a$                         | - 13.02 |
| OPLSA-AA(2.2)                   | - 13.21 |
| MM2$^b$($\varepsilon = 1.0$)    | - 13.47 |
| CHARMM 22                       | - 14.10 |
| MMFF$^a$                        | - 14.97 |
| CHARMM 19                       | - 15.21 |
| MMFF$^b$                        | - 15.38 |
| AMBER94                         | - 16.01 |
| MM2                             | - 16.11 |
| AMBER94($\varepsilon = 1.0$)    | - 16.50 |
| OPLSA-UA(2.8)                   | - 16.70 |
| OPLSA-UA(2.2)                   | - 16.91 |
| OPLS$^a$                        | - 17.63 |

$^a$ Unless otherwise specified, $\varepsilon = 1.0$.

- Quantum chemistry provides “tests” of force fields

M.D. Beachey et al., *JACS*, **119**, 5908 (‘97)
M. Feig et al., *JPCB*, **107**, 2831 (‘03)
A. MacKerell et al., *JCC*, **25**, 1400 (‘04)
Adding Charge Polarization via Charge Equalization

*Electrostatic Potential Energy*

\[
E_{\text{electrostatic}} = \sum_{i=1}^{N} \chi_i Q_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \eta_{ij} Q_i Q_j
\]

- **Equilibrium charge density**: \(Q_i\) and \(Q_j\)
- **Penalty for perturbation from equilibrium due to presence of field**: \(\eta_{ij}\)

**Distance dependent coulomb shielding**

- 1-2 (Bond Atoms)
- 1-3 (Angle Atoms)
- 1-4 (Dihedral Atoms)

**Parameterization of \(\eta\) and \(\chi^0\)**

\[
\eta \Delta \bar{Q} = -\bar{\phi} \quad \Rightarrow \quad \Delta \bar{Q} = -\eta^{-1} \bar{\phi}
\]

\(\Delta Q\) is the difference of the partial charge of an atom due to an applied external potential, \(\phi_k\), relative to vacuum. Decouples fitting of \(\eta\) and \(\chi\)

**Objective Function**

\[
\varepsilon = \left\| \left( \Delta Q_{\text{DFT}} - \Delta Q_{\text{FQ}}^\text{DFT} \right) \right\|
\]

Hardness parameters scaled to reduce condensed phase polarizability—represents confinement of diffuse tails of molecular electronic density due to Pauli repulsion in dense liquid (perhaps a universal need to employ reduced polarizabilities in classical simulations incorporating polarization).

Patel & Brooks, JCC, 24, 1, 2004
Electronic Polarization: Fluctuating Charge Dynamics

- Extended Lagrangian Formulation

\[ L = \sum_{i=1}^{M} \sum_{\alpha=1}^{N_i} \frac{1}{2} m_{i\alpha} \dot{r}_{i\alpha}^2 + \sum_{i=1}^{M} \sum_{\alpha=1}^{N_i} \frac{1}{2} m_{Q,i\alpha} \dot{Q}_{i\alpha}^2 - E(Q,r) - \sum_{i=1}^{M} \lambda_i \sum_{\alpha=1}^{N_i} Q_{i\alpha} \]

- Charge Equations of Motion

\[ m_{Q,i\alpha} \ddot{Q}_{i\alpha} = -\frac{\partial E(Q,r)}{\partial Q_{i\alpha}} - \lambda_i \]

Total charge on molecule ‘i’ constant

\[ \sum_{i=1}^{M} \sum_{\alpha=1}^{N_i} \ddot{Q}_{i\alpha} = 0 \]

\[ \lambda_i = -\frac{\sum_{\alpha=1}^{N_i} Q_{i\alpha}}{N_i} = -\frac{\sum_{\alpha=1}^{N_i} \tilde{\chi}_{i\alpha}}{N_i} \]

average electronegativity of molecule ‘i’

- Charge Evolution/Dynamics

\[ m_{Q,i\alpha} \ddot{Q}_{i\alpha} = -\sum_{\beta=1}^{N_i} (\tilde{\chi}_{i\alpha} - \tilde{\chi}_{i\beta}) \]

force on charge ‘a’ is proportional to the difference between the instantaneous site and average molecular electronegativities
Parameterization of Non-Electrostatic Parameters (Protein Force Field)

\[ U = U_{bonds} + U_{angles} + U_{dihedrals} + U_{cross} + \sum_{\text{non-bonded atom pairs}} \left( U_{\text{dispersion}} + \frac{q_i q_j}{\varepsilon r_{ij}} \right) \]

Electrostatic Parameters

Vacuum Water-Solute Dimers: Geometries, Energies

- CHARMM atom types
- Bulk Liquid Properties: (vaporization enthalpy, density)
  - solute-solvent and solute-solute energetics
  - solvent=TIP4P-FQ (protein ff intimately coupled to solvent model)

van der Waals (dispersion)

Backbone Torsion, Angle, Bond, intra-molecular dispersion interactions

Protein Molecular Dynamics (drift from native structure)

Patel et al., JCC, 24, 1504, 2004

Scaling of polarizability

Electronic Polarization - deriving parameters

Charge Equilibration/Electronegativity Equalization

N-Methylacetamide (vacuum)

\[ Q^0 \]

\[ Q = Q^0 + \Delta Q \]

- Equilibrium distribution of charges (i.e. vacuum)
- Redistribution of charge gives rise to electronic polarization (charge flow maintains electronegativity equalization)
- Directionality of charge flow - atomic electronegativity
- \( \Delta Q \) governed by measure of resistance to charge flow to/from a given site-atomic hardness
- Polarizability \( \rightarrow \) \( \Delta Q \)
Molecular Mechanics
The Basic Algorithms
Demystifying Molecular Mechanics - Energy Minimization

- Minimization follows gradient of potential to identify stable points on energy surface
  - Let $U(x) = \frac{a}{2}(x-x_0)^2$
  - Begin at $x'$, how do we find $x_0$ if we don’t know $U(x)$ in detail?
    - How can we move from $x'$ to $x_0$?
  - Steepest descent based algorithms (SD):
    - $x \rightarrow x' = x + \delta$
    - $\delta = -\kappa \frac{\partial U(x)}{\partial x} = -\kappa a(x - x_0)$
    - This moves us, depending on $\kappa$, toward the minimum.
    - On a simple harmonic surface, we will reach the minimum, $x_0$, i.e. converge, in a certain number of steps related to $\kappa$.

- SD methods use first derivatives only
- SD methods are useful for large systems with large forces
Demystifying Molecular Mechanics - Energy Minimization

• Related Conjugate gradient methods
  – For this algorithm:
    • \( x_n \rightarrow x_{n+1} = x_n + \alpha \delta_n; \delta_n = -\nabla_n U(x) + \delta_{n-1} A \)
  • \( A = \frac{|\nabla_n U|^2}{|\nabla_{n-1} U|^2} \)

  – A related method is the Fletcher-Powell minimizer
• CG and Powell methods use first derivatives only
• Newton-Raphson (NR) and adopted basis NR (ABNR) use 2\textsuperscript{nd} derivatives

NR algorithm: \( x_{n+1} = x_n + \delta_n \)
\( \delta_n = \frac{-(\nabla_n U)}{(\nabla_n \nabla_n U)} \)

  – For our 1-D example:
  – \( \delta_n = -a(x_n - x_0)/a = x_0 - x_n \)
• ABNR approximates 2nd derivatives
• Best near minimum
Demystifying Molecular Mechanics - Molecular Dynamics

- Molecular dynamics
- Objective: \((r_1(t), ..., r_N(t)) \rightarrow (r_1(t+\Delta t), ..., r_N(t+\Delta t))\)

  - expand \(x(t \pm \Delta t)\) in Taylor's series around \(t\)
    \[x(t \pm \Delta t) = x(t) \pm v(t)\Delta t + \frac{1}{2m} f(t)\Delta t^2 \pm \frac{1}{6} \dddot{x}(t)\Delta t^3 + O(\Delta t^4)\]

  - add expansion \(x(t + \Delta t)\) and \(x(t - \Delta t)\) and rearrange
    \[x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{f(t)}{m} \Delta t^2 + O(\Delta t^4)\] (position propagation)

  - add expansion \(x(t - \Delta t)\) and \(x(t + \Delta t)\) and rearrange
    \[v(t) = (x(t + \Delta t) - x(t - \Delta t)/(2\Delta t) + O(\Delta t^3)\] (velocity propagation)
Demystifying Molecular Mechanics - Molecular Dynamics

• Other “summed forms”
  – Leap-frog
    \[ x(t + \Delta t) = x(t) + \Delta t \cdot v(t + \frac{1}{2}\Delta t) \] (position propagation)
    \[ v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + \Delta t \cdot \frac{f(t)}{m} \] (velocity propagation)
  – Velocity Verlet
    \[ x(t + \Delta t) = x(t) + \Delta t \cdot v(t + \Delta t) - \frac{1}{2m} f(t) \Delta t^2 \] (position propagation)
    \[ v(t + \Delta t) = v(t) + \Delta t \cdot \frac{(f(t) + f(t + \Delta t))}{2m} \] (velocity propagation)

• Time step controls accuracy of numerical solution
  – \( \Delta t = 10^{-15} \) sec = 1 fs
  – Fundamental time step determined by high frequency vibrations (bonds)

• Highest frequency motions removed with holonomic constraints (SHAKE)
  – w/SHAKE can increase time step by ~2
Boundary Conditions and Statistical Ensembles
Periodic Boundary Conditions and Solvent Effects
PBCs a Panacea or Not?

Hunenberger & McCammon…
• continuum calculations show artifacts
• reproduced in molecular dynamics simulations

artificial stabilization of α-helix

poly-alanine octapeptide, 2 ns simulations
Controlling Thermodynamic Variables

T and P

• Statistical ensembles connect microscopic to macroscopic/thermodynamic

• NVE (microcanonical - Entropy rules!)

• NVT (Canonical - Helmholtz free energy is relevant, A)
  – $T = \sum m \langle v^2 \rangle / (3k_B)$

• NPT (Isothermal-isobaric - Gibbs free energy is relevant, G)
  – $P =$ kinetic + virial contributions

• Thermostats, barostats, etc., allow one to choose appropriate ensembles
  – Following Nose’, Hoover, Evans and others...[see C.L. Brooks, III, Curr. Opin. Struct. Biol., 5, 211(’95)]
Barriers, Temperature and Size Yield Timescales

• How long should simulations be?

\[ \tau = \tau_0 \exp(\Delta G^\ddagger / k_B T) \]

- \( \tau_0 \sim 10^{-12}, \Delta G^\ddagger \)
  - 1 kcal/mol: \( \sim 1.2 \text{ ps}^{-1} \)
  - 5 kcal/mol: \( \sim 1.5 \text{ ns}^{-1} \)
  - 10 kcal/mol: ms or longer!

• Sampling should exceed timescales of interest by \( \sim 10 \)-fold

• Size and complexity also increase required timescales
  - Equilibration of ions, complex landscapes, multiple minima
Simplifications - eliminating explicit solvent and solvent boundary methods

- Free energy changes are partitioned into internal and external components
- \( \Delta G_{\text{total}} = \Delta E_{\text{internal}} + \Delta S_{\text{conformation}} + \Delta G_{\text{solvation}} \)
- \( \Delta G_{\text{solvation}} = \Delta G_{\text{electrostatics}} + \Delta G_{H-\phi} \)
- \( \Delta G_{H-\phi} = \sum \gamma_i S A_i \)
- \( \Delta G_{\text{electrostatics}} \sim \text{continuum electrostatics} \)