Molecular Dynamics Simulations of Drug-Target Complexes
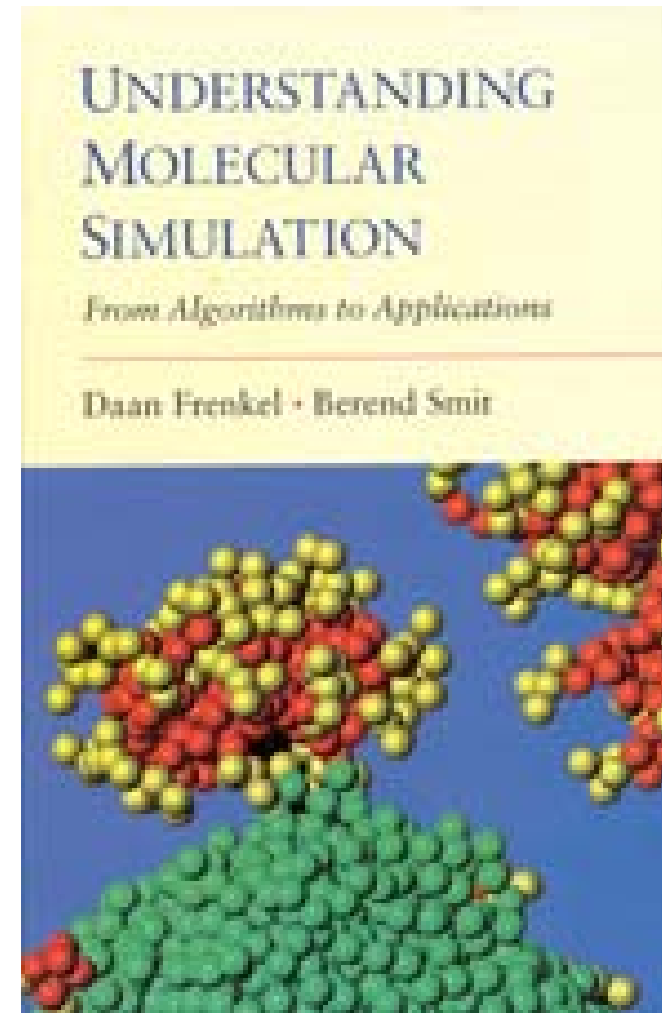
Jung-Hsin Lin (林榮信)

Research Center for Applied Sciences &
Institute of Biomedical Sciences, Academia Sinica
School of Pharmacy, National Taiwan University
College of Engineering, Chang Gung University

http://www.rcas.sinica.edu.tw/faculty/jhlin.html

2018 Frontiers in Computational Drug Design, Academia Sinica, March 16-20, 2018

Daan Frankel

You don't really understand it, if you cannot simulate it.

Boltzmann award, StatPhys at Lyon, 2016

Watching a Protein as it Functions with 150-ps Time-Resolved X-ray Crystallography

Friedrich Schotte,[1] Manho Lim,[2] Timothy A. Jackson,[3] Aleksandr V. Smirnov,[1] Jayashree Soman,[4] John S. Olson,[4] George N. Phillips Jr.,[5] Michael Wulff,[6] Philip A. Anfinrud[1]

We report picosecond time-resolved x-ray diffraction from the myoglobin (Mb) mutant in which Leu[29] is replaced by Phe (L29F mutant). The frame-by-frame structural evolution, resolved to 1.8 angstroms, allows one to literally "watch" the protein as it executes its function. Time-resolved mid-infrared spectroscopy of flash-photolyzed L29F MbCO revealed a short-lived CO intermediate whose 140-ps lifetime is shorter than that found in wild-type protein by a factor of 1000. The electron density maps of the protein unveil transient conformational changes far more dramatic than the structural differences between the carboxy and deoxy states and depict the correlated side-chain motion responsible for rapidly sweeping CO away from its primary docking site.

tions of photolyzed wild-type MbCO showed that CO becomes temporarily trapped in a nearby ligand-docking site before escaping into the surrounding solvent (6). This site presumably mediates the transport of ligands to and from the active binding site. Molecular dynamics simulations suggested a location for this docking site (7), which was found to be consistent with low-temperature crystal structures of photolyzed MbCO (8, 9) where "docked" CO was displaced about 2 Å from the active binding site. The docking site is fashioned by the heme and amino acid side chains valine (Val[68]), isoleucine (Ile[107]), and leucine (Leu[29]), all of which are highly conserved in mammalian Mb. Therefore, site-directed mutagenesis at these sites would likely influence the function of the protein. Indeed, in the L29F mutant (where Leu[29] is replaced by phenylalanine), the O_2 affinity is elevated by an order of magnitude, and the autooxidation rate is lowered by a compara-

Picosecond time-resolved X-ray crystallography: probing protein function in real time

Friedrich Schotte[a], Jayashree Soman[b], John S. Olson[b], Michael Wulff[c], Philip A. Anfinrud[a],*

High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography

Sébastien Boutet,[1]* Lukas Lomb,[2,3] Garth J. Williams,[1] Thomas R. M. Barends,[2,3] Andrew Aquila,[4] R. Bruce Doak,[5] Uwe Weierstall,[5] Daniel P. DePonte,[4] Jan Steinbrener,[2,3] Robert L. Shoeman,[2,3] Marc Messerschmidt,[1] Anton Barty,[4] Thomas A. White,[4] Stephan Kassemeyer,[2,3] Richard A. Kirian,[5] M. Marvin Seibert,[1] Paul A. Montanez,[1] Chris Kenney,[6] Ryan Herbst,[6] Philip Hart,[6] Jack Pines,[6] Gunther Haller,[6] Sol M. Gruner,[7,8] Hugh T. Philipp,[7] Mark W. Tate,[7] Marianne Hromalik,[9] Lucas J. Koerner,[10] Niels van Bakel,[11] John Morse,[12] Wilfred Ghonsalves,[1] David Arnlund,[13] Michael J. Bogan,[14] Carl Caleman,[4] Raimund Fromme,[15] Christina Y. Hampton,[14] Mark S. Hunter,[15] Linda C. Johansson,[13] Gergely Katona,[13] Christopher Kupitz,[15] Mengning Liang,[4] Andrew V. Martin,[4] Karol Nass,[16] Lars Redecke,[17,18] Francesco Stellato,[4] Nicusor Timneanu,[19] Dingjie Wang,[5] Nadia A. Zatsepin,[5] Donald Schafer,[1] James Defever,[1] Richard Neutze,[13] Petra Fromme,[15] John C. H. Spence,[5] Henry N. Chapman,[4,16] Ilme Schlichting[2,3]

Structure determination of proteins and other macromolecules has historically required the growth of high-quality crystals sufficiently large to diffract x-rays efficiently while withstanding radiation damage. We applied serial femtosecond crystallography (SFX) using an x-ray free-electron laser (XFEL) to obtain high-resolution structural information from microcrystals (less than 1 micrometer by 1 micrometer by 3 micrometers) of the well-characterized model protein lysozyme. The agreement with synchrotron data demonstrates the immediate relevance of SFX for analyzing the structure of the large group of difficult-to-crystallize molecules.

[1]Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [2]Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany. [3]Max Planck Advanced Study Group, Center for Free-Electron Laser Science, Notkestrasse 85, 22607 Hamburg, Germany. [4]Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany. [5]Department of Physics, Arizona State University, Tempe, AZ 85287, USA. [6]Particle Physics and Astrophysics, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [7]Department of Physics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA. [8]Wilson Laboratory, Cornell High Energy Synchrotron Source (CHESS), Cornell University, Ithaca, NY 14853, USA. [9]Electrical and Computer Engineering, State University of New York (SUNY) Oswego, Oswego, NY 13126, USA. [10]The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA. [11]Nikhef, National Institute for Subatomic Physics, Science Park 105, 1098 XG Amsterdam, Netherlands. [12]European Synchrotron Radiation Facility, 38043 Grenoble Cedex, France. [13]Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden. [14]PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [15]Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287–1604, USA. [16]University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. [17]Joint Laboratory for Structural Biology of Infection and Inflammation, Institute of Biochemistry and Molecular Biology, University of Hamburg, and Institute of Biochemistry, University of Lübeck, at DESY, Hamburg, Germany. [18]German Centre for Infection Research, University of Lübeck, 23538 Lübeck, Germany. [19]Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden.

*To whom correspondence should be addressed. E-mail: sboutet@slac.stanford.edu

Natively Inhibited *Trypanosoma brucei* Cathepsin B Structure Determined by Using an X-ray Laser

Lars Redecke,[1,2]† Karol Nass,[3,4]† Daniel P. DePonte,[3] Thomas A. White,[3] Dirk Rehders,[1] Anton Barty,[3] Francesco Stellato,[3] Mengning Liang,[3] Thomas R.M. Barends,[5,6] Sébastien Boutet,[7] Garth J. Williams,[7] Marc Messerschmidt,[7] M. Marvin Seibert,[7] Andrew Aquila,[3] David Arnlund,[8] Sasa Bajt,[9] Torsten Barth,[10] Michael J. Bogan,[11] Carl Caleman,[3] Tzu-Chiao Chao,[12] R. Bruce Doak,[13] Holger Fleckenstein,[3] Matthias Frank,[14] Raimund Fromme,[12] Lorenzo Galli,[3,4] Ingo Grotjohann,[12] Mark S. Hunter,[12]* Linda C. Johansson,[8] Stephan Kassemeyer,[5,6] Gergely Katona,[8] Richard A. Kirian,[3,13] Rudolf Koopmann,[10] Chris Kupitz,[12] Lukas Lomb,[5,6] Andrew V. Martin,[3] Stefan Mogk,[10] Richard Neutze,[8] Robert L. Shoeman,[5,6] Jan Steinbrener,[5,6] Nicusor Timneanu,[15] Dingjie Wang,[13] Uwe Weierstall,[13] Nadia A. Zatsepin,[13] John C. H. Spence,[13] Petra Fromme,[12] Ilme Schlichting,[5,6] Michael Duszenko,[10] Christian Betzel,[16]‡ Henry N. Chapman[3,4]‡

The *Trypanosoma brucei* cysteine protease cathepsin B (TbCatB), which is involved in host protein degradation, is a promising target to develop new treatments against sleeping sickness, a fatal disease caused by this protozoan parasite. The structure of the mature, active form of TbCatB has so far not provided sufficient information for the design of a safe and specific drug against *T. brucei*. By combining two recent innovations, in vivo crystallization and serial femtosecond crystallography, we obtained the room-temperature 2.1 angstrom resolution structure of the fully glycosylated precursor complex of TbCatB. The structure reveals the mechanism of native TbCatB inhibition and demonstrates that new biomolecular information can be obtained by the "diffraction-before-destruction" approach of x-ray free-electron lasers from hundreds of thousands of individual microcrystals.

[1]Joint Laboratory for Structural Biology of Infection and Inflammation, Institute of Biochemistry and Molecular Biology, University of Hamburg, and Institute of Biochemistry, University of Lübeck, at Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany. [2]German Centre for Infection Research, University of Lübeck, 23538 Lübeck, Germany. [3]Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany. [4]Department of Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. [5]Max-Planck-Institut für medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany. [6]Max Planck Advanced Study Group, Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany. [7]Linac Coherent Light Source, Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [8]Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden. [9]Photon Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. [10]Interfaculty Institute of Biochemistry, University of Tübingen, Hoppe-Seyler-Strasse 4, 72076 Tübingen, Germany. [11]Photon Ultrafast Laser Science and Engineering (PULSE) Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [12]Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA. [13]Department of Physics, Arizona State University, Tempe, AZ 85287, USA. [14]Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA. [15]Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, SE-75124 Uppsala, Sweden. [16]Institute of Biochemistry and Molecular Biology, University of Hamburg, at DESY, Notkestrasse 85, 22607 Hamburg, Germany.

*Present address: Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA.
†These authors contributed equally to this study.
‡To whom correspondence should be addressed. E-mail: henry.chapman@desy.de (H.N.C.) or christian.betzel@uni-hamburg.de (C.B.)

Deriving the macroscopic world from atoms

$$\frac{n_i}{n_j} = \frac{P(C_i)}{P(C_j)} = \frac{e^{-\varepsilon_i/k_B T}}{e^{-\varepsilon_j/k_B T}}$$

$$Q = \sum_{\text{all configuration } c} e^{-\varepsilon_c/k_B T}$$

$$= e^{-(\varepsilon_i - \varepsilon_j)/k_B T}$$

$$U = \frac{k_B T^2}{Q}\left(\frac{\partial Q}{\partial T}\right)_V = k_B T^2 \left(\frac{\partial \ln Q}{\partial T}\right)$$

Ludwig Eduard Boltzmann (February 20, 1844 – September 5, 1906) was an Austrian physicist whose greatest achievement was in the development of statistical mechanics, which explains and predicts how the properties of atoms (such as mass, charge, and structure) determine the visible properties of matter (such as viscosity, thermal conductivity, and diffusion).

$$S - S_0 = k_B T\left(\frac{\partial \ln Q}{\partial T}\right)_V + k_B \ln Q - k_B \ln Q_0$$

$$A = U - TS$$
$$= -k_B T \ln Q$$
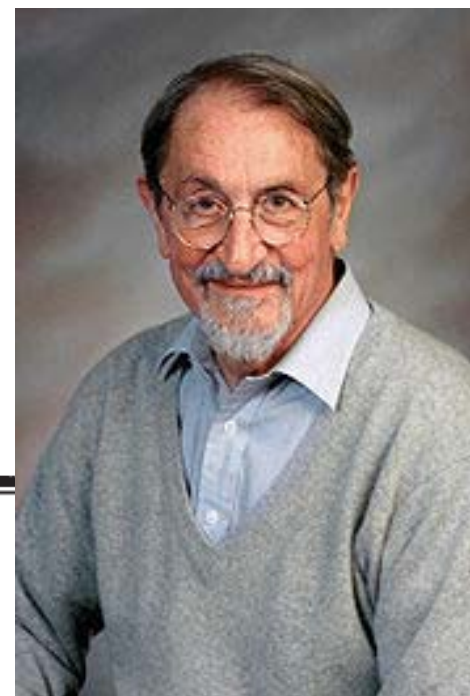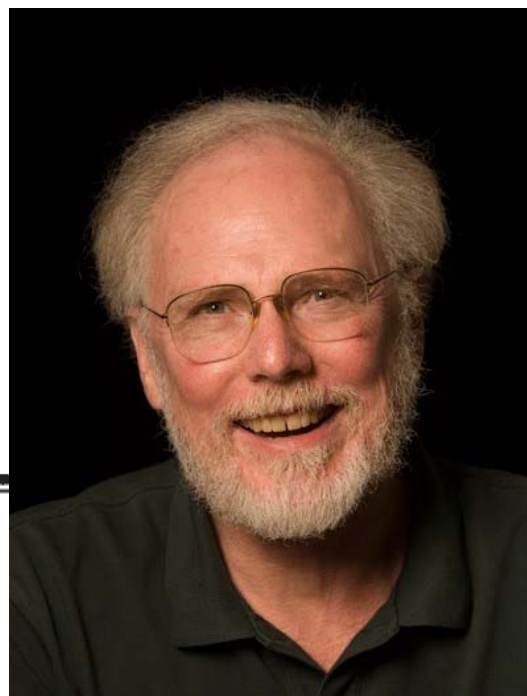
2013 Nobel Prize in Chemistry

articles

Dynamics of folded proteins

J. Andrew McCammon, Bruce R. Gelin & Martin Karplus

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

The dynamics of a folded globular protein (bovine pancreatic trypsin inhibitor) have been studied by solving the equations of motion for the atoms with an empirical potential energy function. The results provide the magnitude, correlations and decay of fluctuations about the average structure. These suggest that the protein interior is fluid-like in that the local atom motions have a diffusional character.

number of interactions which must be calculated and also permits larger steps in the trajectory calculation since the high frequency hydrogen vibrations have been eliminated. Integration of the equations of motion was performed by means of the Gear algorithm[22] with time steps of 9.78×10^{-16} s. X-ray coordinates[21] were used for the initial positions and the initial velocities were set equal to zero. After 100 equilibration steps, the stresses in the initial structure had partly relaxed and the system had an internal kinetic energy corresponding to a temperature of 140 K. At this point, all velocities were multi-

2013 Chemistry Prize

Taking the Experiment to Cyberspace

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*.

▶ Press release
▶ Popular information
▶ Advanced information

Martin Karplus

Martin Karplus, U.S. and Austrian citizen. Born 1930 in Vienna, Austria. Ph.D. 1953 from California Institute of Technology, CA, USA. Professeur Conventionné, Université de Strasbourg, France and Theodore William Richards Professor of Chemistry, Emeritus, Harvard University, Cambridge, MA, USA.

▶ More on Martin Karplus

Michael Levitt

Michael Levitt, U.S., British and Israeli citizen. Born 1947 in Pretoria, South Africa. Ph.D. 1971 from University of Cambridge, UK. Robert W. and Vivian K. Cahill Professor in Cancer Research, Stanford University School of Medicine, Stanford, CA, USA.

▶ Have a look at Michael Levitt's photo gallery

Arieh Warshel

Arieh Warshel, U.S. and Israeli citizen. Born 1940 in Kibbutz Sde-Nahum, Israel. Ph.D. 1969 from Weizmann Institute of Science, Rehovot, Israel. Distinguished Professor, University of Southern California, Los Angeles, CA, USA.

▶ Interviews with Chemistry Laureate Arieh Warshel

# Protein Dynamics: Moore's Law in Molecular Biology

The millisecond barrier has been broken in molecular dynamics simulations of proteins. Such simulations are increasingly revealing the inner workings of biological systems by generating atomic-level descriptions of their behaviour that make testable predictions about key molecular processes.

Michele Vendruscolo
and Christopher M. Dobson

A fundamental understanding of the manner in which a protein molecule functions depends on a detailed knowledge of not just its structure but also its dynamical behaviour [1]. As molecular dynamics simulations carried out on modern computers make it possible to solve the equations that describe the motion of molecular systems [2], they are a supremely powerful way of providing information at atomic resolution about the way in which protein molecules move and interact with their environments [3]. Indeed, after the initial report of the first application of molecular dynamics simulations to studying the structural

(A) Growth in the timescale accessible to molecular dynamics simulations of proteins. After the first (*in vacuo*) simulation of a protein [4] (red point, bottom left), the timescale accessible to all-atom simulations in water has been increasing exponentially, indeed doubling every year, to reach the millisecond range [5] (red point, top right). By 2020, it should therefore be possible to follow the trajectories of small proteins for seconds and beyond. (B) Growth in the size (in number of atoms) of the protein systems studied by all-atom molecular dynamics simulations. The blue line is a theoretical limit derived by assuming that the growth in system size goes as the square root of the increase in timescale, as the number of interactions scales as the square of the number of atoms. The largest post-1990 simulations tend to fall below this limit, partly because of the lack of high-resolution all-atom structures of large macromolecular complexes, and partly because it is often not useful to devote a great deal of resources to simulating a very large system if it can be achieved only for, very short timescale. (C) Change with time in the limits of the timescales and system sizes accessible by molecular dynamics simulations. The blue and red diagonal lines define the boundaries at given times; for example, the current limit (red line labeled as '2010') runs through the millisecond simulation of small proteins [5], and the nanosecond multi-million atoms simulations of the ribosome [8], whose structure is shown in (B). If the current trends continue, it will be possible to simulate the dynamical properties of the ribosome for milliseconds by 2030 (green cross).
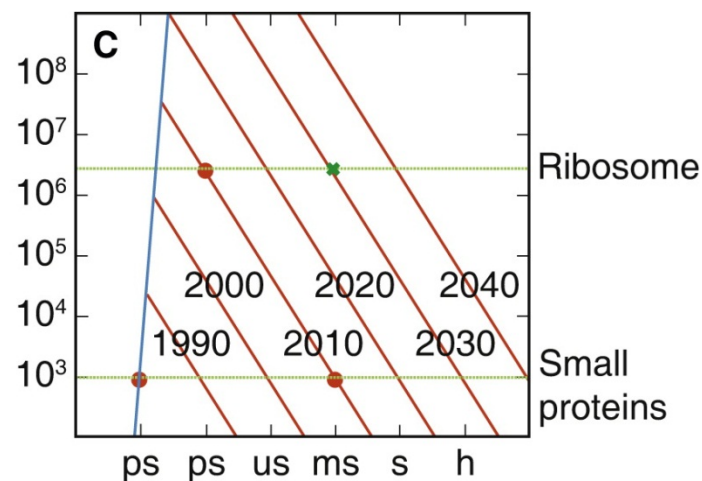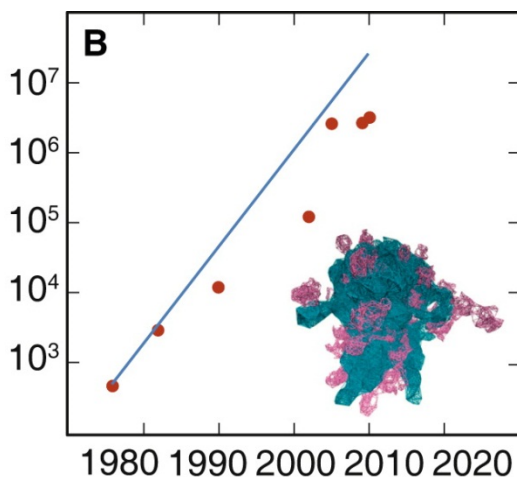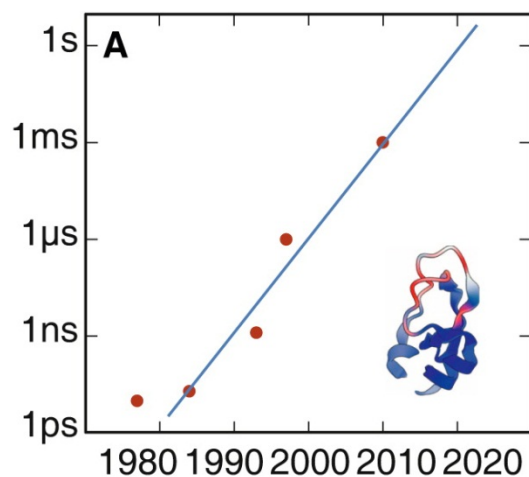
Popular Molecular Dynamics Program Suites

- AMBER
- CHARMm
- GROMACS
- NAMD
- DESMOND

Force Field Molecular Mechanics

What is AMBER?
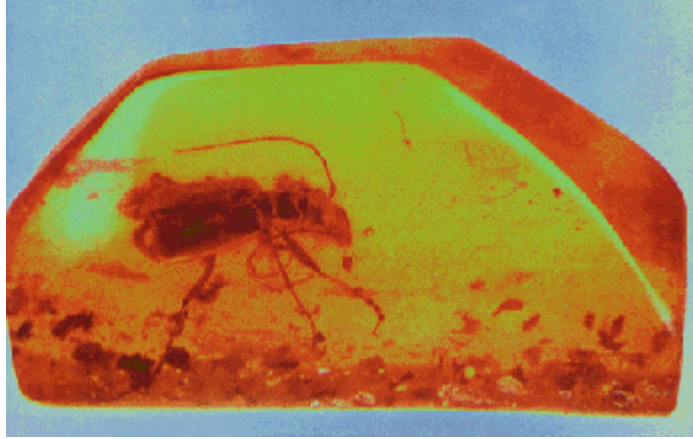
- ***A**ssisted **M**odel **B**uilding with **E**nergy **R**efinement*
- "Amber" refers to two things: a set of molecular mechanical **force fields** for the simulation of biomolecules (which are in the public domain, and are used in a variety of simulation programs); and a package of molecular **simulation programs** which includes source code and demos.
- The current version of the code is Amber version 10, which is distributed by UCSF.

A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins

Scott J. Weiner, Peter A. Kollman,* David A. Case,[†] U. Chandra Singh, Caterina Ghio,[‡] Giuliano Alagona,[‡] Salvatore Profeta, Jr.,[§] and Paul Weiner[⊥]

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143. Received April 28, 1983

Abstract: We present the development of a force field for simulation of nucleic acids and proteins. Our approach began by obtaining equilibrium bond lengths and angles from microwave, neutron diffraction, and prior molecular mechanical calculations, torsional constants from microwave, NMR, and molecular mechanical studies, nonbonded parameters from crystal packing calculations, and atomic charges from the fit of a partial charge model to electrostatic potentials calculated by ab initio quantum mechanical theory. The parameters were then refined with molecular mechanical studies on the structures and energies of model compounds. For nucleic acids, we focused on methyl ethyl ether, tetrahydrofuran, deoxyadenosine, dimethyl phosphate, 9-methylguanine–1-methylcytosine hydrogen-bonded complex, 9-methyladenine–1-methylthymine hydrogen-bonded complex, and 1,3-dimethyluracil base-stacked dimer. Bond, angle, torsional, nonbonded, and hydrogen-bond parameters were varied to optimize the agreement between calculated and experimental values for sugar pucker energies and structures, vibrational frequencies of dimethyl phosphate and tetrahydrofuran, and energies for base pairing and base stacking. For proteins, we focused on Φ,Ψ maps of glycyl and alanyl dipeptides, hydrogen-bonding interactions involving the various protein polar groups, and energy refinement calculations on insulin. Unlike the models for hydrogen bonding involving nitrogen and oxygen electron donors, an adequate description of sulfur hydrogen bonding required explicit inclusion of lone pairs.

The 1984 AMBER Force Field Equation

$$U(\mathbf{r_1}, \mathbf{r_2}, \cdots, \mathbf{r_N}) = \sum_{all\ bonds} k_b (l - l_0)^2$$

$$+ \sum_{all\ angles} k_\theta (\theta - \theta_0)^2$$

$$+ \sum_{all\ dihedrals} \frac{V_n}{2} \left[1 + \cos(n\phi + \gamma) \right]$$

$$+ \begin{cases} \sum_{i<j} \left[\dfrac{A_{ij}}{r_{ij}^{12}} - \dfrac{B_{ij}}{r_{ij}^{6}} \right] \\ \text{or} \\ \sum_{i<j} \left[\dfrac{C_{ij}}{r_{ij}^{12}} - \dfrac{D_{ij}}{r_{ij}^{10}} \right] \end{cases}$$
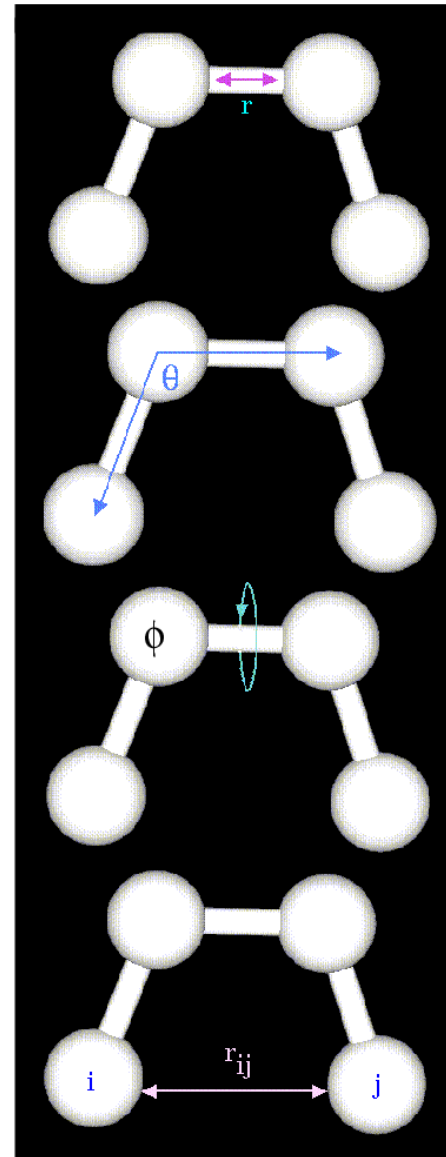
$$+ \sum_{i<j} \left[\frac{q_i q_j}{r_{ij}} \right]$$

$$U_{bond} = \sum_{bonds} K_r (r - r_{eq})^2$$

$$U_{angles} = \sum_{angles} K_\theta (\theta - \theta_{eq})^2$$

$$U_{dihedral} = \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]$$

$$U_{nonbond} = \sum_{i<j} \left\{ \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] \right.$$
$$\text{or} \left. \underbrace{\left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]}_{\text{hydrogen-bonding pair}} \right\}$$
$$+ \sum_{i<j} \frac{q_i q_j}{\epsilon r_{ij}}$$

The van der Waals interactions

Lennard-Jones potential:

$$E_p(r) = \left(\frac{A}{r^{12}} - \frac{B}{r^6} \right)$$

$$= 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

Can the Lennard–Jones 6-12 Function Replace the 10-12 Form in Molecular Mechanics Calculations?

David M. Ferguson and Peter A. Kollman

Department of Pharmaceutical Chemistry University of California, San Francisco, California 94143

A protocol to replace "10-12" hydrogen bonding function with the "6-12" form to reproduce hydrogen bond distances, energies, and geometries in molecular mechanics calculations is described. The 6-12 function was least-squares fit to the normally employed 10-12 form of the function for the hydrogen bond types of the Weiner et al. force field by iterating over the A and B coefficients. A weighting function was used to fit the curves in the most critical areas. The 6-12 hydrogen bond model was compared with the Weiner et al. force field, OPLS/AMBER fore field, and quantum mechanical calculations on two simple systems, the water dimer and the chloride-water interaction. The 6-12 model produced structures, energies, and geometries that were consistent with the other molecular mechanics calculations and showed reasonable agreement to the quantum mechanical results for the water dimer. The 6-12 model was also compared with normal calculations using a 10-12 model on several representative systems. The results indicate that the 6-12 function, when substituted by the procedure outlined in this work, yields structures and hydrogen bond properties that are similar to the normal 10-12 model.

A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules

Wendy D. Cornell,[†] Piotr Cieplak,[‡] Christopher I. Bayly,[§] Ian R. Gould,[⊥] Kenneth M. Merz, Jr.,[‖] David M. Ferguson,[&] David C. Spellmeyer,[#] Thomas Fox, James W. Caldwell, and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Abstract: We present the derivation of a new molecular mechanical force field for simulating the structures, conformational energies, and interaction energies of proteins, nucleic acids, and many related organic molecules in condensed phases. This effective two-body force field is the successor to the Weiner *et al.* force field and was developed with some of the same philosophies, such as the use of a simple diagonal potential function and electrostatic potential fit atom centered charges. The need for a 10−12 function for representing hydrogen bonds is no longer necessary due to the improved performance of the new charge model and new van der Waals parameters. These new charges are determined using a 6-31G* basis set and restrained electrostatic potential (RESP) fitting and have been shown to reproduce interaction energies, free energies of solvation, and conformational energies of simple small molecules to a good degree of accuracy. Furthermore, the new RESP charges exhibit less variability as a function of the molecular conformation used in the charge determination. The new van der Waals parameters have been derived from liquid simulations and include hydrogen parameters which take into account the effects of any geminal electronegative atoms. The bonded parameters developed by Weiner *et al.* were modified as necessary to reproduce experimental vibrational frequencies and structures. Most of the simple dihedral parameters have been retained from Weiner *et al.*, but a complex set of ϕ and ψ parameters which do a good job of reproducing the energies of the low-energy conformations of glycyl and alanyl dipeptides has been developed for the peptide backbone.

The 1995 AMBER Force Field Equation

$$U(\mathbf{r_1}, \mathbf{r_2}, \cdots, \mathbf{r_N}) = \sum_{all\ bonds} k_b (l - l_0)^2$$

$$+ \sum_{all\ angles} k_\theta (\theta - \theta_0)^2$$

$$+ \sum_{all\ dihedrals} \frac{V_n}{2} \left[1 + \cos(n\phi + \gamma) \right]$$

$$+ \sum_{i<j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right]$$

$$+ \sum_{i<j} \left[\frac{q_i q_j}{r_{ij}} \right]$$
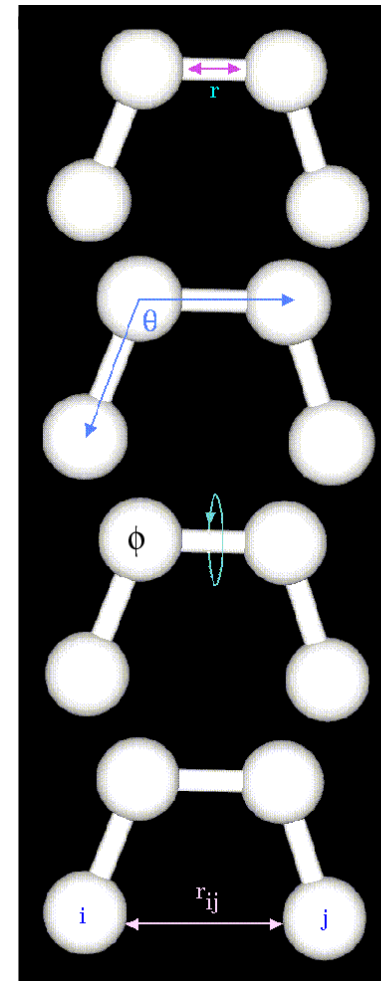
$U_{bond} =$
$\sum_{bonds} K_r (r - r_{eq})^2$

$U_{angles} =$
$\sum_{angles} K_\theta (\theta - \theta_{eq})^2$

$U_{dihedral} =$
$\sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]$

$U_{nonbond} =$
$\sum_{i<j} \left\{ \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] \right.$
or
$\left. \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right] \right\}$
hydrogen-bonding pair
$+ \sum_{i<j} \frac{q_i q_j}{\epsilon r_{ij}}$

Not All the Carbons Are the Same:
Atom Types in Classical Force Field Models

- Atom types are basic building blocks of force field-type molecular modeling.

- Atom types define the connectivity of this type of atom to other atoms.

- Force field parameters, except the partial charges on the atoms, are defined in terms of atom types.

- It is ideal to keep the number of atom types as small as possible, yet reproduce the experimental data as much as possible.

AMBER Atom Types (I)

atom	type	description
carbon	CT	any sp^3 carbon
	C	any carbonyl sp^2 carbon
	CA	any aromatic sp^2 carbon and (Cϵ of Arg)
	CM	any sp^2 carbon, double bonded
	CC	sp^2 aromatic in 5-membered ring with one substituent + next to nitrogen (Cγ in His)
	CV	sp^2 aromatic in 5-membered ring next to carbon and lone pair nitrogen (e.g. Cδ in His (δ))
	CW	sp^2 aromatic in 5-membered ring next to carbon and NH (e.g. Cδ in His (ϵ) and in Trp)
	CR	sp^2 aromatic in 5-membered ring next to two nitrogens (Cγ and Cϵ in His)
	CB	sp^2 aromatic at junction of 5- and 6-membered rings (Cδ in Trp) and both junction atoms in Ade and Gua
	C*	sp^2 aromatic in 5-membered ring next to two carbons (e.g. Cγ in Trp)
	CN	sp^2 junction between 5- and 6-membered rings and bonded to CH and NH (Cϵ in Trp)
	CK	sp^2 carbon in 5-membered aromatic between N and N-R (C8 in purines)
	CQ	sp^2 carbon in 6-membered ring between lone pair nitrogens (e.g. C2 in purines)

AMBER Atom Types (II)

nitrogen	N	sp^2 nitrogen in amides
	NA	sp^2 nitrogen in aromatic rings with hydrogen attached (e.g. protonated His, Gua, Trp)
	NB	sp^2 nitrogen in 5-membered ring with lone pair (e.g. N7 in purines)
	NC	sp^2 nitrogen in 6-membered ring with lone pair (e.g. N3 in purines)
	N*	sp^2 nitrogen in 5-membered ring with carbon substituent (in purine nucleosides)
	N2	sp^2 nitrogen of aromatic amines and guanidinium ions
	N3	sp^3 nitrogen
oxygen	OW	sp^3 oxygen in TIP3P water
	OH	sp^3 oxygen in alcohols, tyrosine, and protonated carboxylic acids
	OS	sp^3 oxygen in ethers
	O	sp^2 oxygen in amides
	O2	sp^2 oxygen in anionic acids
sulfur	S	sulfur in methionine and cysteine
	SH	sulfur in cysteine
phosphorus	P	phosphorus in phosphates

AMBER Atom Types (III)

hydrogen   H   H attached to N

HW   H in TIP3P water

HO   H in alcohols and acids

HS   H attached to sulfur

HA   H attached to aromatic carbon

HC   H attached to aliphatic carbon with no electron-withdrawing substituents

H1   H attached to aliphatic carbon with one electron-withdrawing substituent

H2   H attached to aliphatic carbon with two electron-withdrawing substituents

H3   H attached to aliphatic carbon with three electron-withdrawing substituents

HP   H attached to carbon directly bonded to formally positive atoms (e.g. C next to NH_3^+ of lysine)

H4   H attached to aromatic carbon with one electronegative neighbor (e.g. hydrogen on C5 of Trp, C6 of Thy)

H5   H attached to aromatic carbon with two electronegative neighbors (e.g. H8 of Ade and Gua and H2 of Ade)

AMBER Amino Acids (I)

AMBER Amino Acids (II)

VAL

0.2719 H
0.0969 H1
O −0.5679
−0.4157 N
CT −0.0875
C 0.5973
.0297 HC — CT — CT — HC 0.0791
−0.3192
0.2985
HC 0.0791
HC 0.0791
.0791 HC — CT — HC 0.0791
−0.3192
HC 0.0791

NME

0.2719 H
H1 0.0976
−0.4157 N — CT −0.1490 — H1 0.0976
H1 0.0976
H1 0.0976

HID

0.2719 H
0.0881 H1
O −0.5679
−0.4157 N
CT 0.0188
C 0.5973
0.0402 HC — CT — HC 0.0402
−0.0462
CC −0.0266
0.1147 H4
CV 0.1292
NA −0.3811
H 0.3649
−0.5727 NB
CR 0.2057
H5 0.1392

HIE

0.2719 H
0.1360 H1
O −0.5679
−0.4157 N
CT −0.0581
C 0.5973
0.0367 HC — CT — HC 0.0367
−0.0074
CC 0.1868
0.1862 H4
CW −0.2207
NB −0.5432
−0.2795 NA
CR 0.1635
0.3339 H
H5 0.1435

ACE

0.1123 HC
O −0.5679
−0.3662
0.1123 HC — CT — C 0.5972
0.1123 HC

SER

0.2719 H
0.0843 H1
O −0.5679
−0.4157 N
CT −0.0249
C 0.5973
0.0352 H1 — CT — H1 0.0352
0.2117
−0.6546 OH — HO 0.4275

THR

0.2719 H
0.1007 H1
O −0.5679
−0.4157 N
CT −0.0389
C 0.5973
−0.6761
0.0043 H1 — CT — OH — HO 0.4102
0.3654
0.0642 HC — CT — HC 0.0642
−0.2438
HC 0.0642

CYX

0.2719 H
0.0766 H1
O −0.5679
−0.4157 N
CT 0.0429
C 0.5973
0.0910 H1 — CT — H1 0.0910
−0.0790
S −0.1081

CYS

0.2719 H
0.1124 H1
O −0.5679
−0.4157 N
CT 0.0213
C 0.5973
0.1112 H1 — CT — H1 0.1112
−0.1231
−0.3119 SH — HS 0.1933

HIP

0.2747 H
0.1212 H1
O −0.5894
−0.3479 N
CT −0.1354
C 0.7341
0.0810 HC — CT — HC 0.0810
−0.0414
CC −0.0012
0.2317 H4
CW −0.1141
NA −0.1513
H 0.3866
−0.1718 NA
CR −0.0170
0.3911 H
H5 0.2681

AMBER Amino Acids (III)

MET

PRO

GLY_ct

ARG

LYS

PHE

TYR

TRP

MET_nt

Bond parameters

```
C -C    310.0    1.525    Junmei et al, 1999
C -CA   469.0    1.409    JCC,7,(1986),230; (not used any more
 in TYR)
C -CB   447.0    1.419    JCC,7,(1986),230; GUA
C -CM   410.0    1.444    JCC,7,(1986),230; THY,URA
C -CT   317.0    1.522    JCC,7,(1986),230; AA
C -N    490.0    1.335    JCC,7,(1986),230; AA
C -N*   424.0    1.383    JCC,7,(1986),230; CYT,URA
C -NA   418.0    1.388    JCC,7,(1986),230; GUA.URA
C -NC   457.0    1.358    JCC,7,(1986),230; CYT
C -O    570.0    1.229    JCC,7,(1986),230; AA,CYT,GUA,THY,URA
C -O2   656.0    1.250    JCC,7,(1986),230; GLU,ASP
C -OH   450.0    1.364    JCC,7,(1986),230; (not used any more
 for TYR)
C -OS   450.0    1.323    Junmei et al, 1999
C -H4   367.0    1.080    Junmei et al, 1999
C -H5   367.0    1.080    Junmei et al, 1999
```

Angle parameters

```
C -C -O      80.0      120.00    Junmei et al, 1999 acrolein
C -C -OH     80.0      120.00    Junmei et al, 1999
CA-C -CA     63.0      120.00    changed from 85.0  bsd on C6H6 nmodes;
AA
CA-C -OH     70.0      120.00    AA (not used in tyr)
CB-C -NA     70.0      111.30    NA
CB-C -O      80.0      128.80
CM-C -NA     70.0      114.10
CM-C -O      80.0      125.30
CT-C -O      80.0      120.40
CT-C -O2     70.0      117.00
CT-C -N      70.0      116.60    AA general
CT-C -CT     63.0      117.00    Junmei et al, 1999
CT-C -OS     80.0      115.00    Junmei et al, 1999
CT-C -OH     80.0      110.00    Junmei et al, 1999
```

Torsion parameters

X -C*-CW-X	4	26.10	180.0	2.	intrpol.bsd.on
C6H6					
X -CR-NA-X	4	9.30	180.0	2.	JCC, 7, (1986), 230
X -CR-NB-X	2	10.00	180.0	2.	JCC, 7, (1986), 230
X -CV-NB-X	2	4.80	180.0	2.	JCC, 7, (1986), 230
X -CW-NA-X	4	6.00	180.0	2.	JCC, 7, (1986), 230
X -OH-P -X	3	0.75	0.0	3.	JCC, 7, (1986), 230
X -OS-P -X	3	0.75	0.0	3.	JCC, 7, (1986), 230
N -CT-C -N	1	1.700	180.000	-1.	
N -CT-C -N	1	2.000	180.000	2.	
C -N -CT-C	1	0.850	180.000	-2.	
C -N -CT-C	1	0.800	0.000	1.	
CT-CT-N -C	1	0.50	180.0	-4.	phi,psi,parm94
CT-CT-N -C	1	0.15	180.0	-3.	phi,psi,parm94
CT-CT-N -C	1	0.00	0.0	-2.	JCC, 7, (1986), 230
CT-CT-N -C	1	0.53	0.0	1.	phi,psi,parm94

Van der Waals parameters

O	1.6612	0.2100	OPLS
O2	1.6612	0.2100	OPLS
OW	1.7683	0.1520	TIP3P water model
OH	1.7210	0.2104	OPLS
OS	1.6837	0.1700	OPLS ether
C*	1.9080	0.0860	Spellmeyer
CT	1.9080	0.1094	Spellmeyer
C	1.9080	0.0860	OPLS
N	1.8240	0.1700	OPLS
N3	1.8240	0.1700	OPLS
S	2.0000	0.2500	W. Cornell CH3SH and CH3SCH3 FEP's
SH	2.0000	0.2500	W. Cornell CH3SH and CH3SCH3 FEP's
P	2.1000	0.2000	JCC, 7, (1986), 230;
IM	2.47	0.1	Cl- Smith & Dang, JCP 1994,100:5,3757
Li	1.1370	0.0183	Li+ Aqvist JPC 1990,94,8021. (adapted)
IP	1.8680	0.00277	Na+ Aqvist JPC 1990,94,8021. (adapted)
Na	1.8680	0.00277	Na+ Aqvist JPC 1990,94,8021. (adapted)
K	2.6580	0.000328	K+ Aqvist JPC 1990,94,8021. (adapted)

How to derive the partial charges on the atoms?

An Approach to Computing Electrostatic Charges for Molecules

U. Chandra Singh and Peter A. Kollman*

School of Pharmacy, Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143
Received 10 February 1983; accepted 17 May 1983

We present an approach for deriving net atomic charges from *ab initio* quantum mechanical calculations using a least squares fit of the quantum mechanically calculated electrostatic potential to that of the partial charge model. Our computational approach is similar to those presented by Momany [*J. Phys. Chem.*, **82**, 592 (1978)], Smit, Derissen, and van Duijneveldt [*Mol. Phys.*, **37**, 521 (1979)], and Cox and Williams [*J. Comput. Chem.*, **2**, 304 (1981)], but differs in the approach to choosing the positions for evaluating the potential. In this article, we present applications to the molecules H_2O, CH_3OH, $(CH_3)_2O$, H_2CO, NH_3, $(CH_3O)_2PO_2^-$, deoxyribose, ribose, adenine, 9-CH_3 adenine, thymine, 1-CH_3 thymine, guanine, 9-CH_3 guanine, cytosine, 1-CH_3 cytosine, uracil, and 1-CH_3 uracil. We also address the question of inclusion of "lone pairs," their location and charge.

ARTICLES

A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model

Christopher I. Bayly,† Piotr Cieplak,‡ Wendy D. Cornell,§ and Peter A. Kollman*

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143

We present a new approach to generating electrostatic potential (ESP) derived charges for molecules. The major strength of electrostatic potential derived charges is that they optimally reproduce the intermolecular interaction properties of molecules with a simple two-body additive potential, provided, of course, that a suitably accurate level of quantum mechanical calculation is used to derive the ESP around the molecule. Previously, the major weaknesses of these charges have been that they were not easily transferable between common functional groups in related molecules, they have often been conformationally dependent, and the large charges that frequently occur can be problematic for simulating intramolecular interactions. Introducing restraints in the form of a penalty function into the fitting process considerably reduces the above problems, with only a minor decrease in the quality of the fit to the quantum mechanical ESP. Several other refinements in addition to the restrained electrostatic potential (RESP) fit yield a general and algorithmic charge fitting procedure for generating atom-centered point charges. This approach can thus be recommended for general use in molecular mechanics, molecular dynamics, and free energy calculations for any organic or bioorganic system.

RESP: a two-stage restrained electrostatic fit charge model

1) The geometry of the molecule was taken from experimental structure or theoretical approaches.

2) The quantum mechanical electrostatic potentials (ESP) based on the 6-31G* basis set were evaluated at the shells of points with the density of 1 point/$Å^2$ at each of 1.4, 1.6, 1.8 and 2.0 times the van der Waals radii of the molecule.

3) Atom-centered model charges were derived from minimizing the differences between the reproduced ESP and the original QM ESP plus the deviation from the minimum of a hyperbolic restraint function.

4) In the first-stage fitting process, no forced symmetry is applied and a weak restraints (a=0.0005) is used. In the second-stage fitting, the charges on equivalent atoms are forced to be the same, and a strong restraint (a=0.001) is used.

Restraint functions used in RESP

$$\chi^2 = \chi^2_{esp} + \chi^2_{rstr}$$

$$\chi^2_{esp} = \sum_i \left(V_i - \hat{V}_i \right)^2$$

$$\hat{V}_i = \sum_j \frac{q_j}{r_{ij}}$$

V_i: calculated by QM
i: index for points with ESP values

q_j: model charge at atom j

$$\chi^2_{rstr} = a \sum_j \left(\sqrt{q_j^2 + b^2} - b \right)$$

a: strength of restraint
b: tightness of the hyperbola around its minimum. Default, 0.1 electron charge.

Application of RESP Charges To Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation

Wendy D. Cornell,[†] Piotr Cieplak,[‡] Christopher I. Bayly,[§] and Peter A. Kollman[*]

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Abstract: We apply a new restrained electrostatic potential fit charge model (two-stage RESP) to conformational analysis and the calculation of intermolecular interactions. Specifically, we study conformational energies in butane, methyl ethyl thioether, three simple alcohols, three simple amines, and 1,2-ethanediol as a function of charge model (two-stage RESP *vs* standard ESP) and 1–4 electrostatic scale factor. We demonstrate that the two-stage RESP model with a 1–4 electrostatic scale factor of ~1/1.2 is a very good model, as evaluated by comparison with high-level *ab initio* calculations. For methanol and *N*-methylacetamide interactions with TIP3P water, the two-stage RESP model leads to hydrogen bonds only slightly weaker than found with the standard ESP changes. In tests on DNA base pairs, the two-stage RESP model leads to hydrogen bonds which are ~1 kcal/mol weaker than those calculated with the standard ESP charges but closer in magnitude to the best current available *ab initio* calculations. Furthermore, the two-stage RESP charges, unlike the standard ESP charges, reproduce the result that Hoogsteen hydrogen bonding is stronger than Watson–Crick hydrogen bonding for adenine–thymine base pairs. The free energies of solvation of both methanol and *trans-N*-methylacetamide were also calculated for the standard ESP and two-stage RESP models and both were in good agreement with experiment. We have combined the use of two-stage RESP charges with multiple conformational fitting—recently employed using standard ESP charges as described by Reynolds, *et al.* (*J. Am. Chem. Soc.* **1992**, *114*, 9075)—in studies of conformationally dependent dipole moments and energies of propylamine. We find that the combination of these approaches is synergistic in leading to useful charge distributions for molecular simulations. Two-stage RESP charges thus reproduce both intermolecular and intramolecular energies and structures quite well, making this charge model a critical advancement in the development of a general force field for modeling biological macromolecules and their ligands, both in the gas phase and in solution.

How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?

JUNMEI WANG, PIOTR CIEPLAK,* PETER A. KOLLMAN

Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-0446

ABSTRACT: In this study, we present conformational energies for a molecular mechanical model (Parm99) developed for organic and biological molecules using the restrained electrostatic potential (RESP) approach to derive the partial charges. This approach uses the simple "generic" force field model (Parm94), and attempts to add a minimal number of extra Fourier components to the torsional energies, but doing so only when there is a physical justification. The results are quite encouraging, not only for the 34-molecule set that has been studied by both the highest level *ab initio* model (GVB/LMP2) and experiment, but also for the 55-molecule set for which high-quality experimental data are available. Considering the 55 molecules studied by all the force field models for which there are experimental data, the average absolute errors (AAEs) are 0.28 (this model), 0.52 (MM3), 0.57 (CHARMm [MSI]), and 0.43 kcal/mol (MMFF). For the 34-molecule set, the AAEs of this model versus experiment and *ab initio* are 0.28 and 0.27 kcal/mol, respectively. This is a lower error than found with MM3 and CHARMm, and is comparable to that found with MMFF (0.31 and 0.22 kcal/mol). We also present two examples of how well the torsional parameters are transferred from the training set to the test set. The absolute errors of molecules in the test set are only slightly larger than in the training set (differences of <0.1 kcal/mol). Therefore, it can be concluded that a simple "generic" force field with a limited number of specific torsional parameters can describe intra- and intermolecular interactions, although all comparison molecules were selected from our 82-compound training set. We also show how this effective two-body model can be extended for use with a nonadditive force field (NAFF), both with and without lone pairs. Without changing the torsional parameters, the use of more accurate charges and polarization leads to an increase in average absolute error compared with experiment, but adjustment of the parameters restores the level of agreement found with the additive model. After reoptimizing the Ψ, Φ torsional parameters in peptides using alanine dipeptide (6 conformational pairs) and alanine tetrapeptide (11 conformational pairs), the new model gives better energies than the Cornell et al. (*J Am Chem Soc* 1995, 117, 5179–5197) force field. The average absolute error of this model for high-level *ab initio* calculation is 0.82 kcal/mol for alanine dipeptide and tetrapeptide as compared with 1.80 kcal/mol for the Cornell et al. model. For nucleosides, the new model also gives improved energies compared with the Cornell et al. model. To optimize force field parameters, we developed a program called *parmscan*, which can iteratively scan the torsional parameters in a systematic manner and finally obtain the best torsional potentials. Besides the organic molecules in our test set, *parmscan* was also successful in optimizing the Ψ, Φ torsional parameters in peptides to significantly improve agreement between molecular mechanical and high-level *ab initio* energies.   © 2000 John Wiley & Sons, Inc.   J Comput Chem 21: 1049–1074, 2000

Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method

ARAZ JAKALIAN,[1] BRUCE L. BUSH,[2] DAVID B. JACK,[1] CHRISTOPHER I. BAYLY[3]

[1]*Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada*
[2]*Merck Research Laboratories, Department of Molecular Design and Diversity, Rahway, New Jersey, USA*
[3]*Merck Frosst Canada, Inc., PO Box 1005, Pointe-Claire–Dorval, Québec H9R 4P8, Canada*

ABSTRACT: The AM1-BCC method quickly and efficiently generates high-quality atomic charges for use in condensed-phase simulations. The underlying features of the electron distribution including formal charge and delocalization are first captured by AM1 atomic charges for the individual molecule. Bond charge corrections (BCCs), which have been parameterized against the HF/6-31G* electrostatic potential (ESP) of a training set of compounds containing relevant functional groups, are then added using a formalism identical to the consensus BCI (bond charge increment) approach. As a proof of the concept, we fit BCCs simultaneously to 45 compounds including O-, N-, and S-containing functionalities, aromatics, and heteroaromatics, using only 41 BCC parameters. AM1-BCC yields charge sets of comparable quality to HF/6-31G* ESP-derived charges in a fraction of the time while reducing instabilities in the atomic charges compared to direct ESP-fit methods. We then apply the BCC parameters to a small "test set" consisting of aspirin, D-glucose, and eryodictyol; the AM1-BCC model again provides atomic charges of quality comparable with HF/6-31G* RESP charges, as judged by an increase of only 0.01 to 0.02 atomic units in the root-mean-square (RMS) error in ESP. Based on these encouraging results, we intend to parameterize the AM1-BCC model to provide a consistent charge model for any organic or biological molecule.
© 2000 John Wiley & Sons, Inc.   J Comput Chem 21: 132–146, 2000

Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation

ARAZ JAKALIAN,[1] DAVID B. JACK,[2] CHRISTOPHER I. BAYLY[3]

[1]Boehringer Ingelheim (Canada) Ltd Research and Development, 2100 Rue Cunard,
Laval, Quebec, Canada, H7S 2G5
[2]Department of Chemistry and Biochemistry, Concordia University, 1455 de Maisonneuve Blvd.
Ouest, Montréal, Québec, Canada, H3G 1M8
[3]Merck Frosst Canada & Co., 16711 Trans-Canada Hwy, Kirkland, Québec, Canada, H9H 3L1

J Comput Chem 23: 1623–1641, 2002

Abstract: We present the first global parameterization and validation of a novel charge model, called AM1-BCC, which quickly and efficiently generates high-quality atomic charges for computer simulations of organic molecules in polar media. The goal of the charge model is to produce atomic charges that emulate the HF/6-31G* electrostatic potential (ESP) of a molecule. Underlying electronic structure features, including formal charge and electron delocalization, are first captured by AM1 population charges; simple additive bond charge corrections (BCCs) are then applied to these AM1 atomic charges to produce the AM1-BCC charges. The parameterization of BCCs was carried out by fitting to the HF/6-31G* ESP of a training set of >2700 molecules. Most organic functional groups and their combinations were sampled, as well as an extensive variety of cyclic and fused bicyclic heteroaryl systems. The resulting BCC parameters allow the AM1-BCC charging scheme to handle virtually all types of organic compounds listed in The Merck Index and the NCI Database. Validation of the model was done through comparisons of hydrogen-bonded dimer energies and relative free energies of solvation using AM1-BCC charges in conjunction with the 1994 Cornell et al. forcefield for AMBER.[13] Homo- and hetero-dimer hydrogen-bond energies of a diverse set of organic molecules were reproduced to within 0.95 kcal/mol RMS deviation from the *ab initio* values, and for DNA dimers the energies were within 0.9 kcal/mol RMS deviation from *ab initio* values. The calculated relative free energies of solvation for a diverse set of monofunctional isosteres were reproduced to within 0.69 kcal/mol of experiment. In all these validation tests, AMBER with the AM1-BCC charge model maintained a correlation coefficient above 0.96. Thus, the parameters presented here for use with the AM1-BCC method present a fast, accurate, and robust alternative to HF/6-31G* ESP-fit charges for general use with the AMBER force field in computer simulations involving organic small molecules.
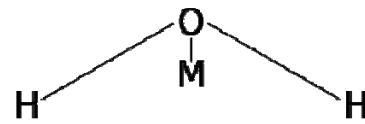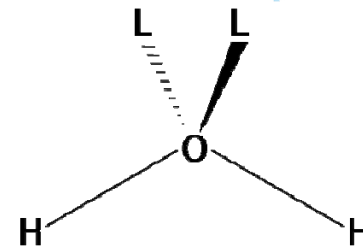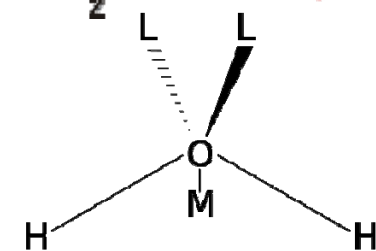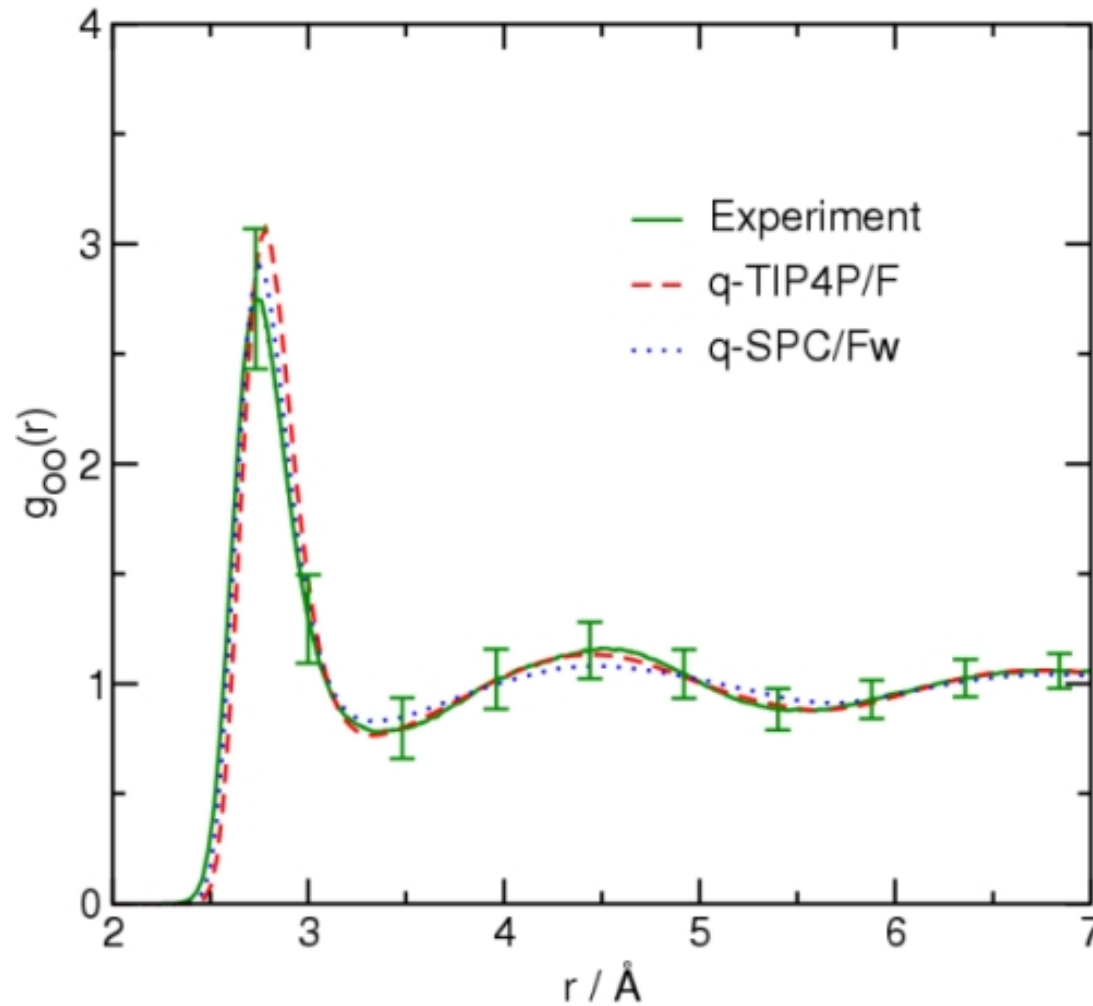
Water Models
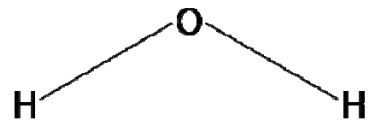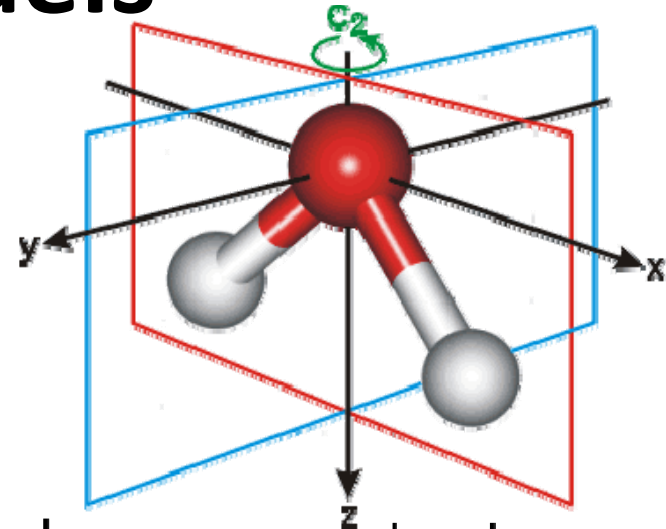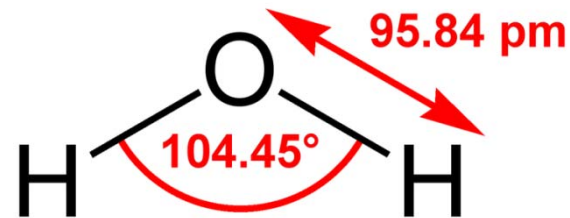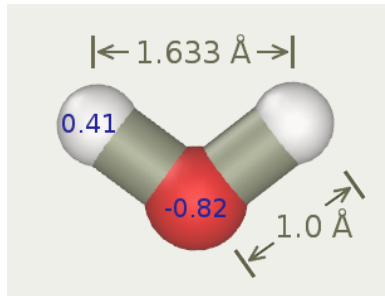
3-site

TIP3P
SPC

4-site

TIP4P

5-site

ST2
TIP5P

6-site

Nada & van der Eerden

$$E_{ab} = \sum_{i}^{on\ a} \sum_{j}^{on\ b} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^{6}}$$

Simulated and Experimental Radial Distribution Function of Water

Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew

Hans W. Horn,[a] William C. Swope, and Jed W. Pitera
IBM Almaden Research Center, San Jose, California 95120

Jeffry D. Madura and Thomas J. Dick
Department of Chemistry and Biochemistry, Center for Computational Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282

Greg L. Hura
Graduate Group in Biophysics, University of California, Berkeley, California 94720

Teresa Head-Gordon
Graduate Group in Biophysics and Department of Bioengineering, University of California, Berkeley, California 94720

A re-parameterization of the standard TIP4P water model for use with Ewald techniques is introduced, providing an overall global improvement in water properties relative to several popular nonpolarizable and polarizable water potentials. Using high precision simulations, and careful application of standard analytical corrections, we show that the new TIP4P-Ew potential has a density maximum at ~1 °C, and reproduces experimental bulk-densities and the enthalpy of vaporization, ΔH_{vap}, from -37.5 to 127 °C at 1 atm with an absolute average error of less than 1%. Structural properties are in very good agreement with x-ray scattering intensities at temperatures between 0 and 77 °C and dynamical properties such as self-diffusion coefficient are in excellent agreement with experiment. The parameterization approach used can be easily generalized to rehabilitate any water force field using available experimental data over a range of thermodynamic points. © 2004 American Institute of Physics. [DOI: 10.1063/1.1683075]

Times Cited: 1109 (2018/3/15)

A modified TIP3P water potential for simulation with Ewald summation

Daniel J. Price and Charles L. Brooks III[a)]
Department of Molecular Biology (TPC-6), The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

The charges and Lennard-Jones parameters of the TIP3P water potential have been modified to improve its performance under the common condition for molecular dynamics simulations of using Ewald summation in lieu of relatively short nonbonded truncation schemes. These parameters were optimized under the condition that the hydrogen atoms do not have Lennard-Jones parameters, thus making the model independent of the combining rules used for the calculation of nonbonded, heteroatomic interaction energies, and limiting the number of Lennard-Jones calculations required. Under these conditions, this model provides accurate density ($\rho = 0.997$ g/ml) and heat of vaporization ($\Delta H_{vap} = 10.53$ kcal/mol) at 25 °C and 1 atm, but also provides improved structure in the second peak of the O–O radial distribution function and improved values for the dielectric constant ($\varepsilon_0 = 89$) and the diffusion coefficient ($D = 4.0 \times 10^{-5}$ cm^2/s) relative to the original parametrization. Like the original parameterization, however, this model does not show a temperature density maximum. Several similar models are considered with the additional constraint of trying to match the performance of the optimized potentials for liquid simulation atom force field to that obtained when using the simulation conditions under which it was originally designed, but no model was entirely satisfactory in reproducing the relative difference in free energies of hydration between the model compounds, phenol and benzene. Finally, a model that incorporates a long-range correction for truncated Lennard-Jones interactions is presented, which provides a very accurate dielectric constant ($\varepsilon_0 = 76$), however, the improvement in this estimate is on the same order as the uncertainty in the calculation. © *2004 American Institute of Physics.*

Times Cited: 341 (2018/3/15)

Validation of molecular dynamics simulations with NMR experiments

Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements

Kyle A. Beauchamp,[†] Yu-Shan Lin,[‡] Rhiju Das,[†,§] and Vijay S. Pande[†,‡,*]

[†]Biophysics Program, [‡]Chemistry Department, Stanford University, Stanford, California, United States
[§]Biochemistry Department, Stanford University, Stanford, California, United States

Ⓢ *Supporting Information*

Times Cited: 128 (2016/4/25)

ABSTRACT: Recent hardware and software advances have enabled simulation studies of protein systems on biophysically relevant time scales, often revealing the need for improved force fields. Although early force field development was limited by the lack of direct comparisons between simulation and experiment, recent work from several laboratories has demonstrated direct calculation of NMR observables from protein simulations. Here, we quantitatively evaluate 11 recent molecular dynamics force fields in combination with 5 solvent models against a suite of 524 chemical shift and J coupling ($^3JH_NH_\alpha$, $^3JH_NC_\beta$, $^3JH_\alpha C'$, $^3JH_NC'$, and $^3JH_\alpha N$) measurements on dipeptides, tripeptides, tetra-alanine, and ubiquitin. Of the force fields examined (ff96, ff99, ff03, ff03*, ff03w, ff99sb*, ff99sb-ildn, ff99sb-ildn-phi, ff99sb-ildn-NMR, CHARMM27, and OPLS-AA), two force fields (ff99sb-ildn-phi, ff99sb-ildn-NMR) combining recent side chain and backbone torsion modifications achieved high accuracy in our benchmark. For the two optimal force fields, the calculation error is comparable to the uncertainty in the experimental comparison. This observation suggests that extracting additional force field improvements from NMR data may require increased accuracy in J coupling and chemical shift prediction. To further investigate the limitations of current force fields, we also consider conformational populations of dipeptides, which were recently estimated using vibrational spectroscopy.

Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States

Stefano Piana,*,† Alexander G. Donchev,† Paul Robustelli,† and David E. Shaw*,†,‡

†D. E. Shaw Research, New York, New York 10036, United States

‡Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States

S *Supporting Information*

ABSTRACT: Many proteins can be partially or completely disordered under physiological conditions. Structural characterization of these disordered states using experimental methods can be challenging, since they are composed of a structurally heterogeneous ensemble of conformations rather than a single dominant conformation. Molecular dynamics (MD) simulations should in principle provide an ideal tool for elucidating the composition and behavior of disordered states at an atomic level of detail. Unfortunately, MD simulations using current physics-based models tend to produce

disordered-state ensembles that are structurally too compact relative to experiments. We find that the water models typically used in MD simulations significantly underestimate London dispersion interactions, and speculate that this may be a possible reason for these erroneous results. To test this hypothesis, we create a new water model, TIP4P-D, that approximately corrects for these deficiencies in modeling water dispersion interactions while maintaining compatibility with existing physics-based models. We show that simulations of solvated proteins using this new water model typically result in disordered states that are substantially more expanded and in better agreement with experiment. These results represent a significant step toward extending the range of applicability of MD simulations to include the study of (partially or fully) disordered protein states.
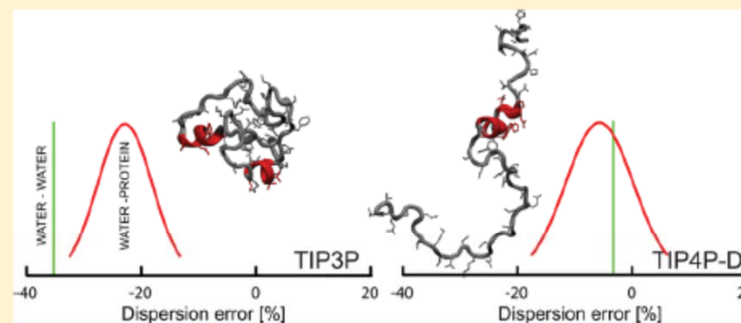
Times Cited: 139 (2018/3/15)

47

Force fields for monovalent and divalent metal cations in TIP3P water based on thermodynamic and kinetic properties

Shavkat Mamatkulov[1] and Nadine Schwierz[2,a]

[1]*Department of Physics, The Centre of Higher Technologies, Tashkent, Uzbekistan*
[2]*Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany*

Metal cations are essential in many vital processes. In order to capture the role of different cations in all-atom molecular dynamics simulations of biological processes, an accurate parametrization is crucial. Here, we develop force field parameters for the metal cations Li^+, Na^+, K^+, Cs^+, Mg^{2+}, Ca^{2+}, Sr^{2+}, and Ba^{2+} in combination with the TIP3P water model that is frequently used in biomolecular simulations. In progressing toward improved force fields, the approach presented here is an extension of previous efforts and allows us to simultaneously reproduce thermodynamic and kinetic properties of aqueous solutions. We systematically derive the parameters of the 12-6 Lennard-Jones potential which accurately reproduces the experimental solvation free energy, the activity derivative, and the characteristics of water exchange from the first hydration shell of the metal cations. In order to reproduce all experimental properties, a modification of the Lorentz-Berthelot combination rule is required for Mg^{2+}. Using a balanced set of solution properties, the optimized force field parameters aim to capture the fine differences between distinct metal cations including specific ion binding affinities and the kinetics of cation binding to biologically important anionic groups. *Published by AIP Publishing.* https://doi.org/10.1063/1.5017694

Comparison of Structural, Thermodynamic, Kinetic and Mass Transport Properties of Mg^{2+} Ion Models Commonly used in Biomolecular Simulations

Maria T. Panteva, George M. Giambaşu, and Darrin M. York*

The prevalence of Mg^{2+} ions in biology and their essential role in nucleic acid structure and function has motivated the development of various Mg^{2+} ion models for use in molecular simulations. Currently, the most widely used models in biomolecular simulations represent a nonbonded metal ion as an ion-centered point charge surrounded by a nonelectrostatic pairwise potential that takes into account dispersion interactions and exchange effects that give rise to the ion's excluded volume. One strategy toward developing improved models for biomolecular simulations is to first identify a Mg^{2+} model that is consistent with the simulation force fields that closely reproduces a range of properties in aqueous solution, and then, in a second step, balance the ion–water and ion–solute interactions by tuning parameters in a pairwise fashion where necessary. The present work addresses the first step in which we compare 17 different nonbonded single-site Mg^{2+} ion models with respect to their ability to simultaneously reproduce structural, thermodynamic, kinetic and mass transport properties in aqueous solution. None of the models based on a 12-6 non-electrostatic nonbonded potential was able to reproduce the experimental radial distribution function, solvation free energy, exchange barrier and diffusion constant. The models based on a 12-6-4 potential offered improvement, and one model in particular, in conjunction with the SPC/E water model, performed exceptionally well for all properties. The results reported here establish useful benchmark calculations for Mg^{2+} ion models that provide insight into the origin of the behavior in aqueous solution, and may aid in the development of next-generation models that target specific binding sites in biomolecules. © 2015 Wiley Periodicals, Inc.

The standard (12-6) potential can be augmented by an additional r^{-4} term to form a (12-6-4) potential[16] as:

$$U_{ij}(r_{ij}) = \varepsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^{6} - 2\kappa R_{ij}^2 \left(\frac{R_{ij}}{r_{ij}} \right)^{4} \right]$$

$$= \frac{A_{ij}}{r_{ij}^{12}} - B_{ij} \left(\frac{1}{r_{ij}^6} + \frac{\kappa}{r_{ij}^4} \right)$$

(3)

where κ is a scaling parameter with units of Å^{-2}. The difference between this potential and the 12-6 potential is the additional attractive term, $\kappa \frac{B_{ij}}{r_{ij}^4}$, which falls off as r^{-4} and mimics the charge-induced dipole interaction.

Table 5. Scaled infinite dilution diffusion coefficients ($\times\ 10^{-5}$ cm^2/s) and hydrodynamic radii (Å) for the Mg^{2+} parameters and the respective solvent viscosities ($\times\ 10^{-4}$ kgm^{-1}s^{-1}). D_o, η and r are D_o^{sim}, η^{sim}, and r^{sim} (Supporting Information Table S5) rescaled by $\frac{D_o^w}{D_o^{\circ}}$ (see discussion in Methods) to correct for errors in the diffusion coefficients of the water models.

Model	Water model	D_o	η	r
Åqvist	SPC/E	0.741 ± 0.057	9.23 ± 0.36	3.19
Li et al. ΔG_{solv}	SPC/E	0.736 ± 0.052	9.16 ± 0.28	3.24
Li et al. $R_{Mg^{2+}-O}$	SPC/E	0.735 ± 0.052	9.67 ± 0.23	3.07
Li et al. CN	SPC/E	0.723 ± 0.052	9.81 ± 0.21	3.08
Li and Merz 12-6-4	SPC/E	0.748 ± 0.044	9.19 ± 0.24	3.18
Mayaan et al.	TIP3P	0.779 ± 0.052	9.18 ± 0.11	3.05
Roux	TIP3P	0.817 ± 0.049	7.64 ± 0.11	3.50
Allnér et al.	TIP3P	0.802 ± 0.068	8.24 ± 0.27	3.30
Babu and Lim	TIP3P	0.798 ± 0.057	8.79 ± 0.28	3.11
Li et al. ΔG_{solv}	TIP3P	0.790 ± 0.065	8.32 ± 0.10	3.32
Li et al. $R_{Mg^{2+}-O}$	TIP3P	0.771 ± 0.057	9.38 ± 0.10	3.02
Li et al. CN	TIP3P	0.783 ± 0.059	9.09 ± 0.29	3.07
Li and Merz 12-6-4	TIP3P	0.813 ± 0.059	8.59 ± 0.41	3.13
Li et al. ΔG_{solv}	TIP4PEw	0.744 ± 0.051	8.70 ± 0.21	3.37
Li et al. $R_{Mg^{2+}-O}$	TIP4PEw	0.779 ± 0.061	7.97 ± 0.18	3.52
Li et al. CN	TIP4PEw	0.740 ± 0.050	8.77 ± 0.10	3.36
Li and Merz 12-6-4	TIP4PEw	0.801 ± 0.052	8.01 ± 0.22	3.40
Experiment		0.706[102]	8.903[103]	

to the minimum energy distances and adiabatic binding energies from gas phase binding energy scans. Most of the models considered either overestimate or underestimate the inner shell water exchange barrier by several kcal/mol. On the other hand, mass transport properties were observed to be somewhat insensitive to the models. The simple 12-6 models were shown to have considerable limitations regardless of the water model used or the specific LJ parameters. These models were not able to simultaneously reproduce both structural and thermodynamic properties with reasonable accuracy. The 12-6-4 models, on the other hand, offer respectable improvement, particularly with respect to matching both the radial distribution function and solvation free energy. One model in particular, the SPC/E 12-6-4 model of Li et al., performs extremely well across all properties (within statistical confidence) despite only being originally parametrized to match structure and thermodynamics.

This detailed analysis of the solution properties predicted by several Mg^{2+} models used in molecular simulations provides a baseline from which to gauge progress and direct future effort. In progressing toward improved force field models for simulations of biomolecules, and in particular RNA, under dif-

Table 3. Summary of structural properties obtained from Mg^{2+}-water pair distribution functions. The first ($i = 1$) and second ($i = 2$) solvation shell peak maximum and minimum positions (r_i^{max}, r_i^{min}), their probabilities [$g(r_i^{max})$, $g(r_i^{min})$] and coordination numbers (CN_i) are compared, along with available experimental data. Distances are in units of Å.

Model	Water Model	r_1^{max}	$g(r_1^{max})$	r_1^{min}	$g(r_1^{min})$	CN_1	r_2^{max}	$g(r_2^{max})$	r_2^{min}	$g(r_2^{min})$	CN_2
Åqvist	SPC/E	1.99	24.9	2.49	0.0	6.0	4.18	2.5	4.93	0.7	13.6
Li et al. ΔG_{solv}	SPC/E	1.95	22.9	2.44	0.0	6.0	4.14	2.5	4.85	0.7	13.1
Li et al. $R_{Mg^{2+}-O}$	SPC/E	2.08	22.9	2.59	0.0	6.0	4.26	2.3	4.96	0.7	13.7
Li et al. CN	SPC/E	2.03	23.8	2.39	0.0	6.0	4.24	2.3	4.9	0.7	13.3
Li and Merz 12-6-4	SPC/E	2.08	26.0	2.48	0.0	6.0	4.22	2.3	4.99	0.8	14.3
Mayaan et al.	TIP3P	2.11	22.8	2.58	0.0	6.0	4.37	2.0	5.28	0.8	16.3
Roux	TIP3P	1.97	24.5	2.44	0.0	6.0	4.16	2.3	4.97	0.7	14.0
Allnér et al.	TIP3P	2.04	23.5	2.53	0.0	6.0	4.29	2.1	5.32	0.7	16.8
Babu and Lim	TIP3P	2.09	23.1	2.54	0.0	6.0	4.33	2.1	5.32	0.7	16.6
Li et al. ΔG_{solv}	TIP3P	1.94	23.8	2.49	0.0	6.0	4.22	2.3	4.99	0.7	14.2
Li et al. $R_{Mg^{2+}-O}$	TIP3P	2.07	23.9	2.49	0.0	6.0	4.28	2.1	5.11	0.7	14.9
Li et al. CN	TIP3P	2.03	24.5	2.49	0.0	6.0	4.25	2.2	5.25	0.7	16.1
Li and Merz 12-6-4	TIP3P	2.08	26.4	2.44	0.0	6.0	4.32	2.1	5.18	0.8	16.3
Li et al. ΔG_{solv}	TIP4PEw	1.89	17.5	2.60	0.0	6.0	4.09	2.5	4.83	0.6	12.9
Li et al. $R_{Mg^{2+}-O}$	TIP4PEw	2.09	22.3	2.54	0.0	6.0	4.31	2.2	5.03	0.7	13.8
Li et al. CN	TIP4PEw	2.04	23.1	2.53	0.0	6.0	4.27	2.3	5.01	0.7	13.8
Li and Merz 12-6-4	TIP4PEw	2.08	26.2	2.49	0.0	6.0	4.26	2.2	4.98	0.8	14.3
Experiment		2.09 ± 0.04 [97]				6.0 [98]	4.1–4.2 [98]				12.0 [98]

Table 4. Summary of data extracted from free energy profiles for inner-sphere water exchange: $R^\dagger$: transition state distance (Å), A: pre-exponential factor (fs^{-1}), $\Delta G^\dagger$: activation free energy (kcal/mol), $k_1^\dagger$: first solvation shell water exchange rate (s^{-1}). Standard deviations come from four consecutive 2 ns segments of data.

Model	Water model	$R^\dagger$	A	$\Delta G^\dagger$	$\log(k_1)$	k_1
Åqvist	SPC/E	2.88 ± 0.01	0.015	13.9 ± 0.1	3.0 ± 0.1	9.5×10^2
Li et al. ΔG_{solv}	SPC/E	2.98 ± 0.01	0.013	12.7 ± 0.3	3.8 ± 0.4	6.4×10^3
Li et al. $R_{Mg^{2+}-O}$	SPC/E	2.76 ± 0.01	0.014	10.9 ± 0.3	5.2 ± 0.4	1.5×10^5
Li et al. CN	SPC/E	2.84 ± 0.01	0.014	12.4 ± 0.1	4.1 ± 0.1	1.1×10^4
Li and Merz 12-6-4	SPC/E	2.66 ± 0.02	0.017	10.2 ± 0.3	5.7 ± 0.4	5.5×10^5
Mayaan et al.	TIP3P	2.72 ± 0.01	0.015	7.8 ± 0.1	7.5 ± 0.1	2.9×10^7
Roux	TIP3P	2.89 ± 0.00	0.014	12.9 ± 0.1	3.7 ± 0.1	4.8×10^3
Allnér et al.	TIP3P	2.75 ± 0.01	0.015	10.9 ± 0.1	5.2 ± 0.1	1.5×10^5
Babu and Lim	TIP3P	2.69 ± 0.02	0.015	8.2 ± 0.1	7.2 ± 0.1	1.5×10^7
Li et al. ΔG_{solv}	TIP3P	2.95 ± 0.01	0.012	12.6 ± 0.2	3.9 ± 0.3	7.2×10^3
Li et al. $R_{Mg^{2+}-O}$	TIP3P	2.70 ± 0.00	0.015	9.6 ± 0.1	6.1 ± 0.1	1.3×10^6
Li et al. CN	TIP3P	2.78 ± 0.00	0.015	11.6 ± 0.1	4.7 ± 0.1	4.5×10^4
Li and Merz 12-6-4	TIP3P	2.68 ± 0.00	0.017	7.5 ± 0.2	7.7 ± 0.3	5.2×10^7
Li et al. ΔG_{solv}	TIP4PEw	3.10 ± 0.01	0.010	8.7 ± 0.1	6.6 ± 0.1	4.2×10^6
Li et al. $R_{Mg^{2+}-O}$	TIP4PEw	2.73 ± 0.01	0.015	9.4 ± 0.2	6.3 ± 0.3	1.9×10^6
Li et al. CN	TIP4PEw	2.84 ± 0.01	0.014	11.5 ± 0.1	4.7 ± 0.1	5.1×10^4
Li and Merz 12-6-4	TIP4PEw	2.63 ± 0.00	0.016	8.4 ± 0.2	7.1 ± 0.3	1.1×10^7
Experiment[101]					5.8	$6.7 ± 0.2 \times 10^5$

Preprocessing the PDB files ……

- Construct the functional unit of the biomolecule(s) of interest.

- Make sure if there are missing residues in the PDB file.

- Check if there are post-translational modifications of amino acids.

- Decide whether the crystal water molecules should be kept.

- Decide whether the metal ion should be loaded, and decide the configuration of the ions.

- Check whether there should be disulfide bonds between cysteines. (CYS:cysteine; CYX: cystine)

- Determine the pH condition for the simulations.

- Predict the protonation states of titratable residues according to the given pH condition.

- Make sure whether the N- and C- terminal should be "capped" or not.

antechamber

- This program suite automates the process of developing force field descriptors for most organic molecules.

- It starts with structures (usually in PDB format), and generates files that can be read into LEaP for use in molecular modeling.

- The force field description that is generated is designed to be compatible with the usual Amber force fields for proteins and nucleic acids.

Automatic atom type and bond type perception in molecular mechanical calculations

Junmei Wang [a,*], Wei Wang [b], Peter A. Kollman [c], David A. Case [d]

[a] College of Chemistry, Peking University, Beijing 100871, China
[b] Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics,
University of California at San Diego, La Jolla, CA 92093-0359, United States
[c] Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, CA 94143-0446, United States
[d] Department of Molecular Biology, TPC15, The Scripps Research Institute, 10550 N. Torrey Pines Rd. La Jolla, CA 92037, United States

Abstract

In molecular mechanics (MM) studies, atom types and/or bond types of molecules are needed to determine prior to energy calculations. We present here an automatic algorithm of perceiving atom types that are defined in a description table, and an automatic algorithm of assigning bond types just based on atomic connectivity. The algorithms have been implemented in a new module of the AMBER packages. This auxiliary module, antechamber (roughly meaning "before AMBER"), can be applied to generate necessary inputs of *leap*—the AMBER program to generate topologies for minimization, molecular dynamics, etc., for most organic molecules. The algorithms behind the manipulations may be useful for other molecular mechanical packages as well as applications that need to designate atom types and bond types.

Development and Testing of a General Amber Force Field

JUNMEI WANG,[1] ROMAIN M. WOLF,[2] JAMES W. CALDWELL,
PETER A. KOLLMAN, DAVID A. CASE[3]

[1]*Encysive Pharmaceuticals Inc., 7000 Fannin, Houston, Texas 77030*
[2]*Novartis Institutes for Biomedical Research, Basle, WSJ-88.10.14, P.O. Box,
CH-4002 Basle, Switzerland*
[3]*Department of Molecular Biology, TPC15, The Scripps Research Institute,
10550 N. Torrey Pines Rd., La Jolla, California 92037*

Abstract: We describe here a general Amber force field (GAFF) for organic molecules. GAFF is designed to be compatible with existing Amber force fields for proteins and nucleic acids, and has parameters for most organic and pharmaceutical molecules that are composed of H, C, N, O, S, P, and halogens. It uses a simple functional form and a limited number of atom types, but incorporates both empirical and heuristic models to estimate force constants and partial atomic charges. The performance of GAFF in test cases is encouraging. In test I, 74 crystallographic structures were compared to GAFF minimized structures, with a root-mean-square displacement of 0.26 Å, which is comparable to that of the Tripos 5.2 force field (0.25 Å) and better than those of MMFF 94 and CHARMm (0.47 and 0.44 Å, respectively). In test II, gas phase minimizations were performed on 22 nucleic acid base pairs, and the minimized structures and intermolecular energies were compared to MP2/6-31G* results. The RMS of displacements and relative energies were 0.25 Å and 1.2 kcal/mol, respectively. These data are comparable to results from Parm99/RESP (0.16 Å and 1.18 kcal/mol, respectively), which were parameterized to these base pairs. Test III looked at the relative energies of 71 conformational pairs that were used in development of the Parm99 force field. The RMS error in relative energies (compared to experiment) is about 0.5 kcal/mol. GAFF can be applied to wide range of molecules in an automatic fashion, making it suitable for rational drug design and database searching.

J Comput Chem 25: 1157–1174, 2004

Explicit solvent vs. implicit solvent
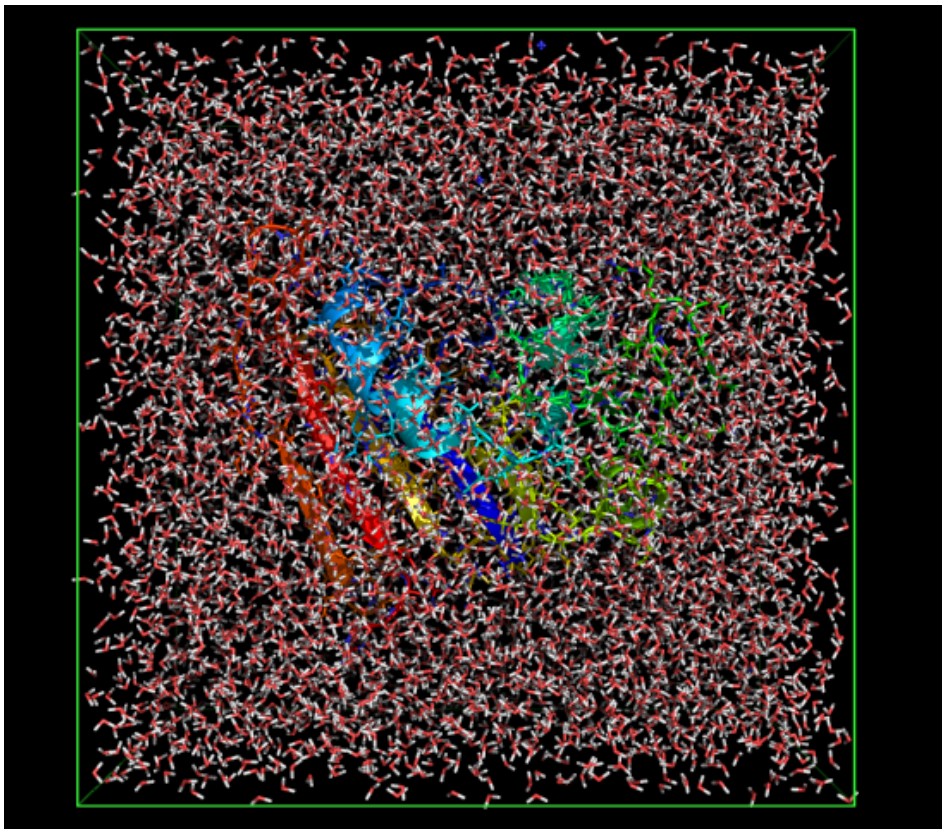
The Poisson-Boltzmann Theory

The Poisson Equation

$$\nabla \cdot \mathbf{E}(\vec{\mathbf{r}}) = \rho_{mol}(\vec{\mathbf{r}})$$

$$\updownarrow$$

$$\nabla \cdot [\varepsilon(\vec{\mathbf{r}}) \cdot \nabla \varphi(\vec{\mathbf{r}})] = -\rho_{mol}(\vec{\mathbf{r}})$$

The Poisson-Boltzmann Equation

$$\nabla \cdot \left(\varepsilon(\vec{\mathbf{r}}) \cdot \nabla \varphi(\vec{\mathbf{r}}) \right) - \kappa^2(\vec{\mathbf{r}}) \sinh\left(\frac{e_c \varphi(\vec{\mathbf{r}})}{k_B T} \right) = \rho_{mol}(\vec{\mathbf{r}})$$

Considering the distribution of ions

$$\nabla \cdot [\varepsilon(\vec{\mathbf{r}}) \cdot \nabla \varphi(\vec{\mathbf{r}})] = -[\rho_{mol}(\vec{\mathbf{r}}) + \rho_{ion}(\vec{\mathbf{r}})]$$

Assuming that the ion distributions obey the Boltzmann law:

$$\rho_{ion}(\vec{\mathbf{r}}) = e \sum_i z_i N_i e^{-Z_i e \varphi / k_B T}$$

$$\approx -\frac{e^2 \varphi}{k_B T} \sum_i z_i^2 N_i$$
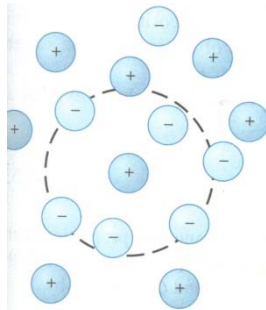
Debye-Hückel screening length

$$d_{DH} = \frac{1}{\kappa} = \left(\frac{\varepsilon k_B T}{e^2 \sum_i c_i z_i^2 N_A} \right)^{1/2}$$
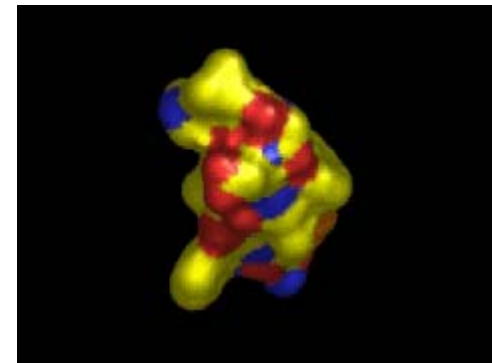
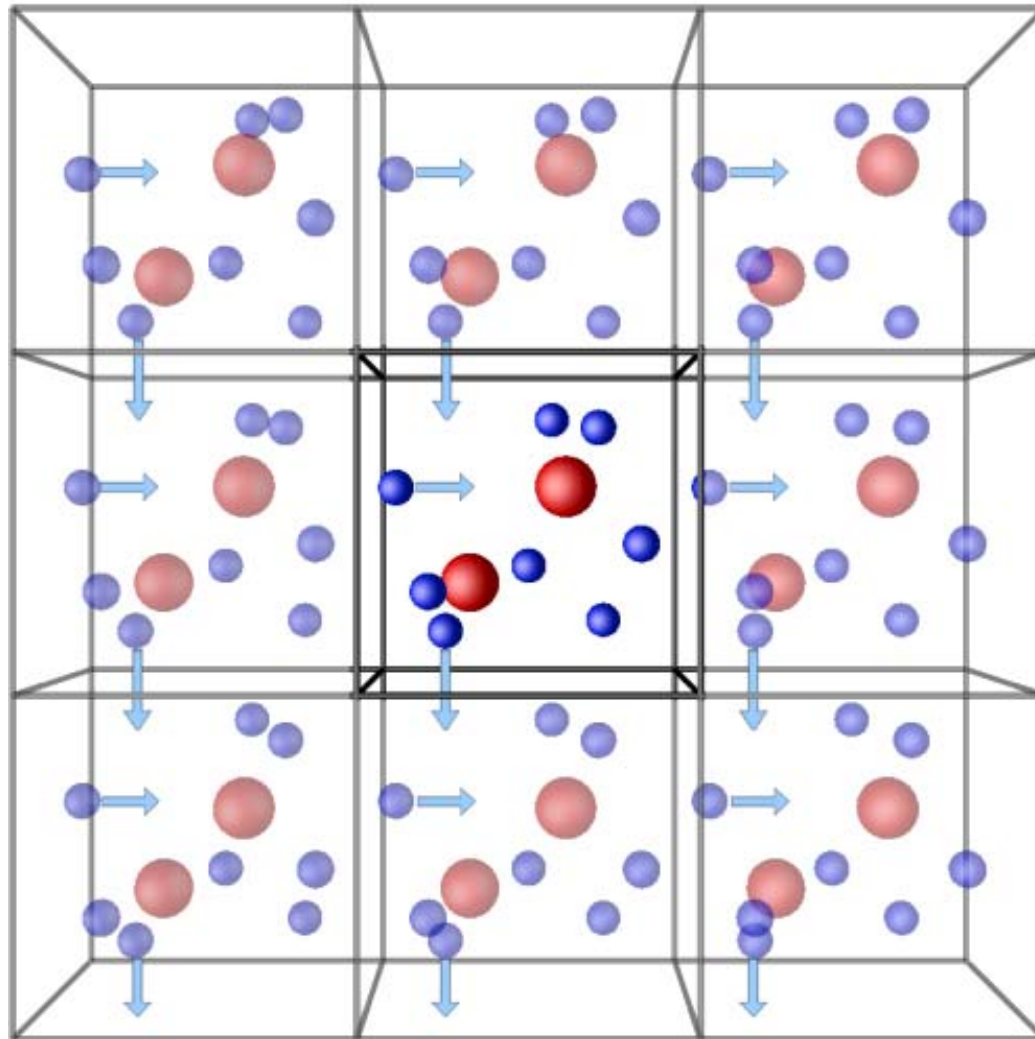# The GB Theory with inclusion of salt effect

The solvation contribution in the linearized Poisson-Boltzmann theory:

$$\Delta G_{pol} = -\left(1 - \frac{e^{-\kappa r_{ij}}}{\varepsilon}\right)\sum_{i=1}^{N-1}\sum_{j=i}^{N}\frac{q_i q_j}{r_{ij}^2}$$

The Generalize Born Equation with inclusion of salt effect:

$$\left(1 - \frac{1}{\varepsilon}\right) \rightarrow \left(1 - \frac{e^{-\kappa f_{GB}}}{\varepsilon}\right)$$

Periodic Boundary Condition

Langevin Dynamics with Implicit Solvent

$$m_i \dot{v}_i = F_i - m_i \xi_i v_i + R_i(t)$$

Fluctuation-dissipation theorem:

$$\langle R_i(t) R_j(t+\tau) \rangle = 2 m_i \gamma_i k_B T_0 \delta(\tau) \delta_{ij}$$

Molecular Dynamics with Explicit Solvent

$$m_i \dot{v}_i = -\nabla_i U + m_i \xi \left(\frac{T_0}{T} - 1 \right) v_i$$
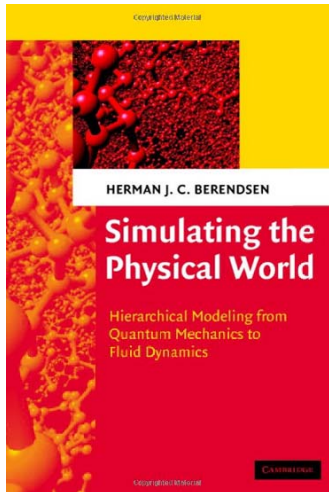
Berendsen weak-coupling thermostat:

$$v_i^{'} = \lambda v_i$$

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right) \right]^{1/2}$$

HERMAN J. C. BERENDSEN

Simulating the Physical World

Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics

Berendsen *et al., J. Chem. Phys.* **81:** 3684-3690 (1984)

Times cited: 19853 (2018/3/16)
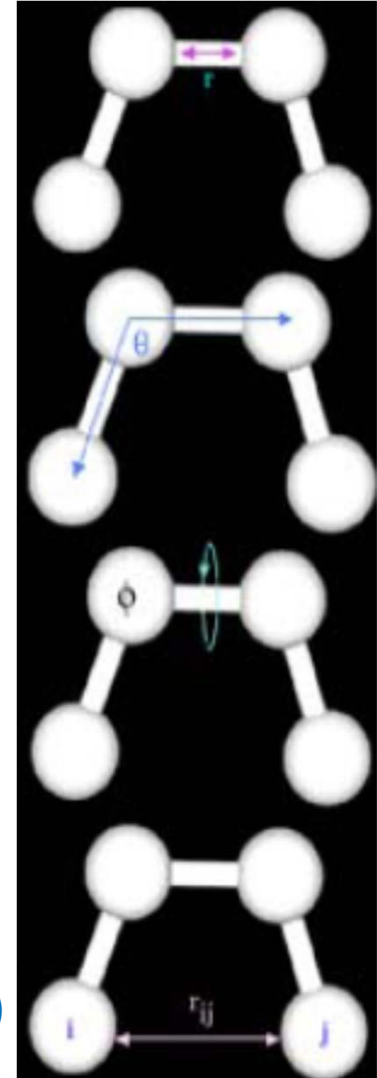
Instantaneous Pressure Tensor

$$\mathcal{P}_{\alpha\beta} = \frac{1}{V}\left(\sum_i \frac{(\mathbf{p}_i)_\alpha (\mathbf{p}_i)_\beta}{m_i} + \sum_i \sum_{j>i} (\mathbf{r}_{ij})_\alpha (\mathbf{F}_{ij})_\beta\right)$$

$$\alpha, \beta = x, y, z$$

For orthorhombic systems:

$$\mathcal{P}_{\alpha\beta} = \begin{cases} 0, & \alpha \neq \beta \\ \mathcal{P}_\alpha, & \alpha = \beta \end{cases}$$

Isotropic Pressure Regulation versus Anisotropic Pressure Regulation

Isotropic scaling:

$$\mu = 1 - \frac{\beta \Delta t}{\tau_P}\left(P - \mathcal{P}_{iso}\right)$$

$$\mathbf{r'} = \mu^{1/3}\,\mathbf{r}$$

$$\mathcal{P}_{iso} = \frac{1}{3}\left(\mathcal{P}_x + \mathcal{P}_y + \mathcal{P}_z\right)$$

Anisotropic scaling:

$$\mu_\alpha = 1 - \frac{\beta \Delta t}{\tau_P}\left(P - \mathcal{P}_\alpha\right)$$

$$\mathbf{r}_\alpha' = \mu_\alpha^{1/3}\,\mathbf{r}_\alpha$$

It is advised to use at least the anisotropic pressure regulation scheme for anisotropic systems like membranes.

Verlet Leapfrog Algorithm

- Good numerical stability
- Time reversible
- Preservation of symplectic form on phase space

Compute velocity at half step
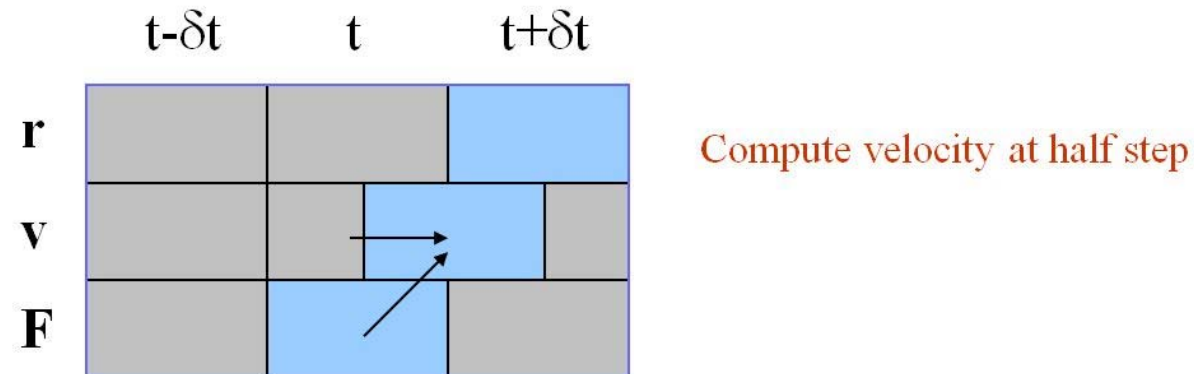
$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2m}\mathbf{F}(t)\delta t^2$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2m}\left[\mathbf{F}(t) + \mathbf{F}(t + \delta t)\right]\delta t$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t + \frac{\delta t}{2}) + \frac{1}{2m}\mathbf{F}(t + \delta t)\delta t$$

$$\mathbf{v}(t + \frac{\delta t}{2}) = \mathbf{v}(t) + \frac{1}{2m}\mathbf{F}(t)\delta t$$

Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*

LOUP VERLET†

Belfer Graduate School of Science, Yeshiva University, New York, New York

(Received 30 January 1967)

The equation of motion of a system of 864 particles interacting through a Lennard-Jones potential has been integrated for various values of the temperature and density, relative, generally, to a fluid state. The equilibrium properties have been calculated and are shown to agree very well with the corresponding properties of argon. It is concluded that, to a good approximation, the equilibrium state of argon can be described through a two-body potential.

thermodynamic quantities will thus be measured in the

¹ W. W. Wood and F. R. Parker, J. Chem. Phys. **27**, 720 (1957).
² B. J. Alder and T. E. Wainwright, J. Chem. Phys. **33**, 1439 (1960).
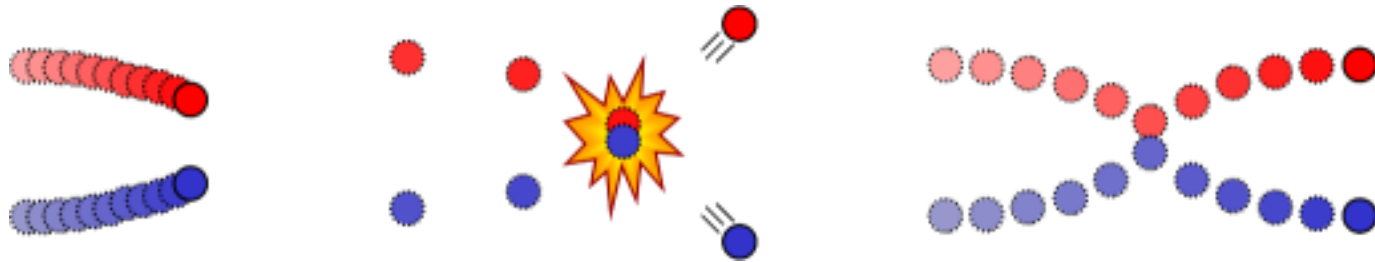³ A. Rahman, Phys. Rev. **136**, A405 (1964).
⁴ L. Verlet (to be published).
⁵ J. L. Lebowitz and J. K. Percus, Phys. Rev. **124**, 1673 (1961).
⁶ J. L. Lebowitz, J. K. Percus, and L. Verlet, Phys. Rev. **153**, 250 (1967).
⁷ A. Michels and H. Wijker, Physica **15**, 627 (1949).

Choosing the time step

Time Scale Limitation

- The time step for integration of Newtonian equation of motion is typically 2 fs.

- Although the time scale of longest simulation being done has been 1 ms, this is still not common for ordinary labs.

- Typical time scales for protein folding are from ms to seconds, if not minutes.

Constrained dynamics: SHAKE

Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes

JEAN-PAUL RYCKAERT*, GIOVANNI CICCOTTI[†], AND
HERMAN J. C. BERENDSEN[‡]

Centre Européen de Calcul Atomique et Moléculaire (CECAM),
Bâtiment 506, Université de Paris XI, 91405 Orsay, France

A numerical algorithm integrating the $3N$ Cartesian equations of motion of a system of N points subject to holonomic constraints is formulated. The relations of constraint remain perfectly fulfilled at each step of the trajectory despite the approximate character of numerical integration. The method is applied to a molecular dynamics simulation of a liquid of 64 n-butane molecules and compared to a simulation using generalized coordinates. The method should be useful for molecular dynamics calculations on large molecules with internal degrees of freedom.

Artifacts caused by constrained dynamics

## Effect of Constraints on the Dynamics of Macromolecules[†]

Wilfred F. van Gunsteren[‡] and Martin Karplus*

*Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138.
Received December 16, 1981*

ABSTRACT: The effects of constraints on the dynamics of macromolecules is investigated by the method of computer simulation. Three molecular dynamics simulations of a protein, bovine pancreatic trypsin inhibitor (BPTI), consisting of 454 heavy atoms, are compared; one is made without applying any constraints, one with bond-length constraints, and one with bond-length and bond-angle constraints. Equilibrium averages and time-dependent molecular properties are examined to determine the effects on the dynamics of the macromolecule of constraining the bond-length and bond-angle degrees of freedom. The use of fixed covalent bond lengths does not significantly alter the dynamical properties of the protein on a time scale longer than 0.05 ps. This makes it possible to obtain a threefold increase of the computational efficiency of macromolecular simulations by the application of bond-length constraints. Constraint of the bond angles has an important effect on the dynamics. The magnitude of the fluctuations (root-mean-square Cartesian and internal coordinate fluctuations) is decreased by a factor of 2 and the dihedral transition rates are dramatically reduced. This makes clear that in a closely packed system, such as a protein in its native configuration, the excluded volume effects due to repulsive van der Waals interactions introduce a strong coupling between the dihedral angle and bond-angle degrees of freedom. A detailed model of the equilibrium or dynamic properties of such systems must therefore take account of the role of bond-angle fluctuations.
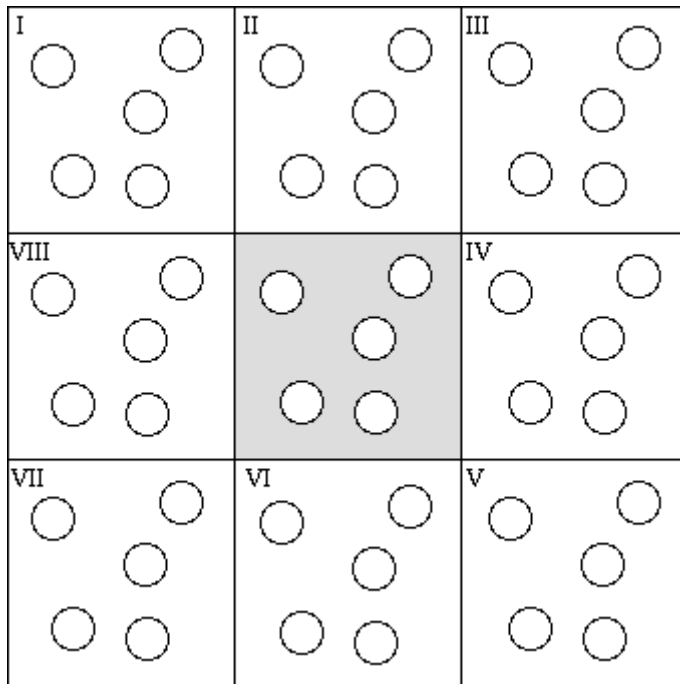
Treating the short-ranged interactions
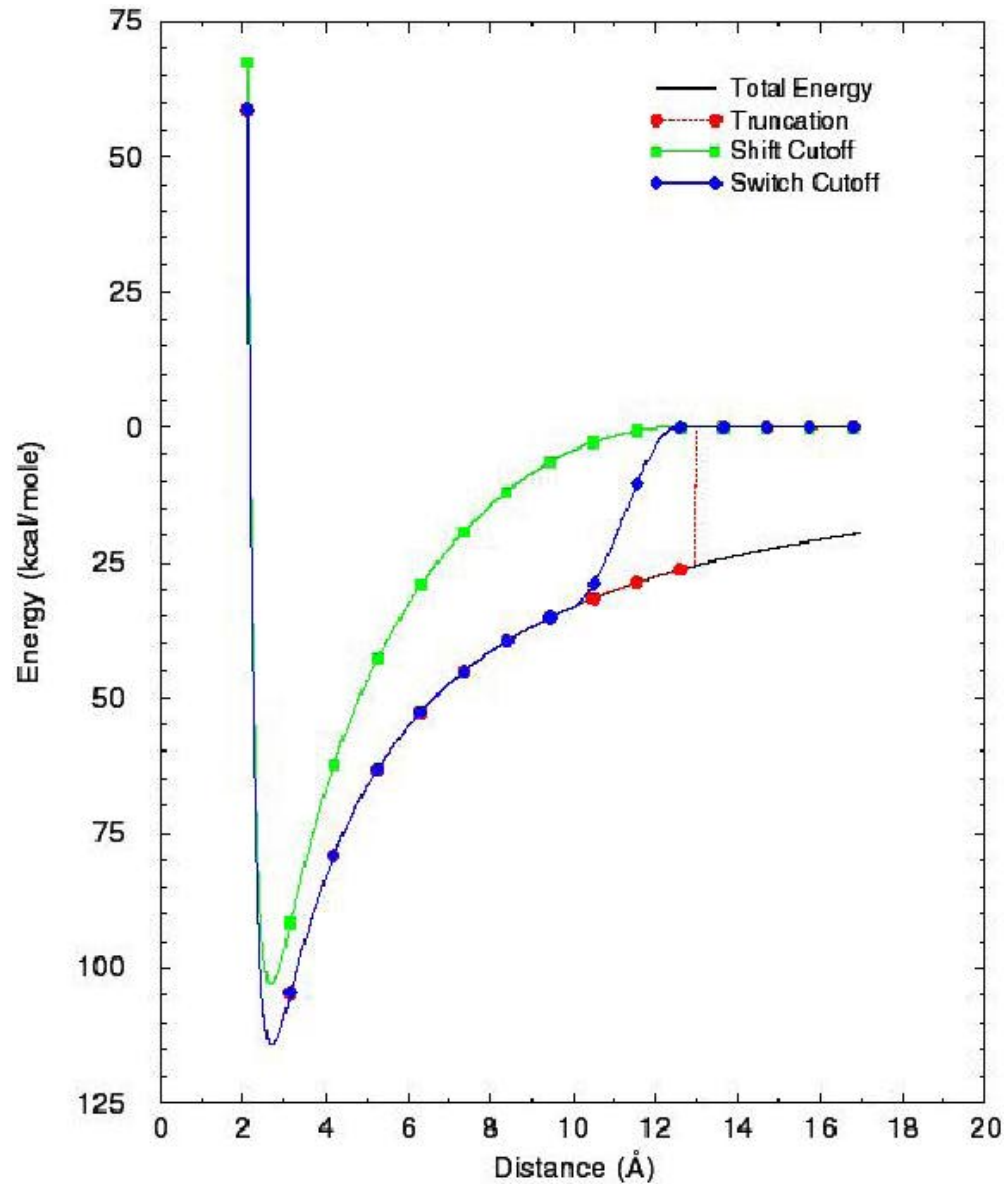
$$U(r) = \begin{cases} U(r), & r \leq r_c \\ 0, & r > r_c \end{cases}$$

How to treat the long-ranged interaction properly with the periodic boundary condition?

$$U_{elec} = \frac{1}{2} \lim_{L \to \infty} \sum_{\mathbf{L}}' \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{i,j,\mathbf{L}}}$$

where prime means that for $\mathbf{L}=0$, $i \neq j$

$$\mathbf{r}_{i,j,\mathbf{L}} = \left| \mathbf{r}_i - \mathbf{r}_j - \mathbf{L} \right|$$

$$\mathbf{L} = \left(n_x L_x, \quad n_y L_y, \quad n_z L_z \right)$$

Conditional convergence

Because the charges could be positive and negative, the series for calculating the electrostatic potential with period boundary condition converges not only very slowly (due to the r^{-1} nature), but also conditionally, which means that incorrect rearrangements of the terms will lead to wrong results. For example:

$$\ln 2 = \frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots$$

$$\frac{1}{2}\ln 2 = \quad\quad \frac{1}{2} \quad\quad - \frac{1}{4} \quad\quad + \frac{1}{6} \quad\quad - \frac{1}{8} + \cdots$$

$$\frac{3}{2}\ln 2 = \frac{1}{1} \quad\quad + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \quad\quad \frac{1}{7} - \frac{1}{4} + \cdots$$

One can see very clearly that the series of 3/2 ln2 is just the reordering of that of ln2!

8. *Die Berechnung optischer und elektrostatischer Gitterpotentiale;*
von P. P. Ewald.

Inhalt: I. 1. Elektrostatische Potentiale. 2. Elektrodynamische Potentiale. 3. Ziel der Arbeit. — II. 1. Thetafunktionen einer Veränderlichen. Zusammenhang zwischen Thetafunktionen und Gitterproblemen. 2. Ableitung der Transformationsformel für Thetafunktionen von 3 Veränderlichen. — III. 1. Potentiale als Summen von Einzelwirkungen (randloser Kristall). 2. Potentialumformung, Trennungsstelle E. Gesamtpotential eines einfachen Gitters. 3. Erregendes Potential eines einfachen Gitters. 4. Gesamtes und erregendes Potential im zusammengesetzten Gitter, Strukturfaktor. — IV. 1. Übergang zu elektrostatischen Potentialen, dimensionslose Größen. 2. Beispiel: Gitterenergie von Steinsalz. 3. Beispiel: Gitterenergie von Flußspat.

I.

1. Bei den Untersuchungen über den Aufbau der Kristalle und ihre Eigenschaften tritt die Notwendigkeit auf, gewisse Potentiale nicht nur als allgemeinen Ausdruck zu kennen, sondern ihren Zahlenwert an irgendeiner Stelle des vom Gitter erfüllten Raumes zu ermitteln. *Das elektrostatische Potential* eines Ionengitters z. B. ist

$$(1) \qquad \varphi(P) = \sum' \frac{\varepsilon_{p'}}{R_{PP'}} ,$$

wo unter P' ein Atom des Gitterverbandes, unter $\varepsilon_{p'}$ seine Ladung, und unter $R_{PP'}$ sein Abstand vom Aufpunkt P verstanden ist und das Summenzeichen sich auf alle Atome P' bezieht. Dies Potential gibt die Arbeit an, die notwendig ist, um eine positive Einheitsladung (wir benutzen gewöhnliche elektrostatische Einheiten) aus dem Unendlichen an den Ort P zu bringen. Wünscht man die Energie zu kennen, die im ganzen Gitterverband aufgespeichert ist, so ist diese[1])

---

1) Vgl. etwa M. Born und A. Landé, Berl. Ber. 45. S. 1048 bis 1068. 1918.

Paul Peter Ewald

Born: 23-Jan-1888, Berlin, Germany;
Died: 22-Aug-1985, Ithaca, NY

Annalen der Physik 369: 253-287 (1921)

Times cited: 5667 (2018/4/25)

The Spirit of the Ewald Sum

Ewald proposed in 1921 to introduce a convergent function, $\varphi(r)$, and split the sum into two parts:

$$U = \frac{1}{2}\lim_{L\to\infty}\underbrace{{\sum_{L}}'\sum_{i=1}^{N}\sum_{j=1}^{N}q_iq_j\frac{\varphi(r)}{r_{i,j,L}}}_{U_{short}} + \frac{1}{2}\lim_{l\to\infty}\underbrace{{\sum_{L}}'\sum_{i=1}^{N}\sum_{j=1}^{N}q_iq_j\frac{1-\varphi(r)}{r_{i,j,L}}}_{U_{long}}$$

The choice of $\varphi(r)$ is then crucial to make the summation efficient. The first term in the above equation should be convergent even if only a few neighbors of a particle are taken in account, and the second term should be convergent in similar way but in the Fourier (reciprocal) space.

1. *$\varphi(r)$* tends to zero rapidly as r becomes large.
2. The Fourier transform of $(1- \varphi(r))/r_{i,j,L}$ should be computable.
3. of $(1- \varphi(r))/r_{i,j,L}$ should be finite as $r \to \infty$.

The Formula of the Ewald Sum

$$4\pi\varepsilon_0 U_{elec} = \frac{1}{2}\lim_{L\to\infty}\sum_{\mathbf{L}}{}'\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{r_{i,j,\mathbf{L}}}$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\mathbf{L}}q_i q_j \frac{erfc\left(\alpha r_{i,j,\mathbf{L}}\right)}{r_{i,j,\mathbf{L}}}$$

$$+ \frac{1}{2\pi V}\sum_{\mathbf{k}\neq 0}\frac{e^{-\left(\pi k/\alpha^2\right)}}{k^2}\left|S(\mathbf{k})\right|^2$$

$$- \frac{\alpha}{\sqrt{\pi}}\sum_{i=1}^{N}q_i^2 + \frac{2\pi}{3V}\left|\mathbf{M}\right|^2$$

erfc(x): complementary error function

$$S(\mathbf{k}) = \sum_{i=1}^{N}q_i e^{-2\pi\mathbf{k}\Box\mathbf{r}_i}$$

V: unit cell volume

$$\mathbf{M} = \sum_{i=1}^{N}q_i \mathbf{r}_i$$

Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems

Tom Darden, Darrin York, and Lee Pedersen
National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

An $N \cdot \log(N)$ method for evaluating electrostatic energies and forces of large periodic systems is presented. The method is based on interpolation of the reciprocal space Ewald sums and evaluation of the resulting convolutions using fast Fourier transforms. Timings and accuracies are presented for three large crystalline ionic systems.

A smooth particle mesh Ewald method

Ulrich Essmann, Lalith Perera, and Max L. Berkowitz
Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599

Tom Darden, Hsing Lee, and Lee G. Pedersen
National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

The previously developed particle mesh Ewald method is reformulated in terms of efficient B-spline interpolation of the structure factors. This reformulation allows a natural extension of the method to potentials of the form $1/r^p$ with $p \geqslant 1$. Furthermore, efficient calculation of the virial tensor follows. Use of B-splines in place of Lagrange interpolation leads to analytic gradients as well as a significant improvement in the accuracy. We demonstrate that arbitrary accuracy can be achieved, independent of system size N, at a cost that scales as $N \log(N)$. For biomolecular systems with many thousands of atoms this method permits the use of Ewald summation at a computational cost comparable to that of a simple truncation method of 10 Å or less. © 1995 American Institute of Physics.

Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins

T. E. Cheatham, III,[†] J. L. Miller,[†] T. Fox,[†]
T. A. Darden,[‡] and P. A. Kollman*,[†]

Department of Pharmaceutical Chemistry
University of California
San Francisco, California 94143-0446
National Institute of Environmental Health Sciences
Research Triangle Park, North Carolina 27709

This communication presents results from molecular dynamics (MD) simulations with AMBER 4.1[1] and the Cornell *et al.*[2] force field of three different, fully solvated, fully charged, macromolecular structures: X-ray-derived structures of d(C-CAACGTTGG)$_2$ DNA[3] and ubiquitin[4] and an NMR-derived r(UUCG) RNA hairpin loop and stem structure.[5] We compare the use of the particle mesh Ewald (PME) method[6] for the treatment of long-range electrostatic interactions to standard charge group based truncation cutoff (CUT) methods used in simulations with periodic boundary conditions.

An accurate representation of long-range electrostatic interactions in MD simulations is extremely important in order to properly represent the structure, dynamics, and energetics of biomolecular systems.[7-9] This is particularly true for highly charged systems, such as DNA and RNA, where it has been difficult to obtain stable trajectories for a fully solvated system without imposing added restraints or artificially modifying the charges on the phosphate.[10] Stable trajectories, ideally without

knowledge, are the first demonstration of stable solvated RNA MD trajectories.

A number of studies to date have applied MD simulations to study macromolecular crystals.[12,13] In these simulations, and more generally, with standard solvated periodic boundary systems (e.g., not true experimentally derived crystal unit cells), it has been found that it is critical to properly treat the electrostatics, preferably through the use of methods which account for the periodicity of the unit cell, such as PME, but minimally through the use of large cutoffs and/or methods which smooth out the potential and forces resulting from a truncated cutoff.[8] To our knowledge, the Ewald[14] method has not been directly applied in standard solvated periodic boundary simulations of bimolecular systems, except for in the work of Smith *et al.*[15] studying a small zwitterionic peptide and Schreiber *et al.*[8] studying helix-forming peptides. Both found that better treatment of the electrostatics was obtained using the Ewald method, in contrast to standard cutoff or switching function techniques, even though the Ewald method is known to introduce long-range correlation of fluctuations.[16]

In each of the macromolecular systems investigated, MD simulations were run[17] using the PME method with tin foil boundary conditions, a charge group based truncation cutoff (CUT), and a group based truncation cutoff coupled with complete evaluation of all the solute—solute interactions (CUTSS). Over the course of the dynamics in each of these cases, the PME structures remained strictly closer to the experimentally observed structures (Figure 1), yet demonstrated significant positional fluctuations.
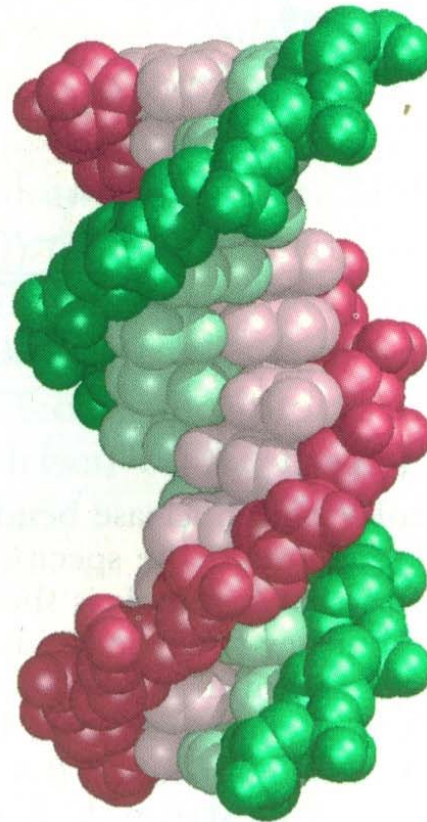
In the case of a 1 ns PME DNA simulation, the structure remained in a strictly B-DNA form with an average rms deviation 3.2 Å away from the crystal structure (2.9 Å away from canonical B-DNA[18]). This simulation, run with a fully charged DNA and explicit counterions, converged after roughly
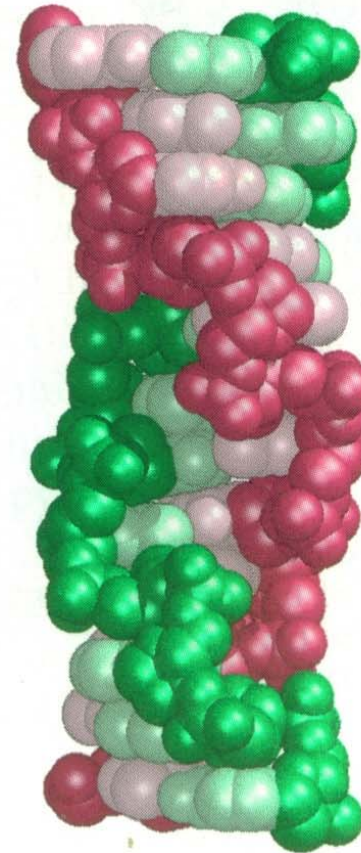
A-DNA, B-DNA and Z-DNA
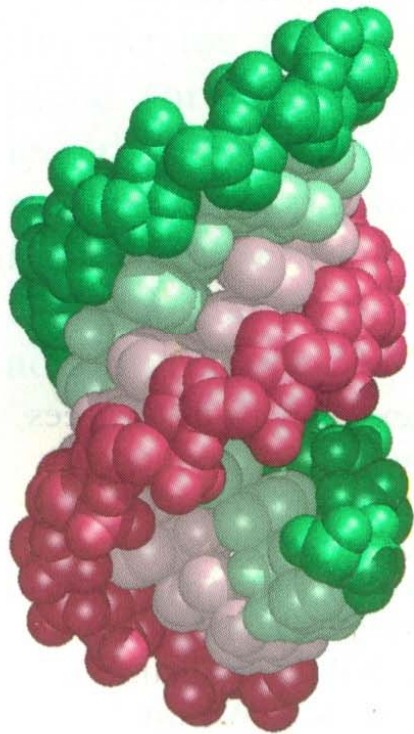
A-DNA   B-DNA   Z-DNA

Comparison of A-, B-, and Z-DNA

Table 31-1
Comparison of A-, B-, and Z-DNA

	Helix type		
	A	*B*	*Z*
Shape	Broadest	Intermediate	Narrowest
Rise per base pair	2.3 Å	3.4 Å	3.8 Å
Helix diameter	25.5 Å	23.7 Å	18.4 Å
Screw sense	Right-handed	Right-handed	Left-handed
Glycosidic bond	*anti*	*anti*	*anti* for C, T *syn* for G
Base pairs per turn of helix	11	10.4	12
Pitch per turn of helix	25.3 Å	35.4 Å	45.6 Å
Tilt of base pairs from normal to helix axis	19°	1°	9°
Major groove	Narrow and very deep	Wide and quite deep	Flat
Minor groove	Very broad and shallow	Narrow and quite deep	Very narrow and deep

JMB

Observation of the *A*-DNA to *B*-DNA Transition During Unrestrained Molecular Dynamics in Aqueous Solution

T. E. Cheatham, III and P. A. Kollman*

*Department of Pharmaceutical Chemistry, University of California, San Francisco CA 94143-0446, USA*

A large challenge in molecular dynamics (MD) simulations of proteins and nucleic acids is to find the correct "experimental" geometry when a simulation is started a significant distance away from it. In this study, we have carried out four unrestrained ~ 1 ns length MD trajectories in aqueous solution on the DNA duplex d(CCAACGTTGG)$_2$, two beginning in a canonical *A*-DNA structure and two beginning in a canonical *B*-DNA structure. As judged by root-mean-squared coordinate deviations, average structures computed from all four of the trajectories converge to within ~ 0.8 to 1.6 Å (all atoms) of each other, which is 1.3 to 1.7 Å (all atoms of the central six residues from each strand) and 3.1 to 3.6 Å (all atoms) away from the *B*-DNA-like X-ray structure reported for this sequence. To our knowledge, this is the first example of multiple nanosecond molecular dynamics trajectories with full representation of DNA charges, solvent and long range electrostatics that demonstrate both internal consistency (two different starting structures and four different trajectories lead to a consistent average structure) and considerable agreement with the X-ray crystal structure of this sequence and NMR data on duplex DNA in aqueous solution. This internal consistency of structure for a given sequence suggests that one can now begin to realistically examine sequence-dependent structural effects in DNA duplexes using molecular dynamics.
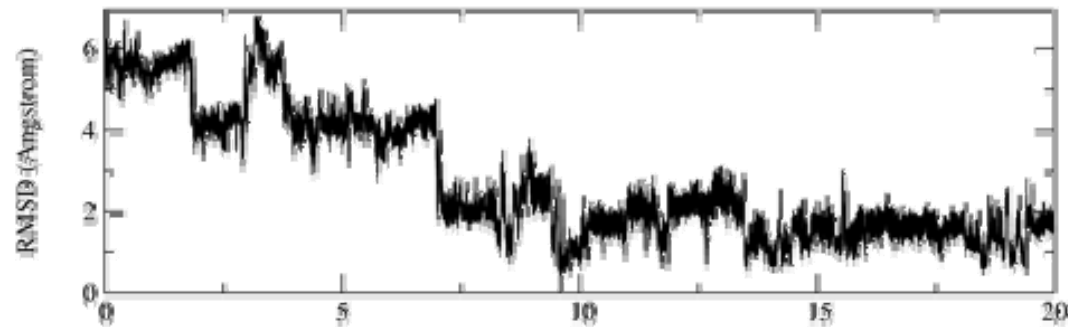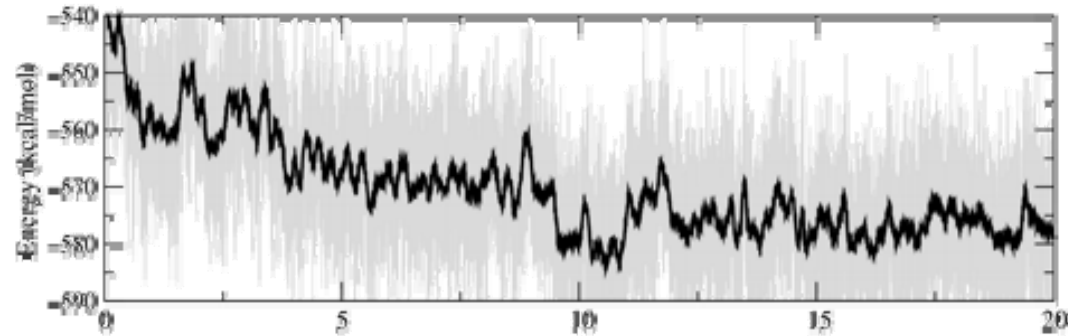
Folding Simulation

Simmerling et al. J. Am. Chem. Soc. **124:** 11258 (2002)

Twenty amino acid Trp cage

- Optimized by Andersen's group in U. Wash. Smallest with 2-state folding
- Sequence: N^{20}LYIQWLKDGGPSSGRPPPS^{39}
- Structure has been determined by NMR
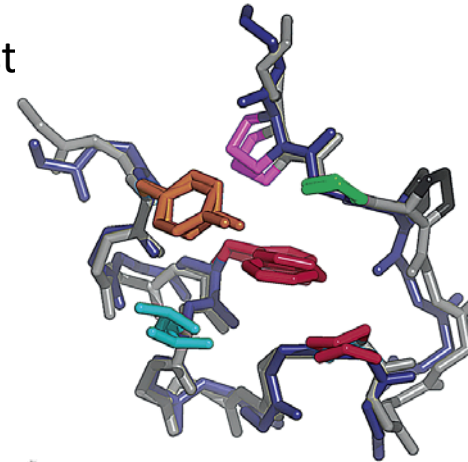
Atomic-Level Characterization of the Structural Dynamics of Proteins

David E. Shaw,[1,2]* Paul Maragakis,[1]† Kresten Lindorff-Larsen,[1]† Stefano Piana,[1]†
Ron O. Dror,[1] Michael P. Eastwood,[1] Joseph A. Bank,[1] John M. Jumper,[1] John K. Salmon,[1]
Yibing Shan,[1] Willy Wriggers[1]

Molecular dynamics (MD) simulations are widely used to study protein motions at an atomic level of detail, but they have been limited to time scales shorter than those of many biologically critical conformational changes. We examined two fundamental processes in protein dynamics—protein folding and conformational change within the folded state—by means of extremely long all-atom MD simulations conducted on a special-purpose machine. Equilibrium simulations of a WW protein domain captured multiple folding and unfolding events that consistently follow a well-defined folding pathway; separate simulations of the protein's constituent substructures shed light on possible determinants of this pathway. A 1-millisecond simulation of the folded protein BPTI reveals a small number of structurally distinct conformational states whose reversible interconversion is slower than local relaxations within those states by a factor of more than 1000.

Many biological processes involve functionally important changes in the three-dimensional structures of proteins. Conformational changes associated with protein folding (*1*), signal transduction (*2*), the catalytic cycles of enzymes (*3*), and the operation of molecular machines and motor proteins (*4*) often involve transitions among two or more structurally distinct states. These states are often characterized as "basins" separated by barriers on an "energy landscape" (*5*).

Substantial progress has been made, using both experimental (*1*, *6*) and computational (*7*, *8*) techniques, in characterizing conformational basins and the ways that proteins move within and among them. It has proven difficult, however, to structurally characterize sparsely populated or disordered states and to elucidate the "basin-hopping" mechanisms involved in the interconversion of various states.

All-atom molecular dynamics (MD) simulations are designed to provide a high-resolution view of the motions of biological macromolecules (*9*), producing continuous trajectories with the potential to connect static structural snapshots generated from experimental data. Computational constraints, however, have limited such

[1]D. E. Shaw Research, 120 West 45th Street, New York, NY 10036, USA. [2]Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA.

*To whom correspondence should be addressed. E-mail: David.Shaw@DEShawResearch.com
†These authors contributed equally to this work.
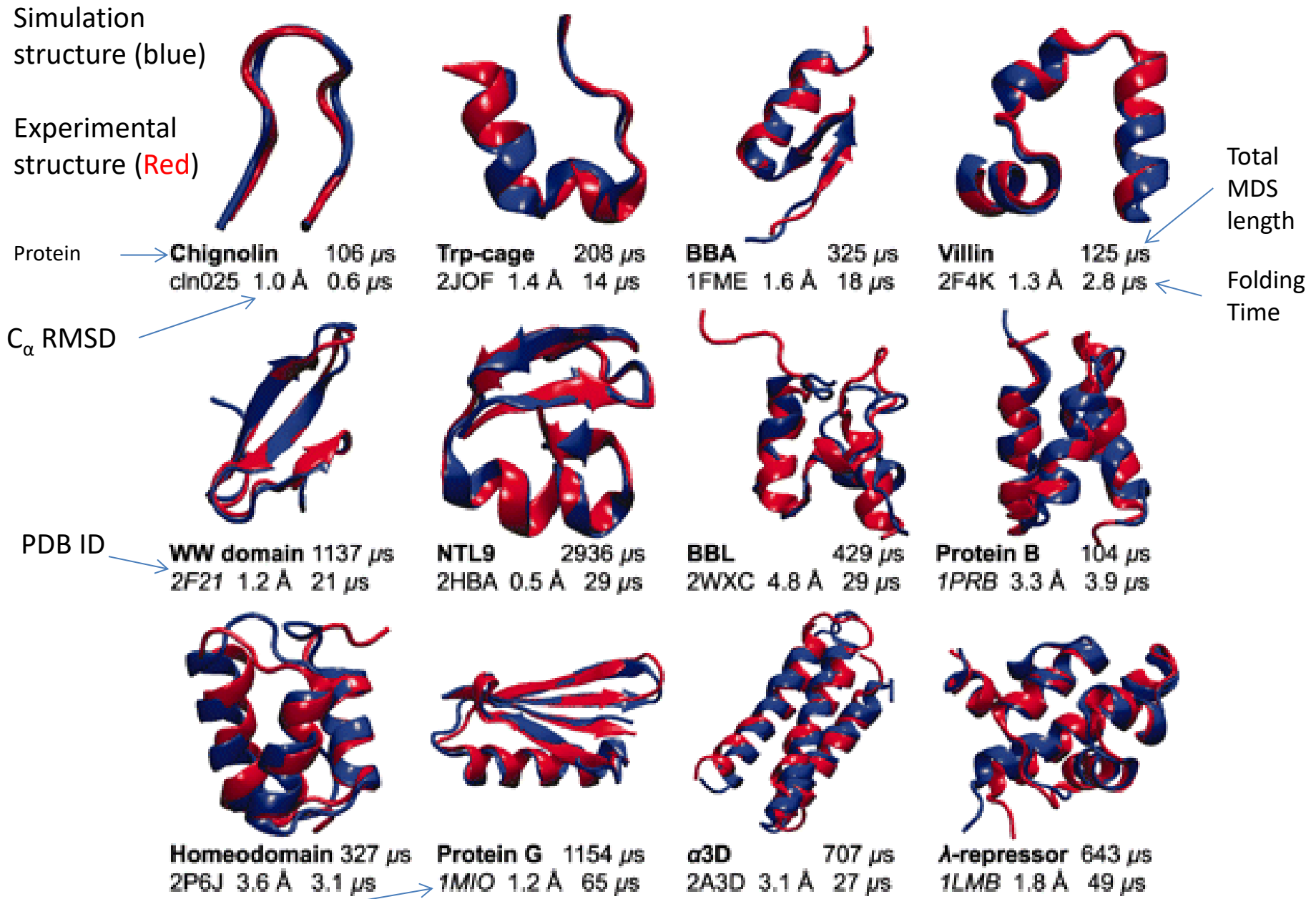
How Fast-Folding Proteins Fold

Kresten Lindorff-Larsen,[1]*† Stefano Piana,[1]*† Ron O. Dror,[1] David E. Shaw[1,2]†

An outstanding challenge in the field of molecular biology has been to understand the process by which proteins fold into their characteristic three-dimensional structures. Here, we report the results of atomic-level molecular dynamics simulations, over periods ranging between 100 μs and 1 ms, that reveal a set of common principles underlying the folding of 12 structurally diverse proteins. In simulations conducted with a single physics-based energy function, the proteins, representing all three major structural classes, spontaneously and repeatedly fold to their experimentally determined native structures. Early in the folding process, the protein backbone adopts a nativelike topology while certain secondary structure elements and a small number of nonlocal contacts form. In most cases, folding follows a single dominant route in which elements of the native structure appear in an order highly correlated with their propensity to form in the unfolded state.

Protein Folding Problem

Levinthal's paradox: consider a small protein with 100 residues. Cyrus Levinthal calculated that if each residue can assume three different positions, the total number of structures is 3^{100}, which is equal to 5×10^{47}. If it takes 10^{-13} s to convert one structure into another, the total time would be $5 \times 10^{47} \times 10^{-13}$ s, which is equal to 5×10^{34} s, or 1.6×10^{27} years! Clearly it would take much too long for even a small protein to fold properly by randomly trying out *all* possible conformations.

Simulation structure (blue)

Experimental structure (Red)

Protein → **Chignolin** 106 μs
cln025 1.0 Å 0.6 μs

C_α RMSD

Trp-cage 208 μs
2JOF 1.4 Å 14 μs

BBA 325 μs
1FME 1.6 Å 18 μs

Villin 125 μs
2F4K 1.3 Å 2.8 μs

Total MDS length

Folding Time

PDB ID → **WW domain** 1137 μs
2F21 1.2 Å 21 μs

NTL9 2936 μs
2HBA 0.5 Å 29 μs

BBL 429 μs
2WXC 4.8 Å 29 μs

Protein B 104 μs
1PRB 3.3 Å 3.9 μs

Homeodomain 327 μs
2P6J 3.6 Å 3.1 μs

Protein G 1154 μs
1MIO 1.2 Å 65 μs

α3D 707 μs
2A3D 3.1 Å 27 μs

λ-repressor 643 μs
1LMB 1.8 Å 49 μs

PDB entries in *italics*: The structure has not been determined for the simulated sequence, instead, compared it with the structure of the closest homolog in the PDB.

86

A protein folding simulation

6 µs
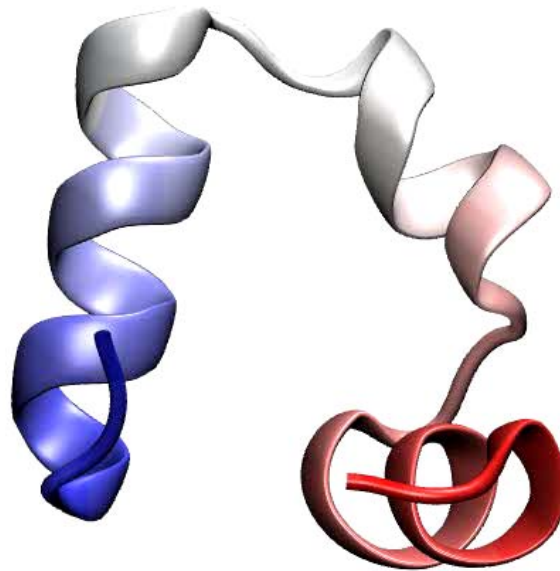
Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent

Hai Nguyen,[†,‡,#] James Maier,[‡,§,#] He Huang,[†,‡] Victoria Perrone,[†] and Carlos Simmerling*[,†,‡,§]

[†]Department of Chemistry, [‡]Laufer Center for Physical and Quantitative Biology and [§]Graduate Program in Biochemistry and Structural Biology, Stony Brook University, Stony Brook, New York 11794-5252, United States

ABSTRACT: The millisecond time scale needed for molecular dynamics simulations to approach the quantitative study of protein folding is not yet routine. One approach to extend the simulation time scale is to perform long simulations on specialized and expensive supercomputers such as Anton. Ideally, however, folding simulations would be more economical while retaining reasonable accuracy, and provide feedback on structure, stability and function rapidly enough if partnered directly with experiment. Approaches to this problem typically involve varied compromises between accuracy, precision, and cost; the goal here is to address whether simple implicit solvent models have become sufficiently accurate for their weaknesses to be offset by their ability to rapidly provide much more precise conformational data as compared to explicit solvent. We demonstrate that our recently developed physics-based model performs well on this challenge, enabling accurate all-atom simulated folding for 16 of 17 proteins with a variety of sizes, secondary structure, and topologies. The simulations were carried out using the Amber software on inexpensive GPUs, providing $\sim 1 \ \mu s$/day per GPU, and >2.5 ms data presented here. We also show that native conformations are preferred over misfolded structures for 14 of the 17 proteins. For the other 3, misfolded structures are thermodynamically preferred, suggesting opportunities for further improvement.

How does a ligand find its binding site?
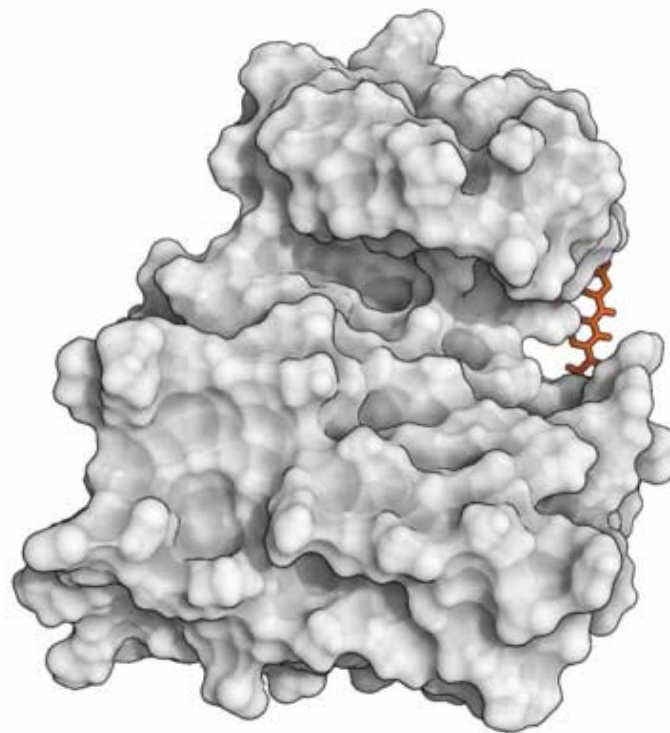
Shan *et al. J. Am. Chem. Soc.,* **133**, 9181-9183 (2011)

The Extended Relaxed Complex Scheme

- Multivalent drug design in a building-block fashion
- Computational analogue of "SAR by NMR"

1. Accommodating receptor flexibility by molecular dynamics

2. Assigning compound molecular properties using rigorous quantum chemical approaches

3. Rapid docking using efficient global search algorithms.

4. Ranking compounds by *binding free energy spectra*, instead of single binding free energy.

5. Molecular dynamics simulations of complexes with top ranked compounds.

6. Automatic free energy calculation with the double decoupling method or with the adaptive umbrella sampling scheme.

Lin *et al. J. Am. Chem. Soc.,* **124**, 5632 (2002) , Lin *et al. Biopolymers,* **68**, 47 (2003)
Amaro *et al. J. Comput-Aid Mol. Des*, **22**, 639 (2008)
Lin, *Curr. Top. Med. Chem.* **11:** 171 (2011) , Lin, *Biopolymers,* **105**, 2 (2016)

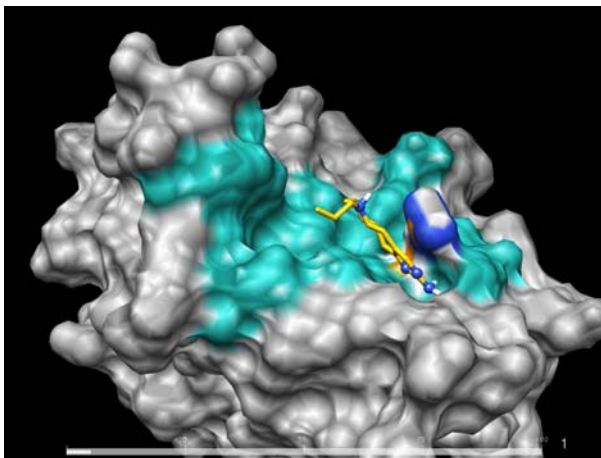# Binding free energy spectra with affinity propagation clustering
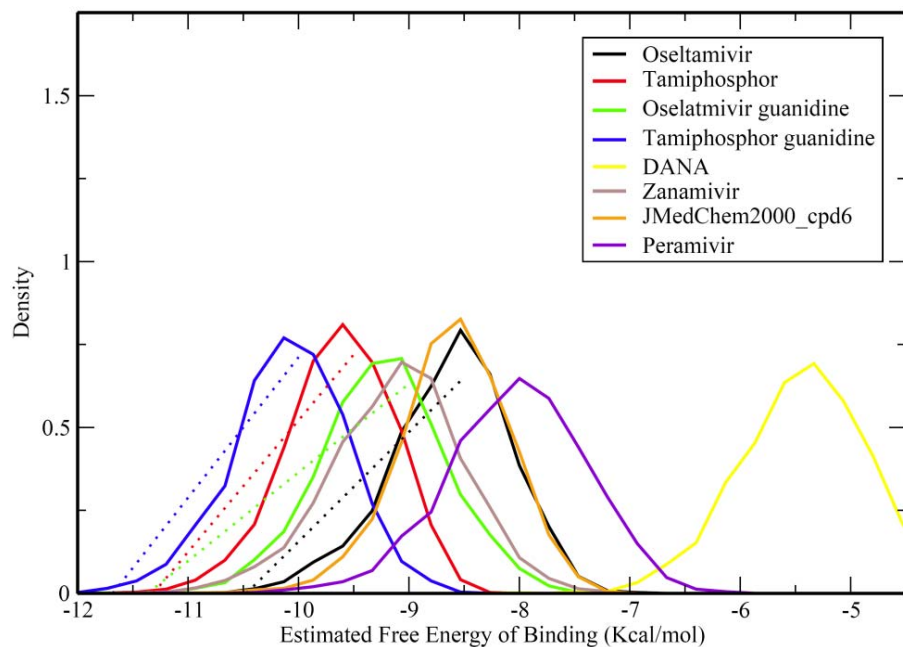
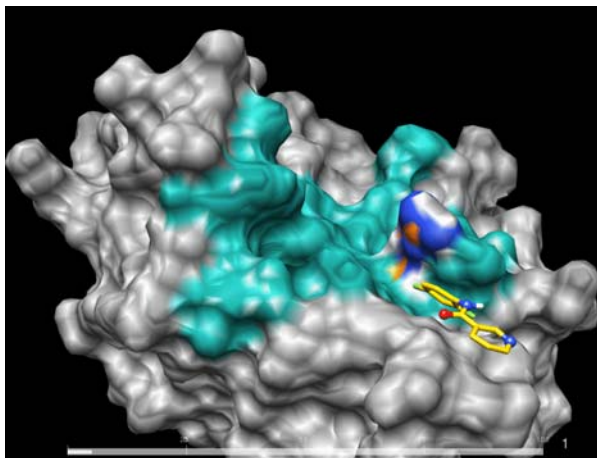RCS binding spectra from 15-ns explicit solvent MD 3000 snapshots, no clustering
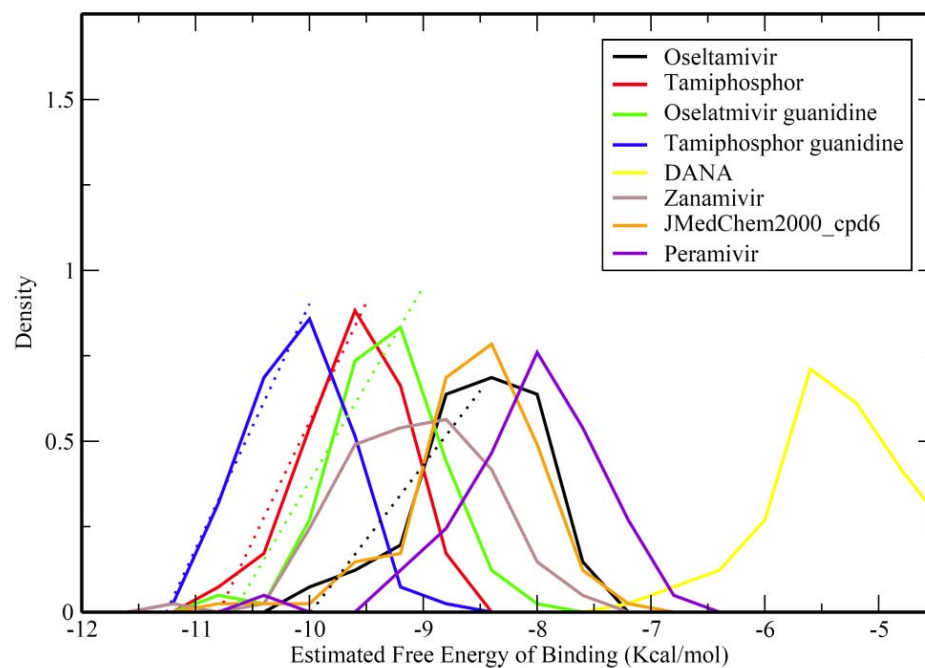
RCS binding spectra from AP-clustered 100 MD snapshots

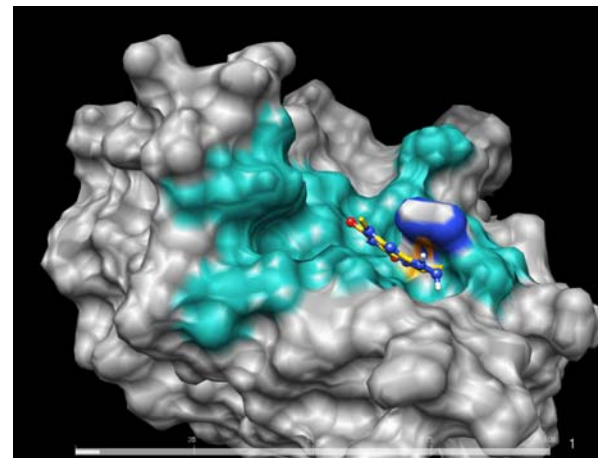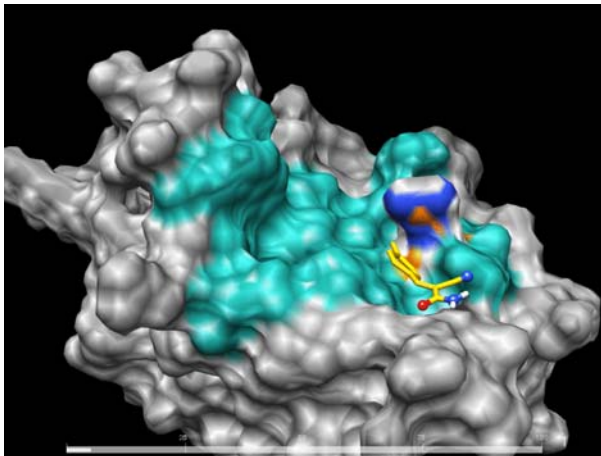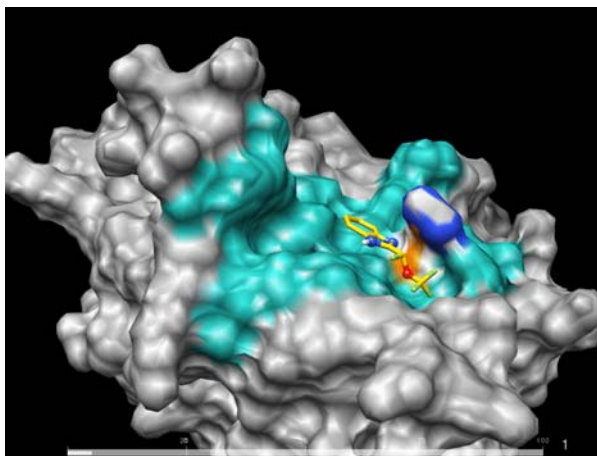Frey *et al. Science* **315**, 972,(2007)

ZINC0000XX41

ZINC0002XX54

ZINC0003XX09

ZINC0004XX25

ZINC0004XX47

ZINC0005XX58