2057-Pos Board B787
Multipoles as Force Field Parameters - Accuracy and Redundancy
Sofie Jakobsen.
Department of Chemistry, Aarhus University, Aarhus C, Denmark.
Simulation of proteins and other large biomolecules rely on a force field representation of the potential energy surface. Fixed charge force fields, like AMBER, CHARMM, OPLS, and GROMOS, are widely used, however, it is well known that they, due to their simplicity provide an inaccurate description of the electrostatic energy component.
In this study, multipoles up to quadrupoles are fitted to reproduce an ab initio electrostatic potential. The accuracy gained by introducing higher order multipoles is significant. However, the inclusion of multipoles as electrostatic parameters is highly associated with additional redundancy among the parameters. In an attempt to resolve redundancy, a large fraction of less important multipoles were identified and eliminated, without affecting the accuracy of the electrostatic potential. Furthermore, it is concluded that the reduced set of chemically important multipoles is transferable to different geometries of the same molecule. This is a promising result with respect to force field development, which highly relies on the assumption of transferability.

2058-Pos Board B788
Assessment of Nonpolar Terms in Implicit Solvent Models to Estimate Small Molecule Hydration Free Energies
Martin Brieg[1], Julia Setzler[2], Wolfgang Wenzel[2].
[1]Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Karlsruhe, Germany, [2]Institute of Nanotechnology, Karlsruhe Institute of Technology, Karlsruhe, Germany.
Solvation effects are of great importance to describe many chemical and biological processes, rendering their understanding an important goal of biophysical research. Here we investigate implicit solvent models, which are desired for applications in which an explicit solvent representation is too demanding from a computational perspective. The estimation of hydration free energies for small organic molecules presents a common test case for all solvent models. Unfortunately, a survey of common implicit solvent models showed that their estimates are not as accurate as estimates based on explicit water models, and further improvement of the nonpolar term in these models has been suggested as a possible solution for this problem.[1] We have optimized model parameters for three different nonpolar terms in combination with a generalized Born model to estimate experimental hydration free energies for a large set of small neutral organic molecules. Our results show that a nonpolar term with atom type depended surface tension coefficients delivers the most accurate estimates for a defined set of atom types, yielding a root mean square error of 0.99 kcal/mol and a squared Pearson correlation coefficient of 0.900. For explicit TIP3P water calculations based on the same molecule set, the corresponding values reported by Mobley et al. are 1.26 kcal/mol and 0.888.[2] Our study provides a thorough overview of the capabilities of these three nonpolar terms. We anticipate that the general conclusions drawn from the analysis of our results will help to improve other existing implicit solvent models.
[1] J.L. Knight, C.L. Brooks III, J. Comput. Chem., 2011.
[2] D.L. Mobley, C.I. Bayly, M.D. Cooper, M.R. Shirts, K.A. Dill, J Chem Theory Comput, 2009.

2059-Pos Board B789
Size-Modified Poisson-Boltzmann Electrostatics for Realistic Biomolecular Systems
Nuo Wang, Peter Kekenes-Huskey, Shenggao Zhou, Bo Li, J. Andrew McCammon.
UC San Diego, La Jolla, CA, USA.
Calculating the electrostatic potential (EP) around a biomolecule is essential for many types of biomolecular modeling. When a biomolecule is solvated by an ionic solution, the EP of the system is typically approximated by the solution of the Poisson-Boltzmann equation (PBE). This is a good approximation when the biomolecule is not highly charged. However, the concentrations of counterions can exceed their maximum packing densities near the highly charged regions of the biomolecule as PBE neglects the finite ion radii. A size-modified Poisson-Boltzmann equation (SMPBE) has previously been formulated to integrate ion sizes into PBE to calculate more accurate EP and ion distributions around biomolecules. Here, we extend the implementation of SMPBE to realistic biomolecular systems that contain an arbitrary number of ion species with non-uniform sizes. Specifically, we apply our method to study the Ca^{++} adsorption to the negatively charged cytoplasmic side of the sarcolemma by electrostatic forces. We use an atomic representation of the sarcolemma for the EP calculation and the solution surrounding the sarcolemma contains Ca^{++}, Na^{+}, K^{+} and Cl- ions. Our calculations indicate that, out of all the counterions, Ca^{++} is the most energetically favorable to be adsorbed to the negatively charged lipid head groups. Our results support the so-called Ca^{++} buffering effect by the sarcolemma and explain, from an electrostatics perspective, how the sarcolemma acts as one of the regulating agents of the free Ca^{++} level in the cardiac myocyte cytosol.

2060-Pos Board B790
Free-Energy Calculations for Semi-Flexible Macromolecules: Applications to DNA Knotting and Looping
Stefan M. Giovan[1], Robert G. Scharein[2], Andreas Hanke[3], Stephen D. Levene[4].
[1]Department of Molecular and Cell Biology, University of Texas at Dallas, Richardson, TX, USA, [2]Hypnagogic Software, Vancouver, BC, Canada, [3]Department of Physics and Astronomy, University of Texas at Brownsville, Brownsville, TX, USA, [4]Department of Bioengineering, University of Texas at Dallas, Richardson, TX, USA.
Obtaining accurate values of the conformational free energy of macromolecular systems is one of the most challenging problems in computational chemistry and biology. Systems involving intermediate length scales such as semi-flexible polymer models of circular DNA molecules are particularly intractable, but nonetheless important. We describe an efficient method to obtain highly accurate conformational free energies of biopolymers having arbitrary ratios of contour length L to persistence length P. Our approach is to use thermodynamic integration (TI) to apply internal constraints until the system behaves harmonically and can be analyzed using normal mode analysis (NMA). We apply this method to a discrete semi-flexible harmonic chain model for circular DNA to compute conformational free energies of prime DNA knots up to six irreducible crossings and an unknotted DNA circle containing a pair of looped domains. We discovered an unanticipated bifurcation transition in the looping free energy as a function of DNA size. This entropy-driven transition is of particular relevance for target-site selection by proteins that bind to multiple DNA sites separated by large linear distances along the genome. Such scenarios arise naturally in mechanisms of gene regulation and the action of type-II topoisomerases. Our procedure is completely general and applicable to multiscale models of any macromolecular system including proteins or other complex polymers.

2061-Pos Board B791
Free Energy Calculation of Protein Conformational Changes using Parallel Cascade Selection Molecular Dynamics Simulation and Markov State Model
Yasutaka Nishihara[1], Ryuhei Harada[2], Akio Kitao[1].
[1]University of Tokyo, Tokyo, Japan, [2]RIKEN, Hyogo, Japan.
Free energy calculation of conformational changes of macromolecules by Molecular Dynamics (MD) simulations often require long time simulations and the analysis of large amounts of simulation data. To accelerate conformational changes and enhance sampling efficiency, many methodologies have been proposed, for instance, steered MD, replica exchange, metadynamics and umbrella sampling. In these methods, optimal parameters such as biasing forces and constraints need to be adjusted for interests.
In this study, we provided an efficient free energy calculation method based on Parallel Cascade Selection MD (PaCS-MD) and Markov State Model (MSM). PaCS-MD is used to generate conformational transition pathway under the condition that a set of reactant and product structures is known a prior. In PaCS-MD, the cycle of short multiple independent MD simulations and the selection of the structures close to the product structure for the next cycle are carried out iteratively until the simulated structures move sufficiently close to the product structure. MSM is used for the studies of folding and conformational changes of macromolecules by MD simulations. In MSM, the configuration space is discretized into microstates, and a transition matrix describing the transition probabilities between microstates is calculated from the simulation data. The free energies can be calculated from the stationary probabilities, which were computed from PaCS-MD trajectories by using MSM.
MSM requires the dynamics of the system only to be Markovian, and no further assumption on the system distribution. This is a significant advantage for analyzing simulation data. To estimate our method, we calculated free energy profiles of folding of mini-proteins and large conformational changes of proteins, for example chignolin, T4-Lysozyme and actin.

2062-Pos Board B792
A Computational Method Including Protein Flexibility to Estimate Affinities with Small Ligands
Ariane Nunes-Alves, Guilherme M. Arantes.
Instituto de Química - Universidade de São Paulo, São Paulo, Brazil.
Computational methods are used to generate protein-ligand complex structures and predict their binding affinities. Usually protein flexibility is not fully

considered, and the energy functions used to estimate binding affinities are poor. This work investigated how implicit ligand and solvation theories as well as the linear response approximation may be combined to better describe the effect of protein flexibility on ligand binding affinities.
T4 lysozyme mutants, HIV-1 reverse transcriptase (HIVRT) and human FK506 binding protein 12 (FKBP) were chosen as model systems. An adaptive energy function based on the linear interaction energy approximation was parametrized and used to estimate partial affinities. Parameters were adapted according to ligand and protein surface polarities. Proteins were represented as an approximate conformational ensemble derived from molecular dynamics simulations. Interaction energies were obtained using the OPLS-AA force field with modified partial charges for ligands. A generalized Born model was used for implicit solvation.
The parametrized energy function resulted in average deviations between experimental and calculated affinities of 1.0 kcal/mol and a correlation coefficient R^2=0.8 for a test set of complexes with known binding sites. Discrimination of false-positive poses was also substantial. Then, approximations to the implicit ligand theory were proposed in order to obtain total binding affinities by combining interaction energies calculated for ligand complexes with the protein conformational ensembles. Several configurations contribute with the same weight for the FKBP protein. But, for lysozyme and HIVRT proteins, total affinities are dominated by one configuration. These results suggest that a faithful representation of protein conformational flexibility and an adequate statistical treatment based on implicit theories may be used to rapidly estimate reliable binding affinities.

2063-Pos Board B793
Development of Efficient Energy Function for Protein-Small Molecule Interactions in Medusadock
Praveen Nedumpully Govindan, Feng Ding.
Physics and Astronomy, Clemson University, Celmson, SC, USA.
MedusaDock [1, 2] is a flexible docking approach in which both the ligand and the receptor conformational flexibilities are modeled simultaneously. A predetermined discrete set of amino acid rotamers is used for representing receptor flexibility while for ligands, stochastic rotamers are generated on-the-fly during the simulation. Previous benchmark studies and CSAR 2012 blind prediction test suggested [2] that MedusaDock is able to rapidly sample the binding poses and accurately predict near-native binding scores with the scoring function, MedusaScore. However, CSAR2012 benchmark results suggested that the MedusaScore cannot satisfactorily predict the binding affinities [3]. In this work, we developed new scoring functions by introducing free energy penalties of both proteins and ligands upon binding. Specifically, we included ligand entropy loss as well as receptor energy strains induced by binding. We benchmarked the ability to reproduce experimentally determined affinities of 148 protein-ligand complexes [4]. With the inclusion of new energy terms and use of new methods, Medusa's performance was significantly improved in terms of recapitulating the binding affinities.
References
1. F. Ding, Yin, S., and Dokholyan, N.V. ''Rapid flexible docking using a stochastic rotamer library of ligands'', Journal of Chemical Information and Modeling, 50:1623-32 (2010).
2. F. Ding and N.V. Dokholyan, ''Incorporating backbone flexibility in MedusaDock improves ligand binding pose prediction in the CSAR2011 docking benchmark'', J. Chem. Inf. Model., 53:1871-1879 (2013).
3. Fourche, D., Muratov, E., Ding, F., Dokholyan, N.V., Tropsha, A. ''Predicting binding affinity of CSAR ligands using both structure-based and ligand-based approaches'', Journal of Chemical Information and Modeling, 53: 1915-1922 (2013).
4. 3. 2012 dataset available on CSAR dock website, www.csardock.org.

2064-Pos Board B794
Conformational Contribution to Thermodynamics of Binding in Protein Complexes Through Microscopic Simulations
Jaydeb Chakrabarti, Amit Das, Mahua Ghosh.
SNBNCBS, Kolkata, India.
The biomacromolecular interactions are primarily governed by the conformational changes. We show that the thermodynamics of these conformational changes in biomacromolecular complexes can be extracted from the distributions of the dihedral angles of the macromolecules. These distributions are obtained from the equilibrium configurations generated via all atom molecular dynamics simulations. The conformational thermodynamics data we obtained for the system of calmodulin bound to different peptide complexes using our methodology corroborate well with the experimentally observed conformational and binding entropies. The conformational free energy changes and its contributions for different peptide binding regions of calmodulin are evaluated microscopically. We also extend the histogram based methods for calculation of conformational thermodynamics to calcium ion binding to calmodulin. This gives the microscopic information on the participation of different residues in the metal binding process.

2065-Pos Board B795
Structure-Based Predictors of Resistance to the HIV-1 Integrase Inhibitor Elvitegravir
Majid Masso, Grace Chuang, Shinar Jain, Kate Hao, Iosif I. Vaisman.
George Mason University, Manassas, VA, USA.
HIV-1 integrase (IN), an enzyme that incorporates reverse transcribed viral DNA into the human host cell genome, is a relatively new drug target. Elvitegravir (EVG) is the second IN inhibitor in its class approved for clinical use, to be used in combination with drugs targeting additional enzymes essential to the HIV-1 replication cycle. Multi-drug regimens designed to maximally block viral replication are the standard of care for treating HIV-infected patients, which minimizes the occurrence of random or drug-selected mutations. Amino acid substitutions in patient viral protein sequences, which may confer resistance to certain drugs, pose a challenge to prescribing appropriate medications. By developing a structure-based model that predicts phenotype (EVG drug susceptibility) from translated IN genotypes, clinicians can better target HIV-1 and avoid drug resistance.
A dataset of 157 mutant IN protein sequences with known susceptibility levels to EVG, each containing only amino acid substitutions relative to native IN (i.e., no indels), were obtained from the Stanford HIV-1 Drug Resistance Database. These data were used to train four classifier (decision tree, random forest, neural network, support vector machine) and two regression (reduced error pruned tree, support vector regression) models with the Weka software package. Each mutant IN sequence was characterized by a distinct feature vector of input attributes, achieved by quantifying ensuing environmental perturbations at sequence positions in the native IN structure upon mutation. Tenfold and leave-one-out cross validation performance reflected balanced accuracy as high as 87% for the classifiers and a correlation coefficient of up to 0.85 for the regression models, indicating promise for this computational mutagenesis approach as a supplementary clinical decision-making tool and as a method to efficiently predict any detrimental effects of unexplored IN mutations on EVG drug susceptibility.

2066-Pos Board B796
Elucidating Ephrin-Induced Intersecting Signaling Pathways in the Nipah Virus G Protein using Machine Learning
Mohsen Botlani, Ralph Leighty, Sameer Varma.
Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, FL, USA.
The fusion of Nipah viruses with host cells is facilitated by two of their membrane proteins, the attachment protein (G) and the fusion protein (F). G binds to specific ephrin receptors on the host membrane. Ephrin binding changes the configurational density of G that activates F, which, in turn, mediates fusion. To understand how ephrin binding causes G to activate F, we use molecular dynamics in conjunction with machine learning1 and filter out the set of residues in the G head domain whose configurational densities are shifted equivalently by different ephrins, B2, B3, and a double mutant of B2. These three ephrins all trigger viral fusion, but with different potencies. We find that these three ephrins induce statistically equivalent shifts in the configurational densities of about one-quarter of the residues in the G head domain. This surprisingly expansive communal change in G includes most of the residues that have been shown experimentally to be important to F activation. This suggests that this set of residues contain the signaling pathways that connect the G-ephrin interface to the G stalk domain that activates F. The distribution of these residues in the G head domain is consistent with two models of signal transduction: one in which the ephrin-binding signal transduces to the F-activating G stalk domain via changes in the head-stalk interface, and the other in which the signal transduces via changes in the G head domain dimer interface. In general, this study shows how machine learning can be utilized along with molecular simulations to filter out quantitatively conserved patterns in changes in protein structure and dynamics.
1) Leighty RE.; Varma S. Quantifying Changes in Intrinsic Molecular Motion Using Support Vector Machines. J Chem. Theor Comput. 2013, 9, 868-875.

2067-Pos Board B797
Design of Druglike Small Molecules with LYN-Specific Binding
D.S. Dalafave.
The College of New Jersey, Newtown, PA, USA.
Glioblastoma multiforme (GBM), a very aggressive brain tumor, has a median survival of only 14 months. LYN, an important kinase involved in regulation of