Exploring theoretically ligands and polymeric linkers to make multivalent ligands
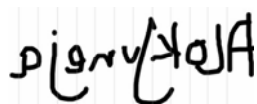
Dissertation zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt in englischer Sprache von

Alok Juneja

aus Mathura (India)

March, 2010

To my **_Teachers_**,

who made me to read and write,

and to my **_Parents_**,

who made me to understand.

In the loving memory of Barbie & Sahiba.

Die vorliegende Arbeit wurde unter Anleitung von Prof. Dr. E.W. Knapp am Institut für Chemie und Biochemie der Freien Universität Berlin im Fachbereich Biologie, Chemie und Pharmazie durchgeführt.

1. Gutachter: Prof. Dr. E. W. Knapp
                  Institute of Chemistry & Biochemistry
                  Freie Universität Berlin
                  Germany

2. Gutachter: Prof. Dr. Thomas Renger
                  Institute for Theoretical Physics
                  Johannes Kepler Universität, Linz
                  Österreich

Tag der Disputation: May 28, 2010

Acknowledgements

Preface

This cumulative thesis summarizing my research work consists of two projects. The first part covers Indirect drug design approach to align the molecules by performing flexible structure alignment without exploiting crystal structure information and its application to HIV-1 protease inhibitors, which have been used as model system in the study. The peer-reviewed published journal paper on this work would be used to focus the findings of this study. The second part addresses implicit solvent force fields optimization for ethers. For this proposition, monomer (1-2 Dimethoxy ethane) (DME) and polymer (polyethylene glycol) (PEG) have been considered as an example to conduct molecular dynamics study to validate the proposed force fields. The research paper summarizing the approach and results obtained from second project has been submitted for publication and is under review.

Statutory Declaration

I hereby testify that this thesis is the result of my own work and research, except of references given in the bibliography. This work contains material that is the copyright property of others, which cannot be reproduced without the permission of the copyright owner.

Alok Juneja

List of Publications

Publications describing parts of this thesis:

Juneja, A., Riedesel, H., Hodoscek, M. and Knapp, E. W. 2009. Bound Ligand Conformer Revealed by Flexible Structure Alignment in Absence of Crystal Structures: Indirect Drug Design Probed for HIV-1 Protease Inhibitors. *J. Chem. Theory Comput.* **5**: 659-673 (DOI**:** 10.1021/ct8004886).
(http://pubs.acs.org/doi/abs/10.1021/ct8004886)

Juneja, A., Numata, J., Nilsson, L. and Knapp, E. W. 2010. Merging implicit with explicit solvent simulations: Polyethylene glycol. *J. Chem. Theory Comput.* (DOI: 10.1021/ct100075m).
(http://pubs.acs.org/doi/abs/10.1021/ct100075m)

List of Abbreviations

2D	two-dimensional
3D	three-dimensional
ABNR	adopted basis Newton-Raphson
BLC	bound ligand conformer
CI	Carbo index
CLC	consensus ligand conformer
CONJ	conjugated gradient
DME	1,2-dimethoxy-ethane
ES	explicit solvent
FSA	flexible structure alignment
GB	generalized Born
GBSW	generalized Born with a simple SWitching
HI	Hodgkin index
HIV	human immunodeficiency virus
HIV-P	HIV-1 protease
IS	implicit solvent
LJ	Lennard-Jones
MD	molecular dynamics
PB	Poisson-Boltzmann
PEG	polyethylene glycol
PDB	protein data bank
RMSD	root mean square deviation
RSA	rigid structure alignment
SASA	solvent accessible surface area
SD	steepest decent
vdW	van der Waals

List of Figures & Tables

Part I

Part II

List of Symbols

a, b	indices for molecules
m, n	indices for atoms
R_m	atomic radius of atom m
V_m	vdW volume of atom m
q_m	atomic partial charge of atom m
$\vec{r}$	spatial position
$\vec{r}_m$	position vector of atom m
P	atom based molecular property (e.g. volume, charge, electrostatic potential)
n_a	number of atoms in molecule a
σ_m	width of Gaussian located at spatial position $\vec{r}_m$ of atom m
k	Gaussian width factor
$\rho_a(\vec{r})$	first order charge density function of molecule a at spatial position $\vec{r}$
$\rho_a^P(\vec{r})$	molecular distribution function of molecule a for property P at spatial position $\vec{r}$
$g_m(\vec{r})$	Gaussian distribution function of atom m at spatial position $\vec{r}$
$g_m^P(\vec{r})$	Gaussian distribution function of atom m for property P at spatial position $\vec{r}$
$HI(a,b)$	Hodgkin index of molecule pair a, b for all properties
$CI(a,b)$	Carbo index of molecule pair a, b for all properties
$HI^{(P)}(a,b)$	Hodgkin index of molecule pair a, b for property P
$S^{(P)}(a,b)$	overlap integral of distribution functions of molecules (a, b) for property P
$S^{(P)}(a,a)$	self-overlap of distribution function of molecule a for property P
$HI_{combined}$	overall Hodgkin index
w_{vol}, HI_{vol}	weight and Hodgkin index of volume similarity
w_{charge}, HI_{charge}	weight and Hodgkin index of charge similarity
w_{epot}, HI_{epot}	weight and Hodgkin index of electrostatic potential similarity
$E(a,b)$	effective potential energy function
$E_{intra}(a)$	intramolecular potential energy function within molecule a (both bonded and non-bonded terms)
e_S	similarity parameter
$HI_{combined}(a,b)$	overall Hodgkin index of molecule pair (a, b)
S_{mn}	overlap integral of spherical Gaussian functions of atoms m and n

This page intentionally left blank

Contents

Part I Indirect drug design: Ligand structure alignment by MD

Part II Polyethylene glycol implicit solvent MD

This page intentionally left blank

Part I

Indirect drug design:
Ligand structure alignment by MD

"Research is what I'm doing when I don't know what I'm doing"

~Wernher Von Braun

1 Introduction

Drug discovery is the core focus of research in pharmaceutical industry and several organizations are working towards developing new drug molecules. Tremendous developments in the fields of genetics, genomics, molecular and cell biology, and proteomics have helped modernize the drug discovery process, resulting in the design of many disease-specific drugs. Developing safe and effective new drugs is a time consuming and expensive process. Better target selection and the appropriate use of predictive technologies are among the approaches currently being pioneered for an improvement in the discovery of new drug candidates. The recent advent of high throughput in-silico (McInnes, C. 2007; Muegge, I. 2008; Kirchmair, J. et al. 2008) and in-vitro (Wan, H. et al. 2009) screenings have addressed some, but not all of these challenges by providing an efficient and effective way for developing safer and better drugs. Extensive technological upgradation has also shortened the time frame for innovative drug design and improved cost-effectiveness. Researchers focused on drug discovery have made remarkable progress in understanding the molecular mechanisms of disease.

The optimal, novel and rational drug design is one of the major challenges in structural and computational biology. It is not only confined to the physiological, biochemical and structural properties of the biological system where the drug is going to effect but involves also the chemistry of the ligand-receptor interaction. Most of the known theoretical approaches on drug design are based on knowledge of the structures of the biological targets and its active sites where the drug binds (Kuntz, I. D. 1992; Greer, J. et al. 1994; DesJarlais, R. L. et al. 1990; Richards, W. G. 1994). With the help of the three-dimensional structure of the target-ligand complex, it is possible to design novel therapeutic compounds. It is however essential to have knowledge about the molecular mechanism of the disease and the biological structure of the target to accomplish this task. When the target structure is unknown, purely ligand-based indirect drug design is still possible in case several ligands of chemically different architecture are known to bind to the same target (Kuntz, I. D. 1992; Greer, J. et al. 1994; DesJarlais, R. L. et al. 1990; Richards, W. G. 1994).

Apart from these techniques, combinatorial techniques, high-throughput screening and automated methods are promising to screen thousands of compounds at a given time (Kontoyianni, M. et al. 2008; McInnes, C. 2007; Reddy, A. S. et al.

2007; Villoutreix, B. O. et al. 2007; Jain, A. N. 2004; Klebe, G. 2006). Such approaches use conventional force fields from molecular dynamics (MD) simulations as well as empirical force fields derived from data bases containing structures of different drugs bound to different targets where by complex modeling procedures a drug molecule is built into the binding site of the target molecule. It involves lots of statistics, computation techniques along with the electrochemistry and the energetic discrimination of the target structure. However, in contrast to the present approach, these methods generally do not yield coordinates of the target bound ligand conformer (BLC).
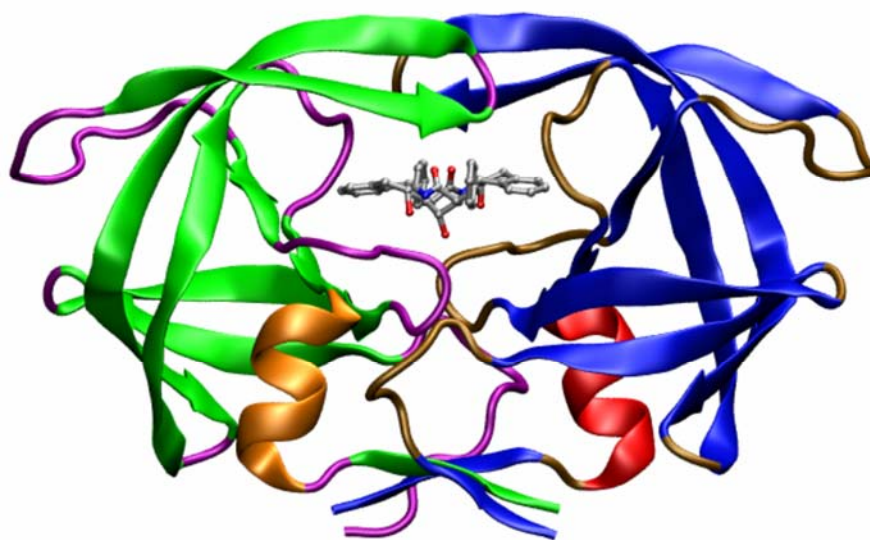
Figure 1: Crystal structure of HIV-1 protease (Bone, R. et al. 1991) with the HIV-P ligand L-700,417 (VAC) (PDB: 4PHV) (see ligand 10 in Fig. 2) as ball and stick model. The homo-dimer is represented as rubber band model possessing a C2 rotational symmetry with rotation axis oriented vertically in the drawing plane. The binding pocket extends over both monomers.

The proposed indirect drug design approach tries to circumvent the above stated hurdles. This computational approach is suitable to model drugs, if the knowledge on their binding sites is absent, assuming that drugs binding in the same pocket have common properties, which can be elucidated by appropriate similarity measures. Hence, by exploring the similarities between chemical architecture and composition of ligands that bind to the same target, one may also obtain information on the conformation of the ligand in the binding pocket. This study uses the conformational variability of a set of ligands that bind to the same target protein to

gain information on the BLC. Taking all together, this would lead to the identification and application of ideas to come out with optimal, accurate and reliable structure of the drug.

The maximum OVERLAP method (implemented in CHARMM-version 29a1) that can perform flexible structure alignment (FSA) of drug molecules on the basis of specialized energy function has been employed in the study to yield approximate BLC coordinates. HIV-1 protease (HIV-P) (Fig. 1), one of the best studied protein-ligand model system for which a large number of co-crystallized structures of ligands are available, has been considered as test case to validate the FSA approach. Here, 44 co-crystallized structures of HIV-P with chemically different inhibitors that are available in the PDB (Berman, H. M. et al. 2000) (Fig. 2) have been studied. The human immunodeficiency virus (HIV) is constituted from immature poly-proteins that contain HIV-P (Gandhi, R. et al. 1999). During viral replication, HIV-P cleaves the peptide bonds of the Gag and Pol poly-proteins yielding the appropriately sized active protein components of the mature virus. HIV-P is proteolytically active as a homo-dimer possessing C2 rotational symmetry (Fig. 1) (Toh, H. et al. 1985; Pearl, L. H. et al. 1987) with the binding cavity extending symmetrically over both subunits of the homo-dimer.

The FSA approach employs MD simulations of ligand pairs using a modified energy function. This energy function uses conventional intramolecular interactions, but instead of conventional intermolecular interactions it uses energy terms based on molecular similarity allowing molecular groups from different ligands to attract each other in proportion to their similarities and to adopt the same space. The similarity measures are based on three features: (i) volume/shape, (ii) charge and (iii) electrostatic potential.
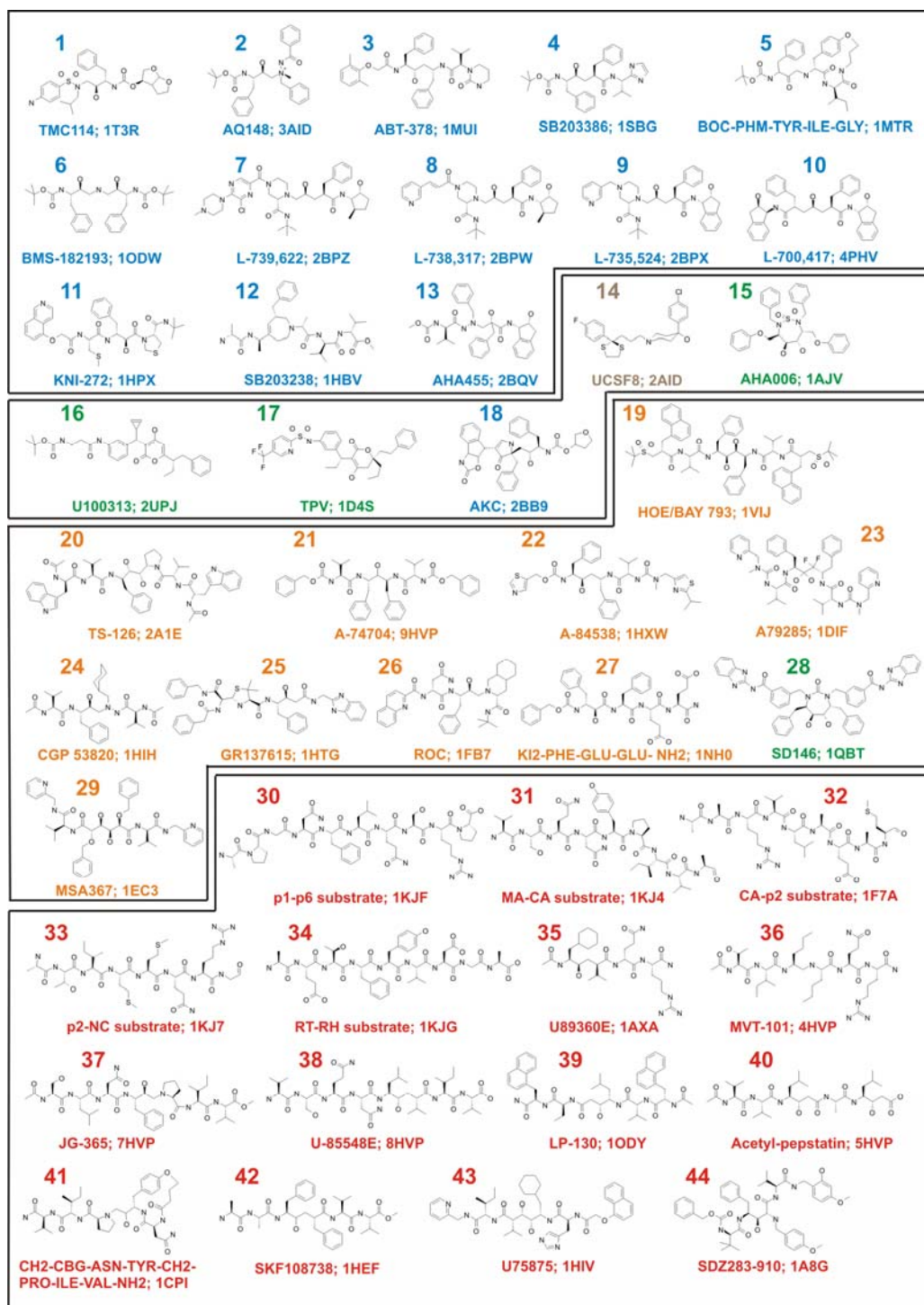
Figure 2: 2D structures of 44 HIV-P ligands for which crystal structures with HIV-P are available in PDB (Berman, H. M. et al. 2000). Trivial names, PDB code and the ligand number Lx as used in the study. The color code corresponds to the four clusters found by analysis of the H-bond pattern (see Fig. 4). The boxes enclosing the ligands correspond to the four clusters found on the basis of similarity (see Fig. 5).

FSA has been carried out between randomly generated structures of HIV-P inhibitors, to superimpose their structures optimally. To avoid any bias of crystal structure information on ligand conformers used in this study, the initial conformers

for the MD simulations are made to differ as much as possible from the conformers they adopt in the HIV-P binding pocket. The average RMSD of the initial conformers relative to the BLCs is 4.36 Å. The complicated procedure of pair-wise FSA by MD simulation is summarized schematically by an example of two artificial molecules, each possessing three atoms (Fig. 3).
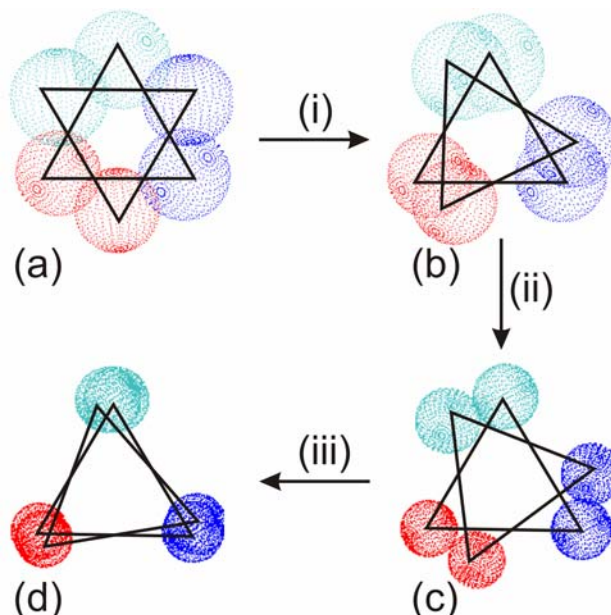
Figure 3: Schematic demonstration of the FSA procedure for two molecules that consist of three different atoms oriented in the same plane. For the sake of simplicity the two molecules to be aligned are considered to be identical, rigid, and to adopt the same conformation. The radii of the dotted spheres correspond to the widths of the Gaussians representing the atoms: (a) shows the initial orientation of the two molecules. The transition (i) from (a) to (b) corresponds to the first minimization and the MD simulation. (b) displays the conformer of the pair of molecules with maximum similarity obtained by the MD simulation. (c) displays the same molecule pair conformer as (b), but the Gaussian width factor (k) is reduced from 1.0 to 0.25. (d) shows the result after the final energy minimization with reduction of the atomic Gaussian width factor to enhance the molecular similarity.

With similarity based energy function, MD simulations of a ligand pair with subsequent energy minimization yield superimposed structures of ligand pairs in their lowest energy conformer. During this structure optimization, the inhibitors exchange information they have about the binding site and assist each other to attain the conformations that show maximum similarity amongst them. By an extensive search of structure alignments we succeeded to assign the inhibitors to 4 different clusters (Fig. 5), which is a hallmark of different binding modes in the same binding site. Our clustering approach based on the similarity measure was related to the structure data, where clustering has been done looking at hydrogen bonding pattern between inhibitor and target, where four binding modes are also observed (Fig. 4).

Figure 4: HIV-P H-bond pattern. Column 1: ligand number; Row 1: polypeptide chain A or B of HIV-P; Row 2: formal numbering of the H-bond partner in HIV-P (an H-bond is formed if the participating non-hydrogen atoms are closer than 3.5 Å); Row 3: specification of H-bond partners in format (atom_name)-(amino_acid_type)-(amino_acid_number)-(backbone(B)/side_chain(S)). Each column has a color code referring to the H-bond partner in HIV-P. Each line characterizes the H-bond pattern for a specific HIV-P ligand labeled in column 1. For the ligands (L3, L10, L14, L25, L26, L35, L40, L42) alternate conformations were found in the HIV-P crystal structures. For these ligands both H-bond patterns are given in subsequent lines. These cases are highlighted by black boxes. A filled cell denotes an H-bond of the ligand with the particular group of HIV-P; the digits (1 or 2) count the number of H-bonds formed between the ligand and this group. Empty cells denote absence of possible H-bonds. H-bond patterns persistent for a group of ligands are highlighted by ellipses. Red (blue) ellipses denote strict absence (presence) of specific H-bonds. Green ellipses indicate dominant occurrences of a particular H-bond. The five substrate HIV-P ligands (L30 – L34) belong to the red cluster. The clustering of the HIV-P ligands is based on the similarity of the H-bond pattern and indicated by the color code in column 1.
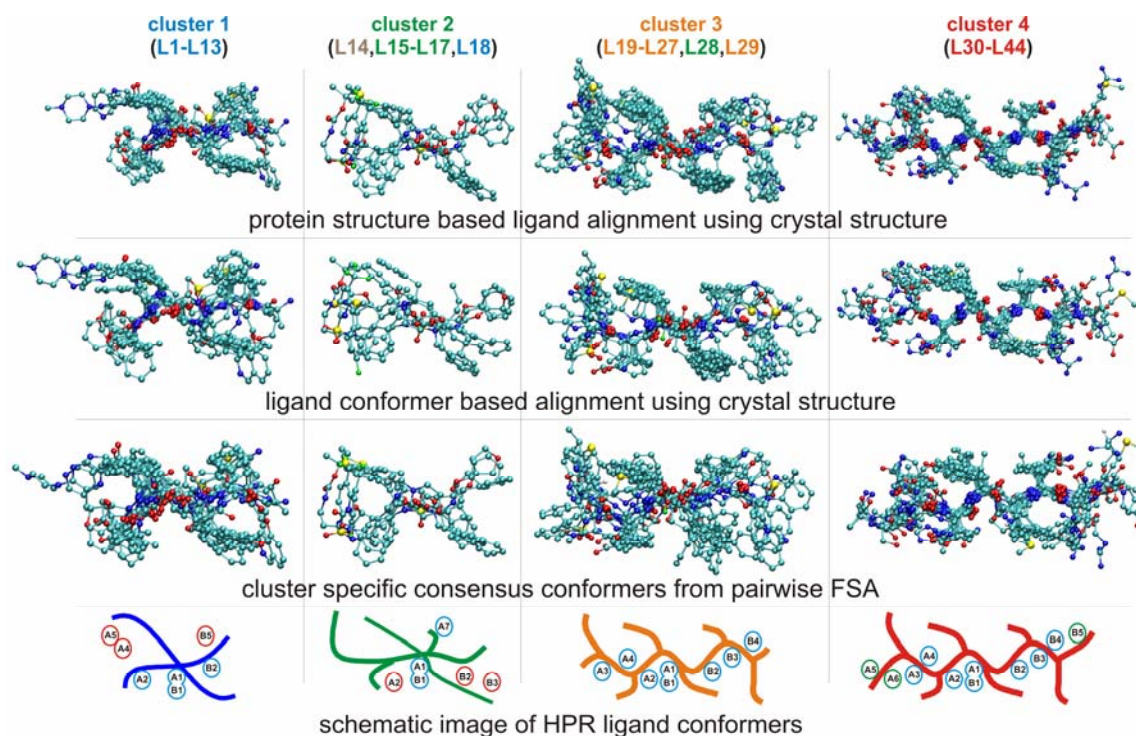
Figure 5: Structure alignments of HIV-P ligands for four clusters found from pairwise similarity based on FSA. First row: protein structure-based RSA of ligand conformers in crystal structures. Second row: ligand conformer-based RSA using ligand conformers from crystal structures. Third row: structure alignment of consensus conformers from pairwise FSA. Last row: schematic representation of ligand conformers from protein structure-based RSA of first row. Circles indicate dominant H-bond patterns found from analysis of crystal structures. Labels in circles refer to residue identified in Fig. 4. Red circles indicate strict absence; blue circles strict presence of H-bonds; green circles dominant occurrences of H-bonds (circles obey same color code as ellipses in Fig. 4).

Flexible structure alignment (FSA) leads to consensus ligand conformers (CLC), which need to be compatible with the diverse chemical architectures of the ligands found in the binding pocket. The conformations of ligands obtained after FSA are then compared with their respective crystal structures. Apparently, the drug conformations modeled by FSA are close to the conformations that the drugs acquire to fit in the binding site of the crystal structure. The ensemble of obtained CLCs from FSA can be used as integral part of the pharmacophore to design new drugs or to perform virtual screening on molecular libraries (Kontoyianni, M. et al. 2008; Mclnnes, C. 2007; Reddy, A. S. et al. 2007; Villoutreix, B. O. et al. 2007; Jain, A. N. 2004; Klebe, G. 2006).

2 Molecular similarity

Carbo proposed the concept of molecular similarity to compare charge densities $\rho_a(\vec{r})$ and $\rho_b(\vec{r})$ of molecules a and b respectively, and coined it as Carbo Index (*CI*) (Carbo, R. et al. 1980; Good, A. C. 1992), which is defined as

$$CI(a,b) = \frac{\int \rho_a(\vec{r}) \, \rho_b(\vec{r}) d\vec{r}}{\left(\int \rho_a^2(\vec{r}) d\vec{r}\right)^{1/2} \left(\int \rho_b^2(\vec{r}) d\vec{r}\right)^{1/2}} \tag{1}$$

Because of the inability of *CI* to quantify similarities in magnitude, an alternative definition of molecular similarity was proposed by Hodgkin and Richard (Good, A. C. 1992; Hodgkin, E. E. et al. 1987; Good, A. C. et al. 1998) that in addition to the shape similarity as considered in *CI*, is also sensitive to similarity in magnitude and was termed Hodgkin Index (*HI*).

$$HI(a,b) = \frac{2\int \rho_a(\vec{r}) \, \rho_b(\vec{r}) d\vec{r}}{\int \rho_a^2(\vec{r}) d\vec{r} + \int \rho_b^2(\vec{r}) d\vec{r}} \tag{2}$$

Both similarity indices vary between $+1$ and -1, extremes which correspond to identity and complementarity, respectively. A combination of shape and volume, distribution of atomic partial charges and electrostatic potential are used to characterize the structure of molecules. Atoms (m) of molecule are represented by three-dimensional Gaussian distribution functions $g_m(\vec{r})$ (Good, A. C. et al. 1992; Good, A. C. et al. 1993) to exemplify the shape and volume of a molecule:

$$g_m(\vec{r}) = \frac{\sqrt{V_m}}{\sigma_m^{3/2} \pi^{3/4}} \cdot \exp\left(\frac{-(\vec{r} - \vec{r}_m)^2}{2\sigma_m^2}\right), \tag{3}$$

where V_m is the van der Waals (vdW) volume of atom m, $\vec{r}_m$ is the position of the center of atom m, σ_m is the width of the Gaussian accounting approximately for the atomic radius $R_m \approx \sigma_m$. The volume common to two atoms m and n is given by the overlap integral:

$$S_{mn} = \int \left(g_m(\vec{r}) \ g_n(\vec{r}) \right) d\vec{r} = \sqrt{V_m \ V_n} \left(\frac{2 \ \sigma_m \ \sigma_n}{\sigma_m^2 + \sigma_n^2} \right)^{3/2} \exp\left(\frac{-\left(\vec{r}_m - \vec{r}_n\right)^2}{2\left(\sigma_m^2 + \sigma_n^2\right)} \right), \qquad (4)$$

The molecular similarity approach is not restricted to volume and shape similarity (Meyer, A. Y. et al. 1991) and can be applied to any atom based molecular property *P*, like volume/shape, charge, electrostatic potential considered in this study. A molecular distribution function, $\rho_a^{(P)}(\vec{r})$ referring to property *P* of molecule *a*, which consists of n_a atoms is represented as sum of $g_m^{(P)}(\vec{r})$ referring to the corresponding atomic property *P*:

$$\rho_a^{(P)}(\vec{r}) = \sum_{m=1}^{n_a} g_m^{(P)}(\vec{r}). \qquad (5)$$

Similarly, $\rho_b^{(P)}(\vec{r})$ would be the molecular distribution function referring to property *P* of molecule *b*. The overlap integral of the two corresponding molecular distribution functions of molecules (*a*, *b*) would provide an estimate of the similarly between them with respect to property *P*:

$$S^{(P)}(a,b) = \int \rho_a^{(P)}(\vec{r}) \ \rho_b^{(P)}(\vec{r}) d\vec{r} = \sum_m^{n_a} \sum_n^{n_b} S_{mn}^{(P)}. \qquad (6)$$

Hence, the atom-based form of *HI* that measures the similarity with respect to property *P* for a molecule pair (*a*, *b*) is given by

$$HI^{(P)}(a,b) = \frac{2 \cdot S^{(P)}(a,b)}{S^{(P)}(a,a) + S^{(P)}(b,b)} \qquad (7)$$

where $S^{(P)}(a,b)$ is the overlap integral for the molecule pair (*a*, *b*) and $S^{(P)}(a,a)$ and $S^{(P)}(b,b)$ are the self-overlaps of molecules *a* and *b*, respectively.

The $HI_{combined}$ which estimates the overall similarity for the molecule pair (a, b), considering contributions from all three properties (volume/shape, charge and electrostatic potential) is the weighted sum of the individual HI (HI_{vol}, HI_{charge}, and HI_{epot}). The resulting similarity index $HI_{combined}$ can in principle vary between -1 and $+1$.

$$HI_{combined} = \frac{w_{vol}}{w_{sum}} HI_{vol} + \frac{w_{charge}}{w_{sum}} HI_{charge} + \frac{w_{epot}}{w_{sum}} HI_{epot}, \qquad (8)$$

where $w_{sum} = w_{vol} + w_{charge} + w_{epot}$; w_{vol}, w_{charge}, w_{epot} are the relative weights for volume/shape, charge and electrostatic potential overlap, respectively. The $HI_{combined}$, value is constantly monitored as the conformers of a molecule pair are being adjusted to maximize their similarity where each of the two molecules at the same time must try to assume a reasonable low-energy structure. It is possible to achieve a compromise between these two criteria by an effective potential energy function that simultaneously accounts for both influences: intramolecular interaction and similarity.

$$E(a,b) = E_{intra}(a) + E_{intra}(b) - e_S \, HI_{combined}(a,b). \qquad (9)$$

The intramolecular interactions $E_{intra}(a)$, $E_{intra}(b)$ are the true physical ones (i.e. bonded interactions depending on bond length, bond angle, torsions; and non-bonded interactions depending on van der Waals and Coulomb terms of atom pairs within the same molecule). The true intermolecular contribution is replaced by a non-physical term proportional to $HI_{combined}(a,b)$. The parameter e_S is set so that all three terms are of comparable magnitude such that both molecules are structurally aligned in a low energy conformation with similar parts of the two molecules superimposed.

3 Publication

Article   Bound Ligand Conformer Revealed by Flexible Structure Alignment in Absence of Crystal Structures: Indirect Drug Design Probed for HIV-1 Protease Inhibitors

Authors   <u>Alok Juneja</u>, Henning Riedesel, Milan Hodoscek and E. W. Knapp

Bibliography   J. Chem. Theory Comput.*,* 2009, 5 (3), pp 659–673

DOI   10.1021/ct8004886

Title Page Figure
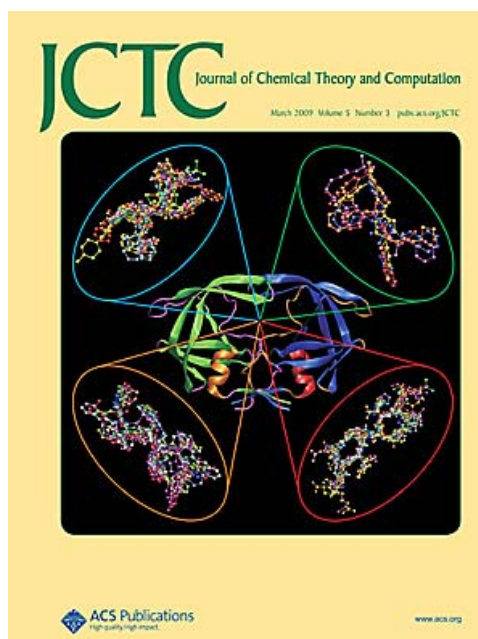
Indirect drug design is performed in the absence of structural information solely using several ligands known to bind at the same target protein. The structure of ligand pairs is aligned by MD simulation using additional similarity based energy terms. Application to 44 inhibitors of HIV-protease (center) yield consensus ligand conformers (CLCs), which based on similarity belong to four different clusters (corners). These CLCs are close to the ligand conformers in the crystal structure, and the clusters refer to different ligand binding modes.

<p align="center">The full article is attached at the end of document</p>

4 Conclusions

The present study uses MD simulations on ligand pairs from a set of ligands binding to the same target protein to perform flexible structure alignment (FSA). The FSA approach has been explored using a test set of 44 HIV-P crystal structures. A specialized energy function has been employed for this procedure, which includes conventional intramolecular interactions, while the true intermolecular interactions are replaced by attractive interactions derived from similarity measures accounting for volume/shape, atomic partial charges and electrostatic potential of the ligands. These attractive terms allow equivalent molecular groups from a pair of flexible ligands to be superimposed, yielding consensus ligand conformers (CLC) that are most similar to each other. Providing a sufficient number of ligands of different chemical architecture binding to the same target that are available and with the assumption that the CLCs contain information on the true bound ligand conformer (BLC), it should be possible to obtain approximate BLCs without using any experimental structural information of the target protein.

Four different ligand clusters based on the similarity of the HIV-P ligand conformers obtained from the pairwise FSA have been reported. Two clusters of small and two clusters of large ligands were found. The five natural substrate oligopeptides (L30 − L34), all belong to the same red cluster of large ligands. Classification of clusters based on H-bond pattern of the ligands in the HIV-P binding pocket supports the four clusters reported by similarity search measure on structure alignments of HIV-P ligands. The consensus ligand conformers (CLCs) agree well with the BLC of the crystal structures with an RMSD of about 2 Å. Hence, the present approach of indirect drug design offers a possibility to generate coordinates of the BLCs without using any structural information from the target of the ligand.

The application to 44 HIV-P ligands demonstrated that the procedure works well although the ligands possess four different binding modes. The ensemble of coordinates of CLCs belonging to the same binding mode that are generated by the present FSA approach approximate the true BLCs. For the FSA approach of indirect drug design to work, several ligands of different chemical architecture binding to the same target that possess the same binding mode are needed. The results for the test case of HIV-P inhibitors, the introduced procedure of indirect drug design looks promising, which demonstrates to find the most probable conformation of a drug in

the binding pocket without knowledge on the binding site. Thus, in combination with the chemical composition (i.e. electrostatics, H-bond pattern and hydrophobicity) of the ligands the ensemble of CLCs can be used to define pharmacophores. In a nutshell, this indirect drug design approach will not only lay foundations of alternative ways for next generation drug design but may also be helpful to optimize lead compounds in an efficient way.

5 References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235-242.

Bone, R., Vacca, J. P., Anderson, P. S. and Holloway, M. K. 1991. X-ray crystal structure of the HIV protease complex with L-700,417, an inhibitor with pseudo C2 symmetry. *J. Am. Chem. Soc.* **113**: 9382-9384.

Carbo, R., Leyda, L. and Arnau, M. 1980. How similar is a molecule to another? An electron density measure of similarity between two Molecular Structures. *Int. J. Quantum Chem.* **17**: 1185-1189.

DesJarlais, R. L., Seibel, G. L., Kuntz, I. D., Furth, P. S., Alvarez, J. C., Montellano, P. R. O. d., et al. 1990. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. USA* **87**: 6644-6648.

Gandhi, R., Bartlett, J. G. and Linkinhoker, M. 1999. Life Cycle of HIV Infection. *Johns Hopkins University Division of Infectious Diseases and AIDS Service*.http://hopkins-aids.edu/hiv_lifecycle/hivcycle_txt.html

Good, A. C. 1992. The calculation of molecular similarity: alternative formulas, data manipulation and graphical display. *J. Mol. Graphics* **10**: 144-151.

Good, A. C., Hodgkin, E. E. and Richards, W. G. 1992. Utilization of Gaussian functions for the rapid evaluation of molecular similarity. *J. Chem. Inf. Comput. Sci.* **32**: 188-191.

Good, A. C. and Richards, W. G. 1993. Rapid evaluation of shape similarity using Gaussian functions. *J. Chem. Inf. Comput. Sci.* **33**: 112-116.

Good, A. C. and Richards, W. G. 1998. Explicit Calculation of 3D Molecular Similarity. 3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity. Kubinyi, Folkers and Martin. The Netherlands, Kluwer Academic. **2:** 321-338.

Greer, J., Erickson, J. W., Baldwin, J. J. and Varney, M. D. 1994. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *J. Med. Chem.* **37**: 1035-1054.

Hodgkin, E. E. and Richards, W. G. 1987. Molecular Similarity based on electrostatic potential and electric field. *Int. J. Quantum. Chem.*: 105-110.

Jain, A. N. 2004. Virtual screening in lead discovery and optimization. *Curr. Opin. Drug. Discov. Devel.* **7**: 396-403.

Kirchmair, J., Distinto, S., Schuster, D., Spitzer, G., Langer, T. and Wolber, G. 2008. Enhancing Drug Discovery Through In Silico Screening: Strategies to Increase True Positives Retrieval Rates. *Curr. Med. Chem.* **15**: 2040-2053.

Klebe, G. 2006. Virtual ligand screening: strategies, perspectives and limitations. *Drug. Discov. Today* **11**: 580-594.

Kontoyianni, M., Madhav, P., Suchanek, E. and Seibel, W. 2008. Theoretical and Practical Considerations in Virtual Screening: A Beaten Field? *Curr. Med. Chem.* **15**: 107-116.

Kuntz, I. D. 1992. Structure-Based Strategies for Drug Design and Discovery. *Science* **257**: 1078-1082.

McInnes, C. 2007. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **11**: 494-502.

Mclnnes, C. 2007. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **11**

Meyer, A. Y. and Richards, W. G. 1991. Similarity of molecular shape. *J. Comput.-Aided Mol. Des.* **5**: 427-439.

Muegge, I. 2008. Synergies of Virtual Screening Approaches. *Mini-Rev. Med. Chem.* **8**: 927-933.

Pearl, L. H. and Taylor, W. R. 1987. A structural model for the retroviral proteases. *Nature* **329**: 351-354.

Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N. and Sastry, G. N. 2007. Virtual screening in drug discovery -- a computational perspective. *Curr. Protein Pept. Sci.* **8**: 329-351.

Richards, W. G. 1994. Computer-aided drug design. *Pure Appl. Chem.* **66**: 1589-1596.

Toh, H., Ono, M., Saigo, K. and Miyata, T. 1985. Retroviral protease-like sequence in the yeast transposon Ty 1. *Nature* **315**: 691-692.

Villoutreix, B. O., Renault, N., Lagorce, D., Sperandio, O., Montes, M. and Miteva, M. A. 2007. Free resources to assist structure-based virtual ligand screening experiments. *Curr. Protein. Pept. Sci.* **8**: 381-411.

Wan, H. and Holmen, A. G. 2009. High Throughput Screening of Physicochemical Properties and In Vitro ADME Profiling in Drug Discovery. *Comb. Chem. High Throughput Screening* **12**: 315-329.

Part II

Polyethylene glycol implicit solvent MD

"…Simulations are fiction aspiring to emulate the reality. Pretty pictures and even a few good numbers do not guarantee good science"

~Peter J. Steinbach

1 Introduction

The biological environment of most biomolecules is water. Interactions with the surrounding water are important for the stability and proper function of biomolecules. The behavior of molecules in solution depends fundamentally on the interplay of solute-solute versus solute-solvent and solvent-solvent interactions. The type of solvent used determines solvation and association behavior of molecules as well as their protonation and redox states in case of titratable or redox-active molecules and their conformations in case of flexible molecules. The native structure of biomolecules and in particular of proteins is governed by interactions with water (Gregory, R. B. 1995; Franks, F. 1972-1982). Biological macromolecules need to be under physiological conditions (i.e. to be in aqueous solution at specific temperatures, pHs and ionic strengths) to adopt their native structure, which is a prerequisite to its proper function (Eisenberg, D. et al. 1986; Teeter, M. M. 1991; Soares, C. M. et al. 2003; Privalov, P. L. et al. 1993; Honig, B. et al. 1995). Binding strengths of drugs, substrates and inhibitors to proteins are strongly influenced by water (Klebe, G. 2006; Corbeil, C. R. et al. 2009; Franks, F. 1972-1982; Froloff, N. et al. 1997). Conformational dynamics and function of proteins is also solvent controlled (Frauenfelder, H. et al. 2006; Dill, K. A. 2002; Ben-Naim, A. 1980). Also the neighborhood of a protein surface influences the behavior of the surrounding water. It restricts their conformational variability leading to reduced entropy (Vaitheeswaran, S. et al. 2004; Dill, K. A. 2002; Southall, N. T. et al. 2001) and slows down its dynamics by about two to four fold (Makarov, V. A. et al. 1998; Knapp, E. W. et al. 1993).

Modeling and simulation of structure and dynamics of molecules and their interactions in solutions generally requires the consideration of interactions with solvent molecules in atomic detail. Therefore, the effects of solvent environment must be taken into account for the realistic modeling of biomolecules in atomic detail employing an atom based molecular force field. This is why in approaches using molecular dynamics (MD) simulations the solvent environment is represented in an explicit fashion by placing sufficiently large number of individual water molecules around the solute and simulating their motion on an equal footing with the solute molecule of interest.

The explicit solvent (ES) methodology, which is most realistic theoretical approach though provides a representative picture of how biomolecules behave in a biological cell but at the same time suffers from considerable computational cost, the significant part of which comes from solvent-solvent interactions. Even in the best case, the computational costs of such molecular systems increase faster than linear with the number of atoms making MD simulations with ES models rather expensive. Molecular dynamics studies based on this methodology are well established for the study of single biomolecules over nanosecond time scales but are often prohibitively slow, especially for the large biomolecular systems and long-time scale processes such as those involved in protein folding (Karplus, M. et al. 2002).

Different procedures have been applied to reduce the CPU time required for MD simulations. One approach to accelerate simulations of biomolecules involves the application of mean-field description of solute-solvent interactions rather than explicit solvent (Roux, B. et al. 1999). This implicit solvent (IS) methodology (Warwicker, J. et al. 1982; Cramer, C. J. et al. 1999; Roux, B. et al. 1999; Gilson, M. K. et al. 1993; Baker, N. A. 2005; Lopes, A. et al. 2007; Ferrara, P. et al. 2002; Chen, J. et al. 2008; Vorobjev, Y. N. et al. 1998; Honig, B. et al. 1995; Gilson, M. K. 1995; Scarsi, M. et al. 1997) which diminishes the number of atoms to be considered enormously, represents the aqueous solvent environment by high dielectric continuum medium with electrostatic, entropic (hydrophobic) and viscous properties that match water. This approximation eliminates a number of solvent effects and also does not represent the effects of tightly bound water molecules, which might be important for the stability and function of the structure of interest, thus introduce serious artifacts.

Implicit solvent simulations represent a compromise between computational efficient and the level of realism that can be achieved with a mean-field solvent representation. An additional efficiency bonus of IS models is the absence of solvent viscosity. As a consequence the actually used elementary time step of MD simulations with an IS model corresponds in reality to a larger time interval. Hence, it allows exploring the conformational space of a solute molecule in shorter simulation times, but bears the disadvantage that the dynamics is unrealistically fast and is not reliable. With this advantage of exploring large conformational transitions in IS models, several all-atom MD simulations of *ab initio* folding of small proteins are reported to gain better insight into the protein folding mechanism (Simmerling, C. et al. 2002;

Pitera, J. W. et al. 2003; Zagrovic, B. et al. 2002) and to study large-scale motions in proteins and their complexes (Hornak, V. et al. 2006).

2 Theoretical background

2.1 Explicit solvent framework

In conventional MD simulations of solute-water systems, water molecules are described explicitly in atomic detail employing an atom based molecular force field. Such MD simulations involve routinely a large number of atoms, where most of them belong to solvent molecules. Water interacts with solute molecules in three ways: two direct and one indirect type of interaction. The former two are electrostatic (Warshel, A. et al. 1991) and van der Waals (vdW) interactions (Dill, K. A. 2002), the latter is due to the hydrophobic effect (Kauzmann, W. 1959; Tanford, C. 1968; Southall, N. T. et al. 2001; Hummer, G. et al. 1996; Franks, F. 1972-1982; Ben-Naim, A. 1980). The influence of hydrogen bonds (H-bonds) is generally accounted for by a suitable combination of electrostatic and vdW interactions. In ES models, the solvent, modeled as explicit water, screens Coulomb (and also vdW) interactions between solute atoms and competes with the direct solute-solute atom pair interactions (Coulomb as well as vdW).

2.2 Implicit solvent framework

In an IS model, the solute-solvent electrostatic interactions is approximated using a dielectric continuum model (Honig, B. et al. 1995). In this formalism, a solvated biomolecule may best be described as a set of explicit solute charges embedded in a low-dielectric cavity (generally $\varepsilon = 1$) that is surrounded by a high-dielectric continuum environment ($\varepsilon = 80$ for water). The boundary between low and high dielectric medium is essentially defined by the surface separating the vdW solute volume (given by the merged volumes of solute atoms) from the solvent. In this approach, the solute atomic partial charges interact with virtual surface charges induced at the solute-solvent boundary. The corresponding electrostatic energies can

be evaluated with the Poisson or for non-vanishing ionic strength with the Poisson-Boltzmann equation (Feig, M. et al. 2004; Baker, N. A. 2005; Im, W. et al. 1998).

Since solving the Poisson equation at each time step of MD slows down the simulations considerably, more approximate electrostatic models are used. These are for instance the generalized Born approximation (GB) (Still, W. C. et al. 1990), even more simplifying approaches such as a distance dependent dielectric model (Ramstein, J. et al. 1988; Wang, L. et al. 2002; Kosikov, K. M. et al. 2002), neutralized charged groups method (Lazaridis, T. et al. 1999; Feig, M. et al. 2004; Chen, J. et al. 2008; Tomasi, J. 2004), generalized reaction field model (Tironi, I. G. et al. 1995), analytical linearized Poisson-Boltzmann method (ALPB) (Sigalov, G. et al. 2006) and image-charge solutions based model (Cai, W. et al. 2007; Abagyan, R. et al. 1994). Very few PB-based molecular dynamics have been reported in (Luo, R. et al. 2002; Grant, J. A. et al. 2001; Prabhu, N. V. et al. 2004). As an efficient alternative the analytical GB model has become popular in MD simulations (Bashford, D. et al. 2000; Dominy, B. N. et al. 1999; Calimet, N. et al. 2001; Tsui, V. et al. 2000) because of its computational efficiency and relative simplicity compared to numerical solutions of the PB equation.

In addition to the electrostatic energy contributions, IS force fields contain a non-polar energy term that accounts for the entropic costs for water to be in contact with the surface of a solute. In touch with the surface of a solute a water molecule can no longer adopt as many different H-bond patterns as in bulk water. This goes along with a loss of entropy of these water molecules, the so-called hydrophobic effect. Consequently, water has positive contact energies with solute molecules, forcing different solute molecules to aggregate and individual solute molecules to assume a compact conformation with minimal surface. This non-polar contribution to the solvation free energy consists of favorable van der Waals attraction between the solute and solvent molecules and unfavorable cost of breaking the structure of solvent (water) around the solute i.e. cost of cavity formation. These two terms may be considered separately but are often combined into a single simple term that is assumed to be directly proportional to the solvent accessible surface area (SASA) (Lee, B. et al. 1971; Richmond, T. J. 1984; Fraczkiewicz, R. et al. 1998).

In IS model, the hydrophobic effect is normally accounted for by an artificial energy term that is proportional to the solvent accessible solute surface area (Lee, B. et al. 1971; Richmond, T. J. 1984; Fraczkiewicz, R. et al. 1998) with a proportionality

constant γ varying between 5 and 40 cal/(mol Å^2) (Onufriev, A. et al. 2002; Im, W. et al. 2003; Zhu, J. et al. 2002; Mongan, J. et al. 2007; Spassov, V. Z. et al. 2002).

3 Limitations of Implicit solvent model

There have been a number of efforts in recent times to advance IS models (Ferrara, P. et al. 2002; Feig, M. et al. 2004; Feig, M. et al. 2004; Chen, J. et al. 2008; Lopes, A. et al. 2007) to simulate dynamics of biomolecules efficiently and to come close to results of computationally more expensive conventional MD simulations with an ES model. Although the continuum dielectric medium models explicit electrostatic solute-solvent interactions generally faithfully, it was occasionally observed that solute-solute H-bonds and salt-bridges are too persistent in IS model with GB, as observed in CHARMM22/GBSW model (Chen, J. et al. 2006). This discrepancy with the explicit water model manifests as an overstabilization of non-native conformations of peptides using OPLS (Jorgensen, W. L. et al. 1988; Jorgensen, W. L. et al. 1996) all atom force field and surface-generalized Born (SGB) (Zhou, R. et al. 2002) and overestimation of the melting temperature of proteins using AMBER (Weiner, S. J. et al. 1986; Weiner, S. J. et al. 1984; Cornell, W. D. et al. 1995) parm94 force field and GB/SA (Pitera, J. W. et al. 2003).

In ES simulations (and also in real molecular systems) the attractive solute-solute and solute-solvent vdW interactions compete with each other and thus balance the composition of solute and solvent atoms in the neighborhood of solutes. In absence of explicit solvent, only the always attractive solute-solute vdW interactions are left and as a result the neighborhood of solute molecules is unbalanced. These solute-solute interactions occur between atoms of different solute molecules (intermolecular) but also between atoms of the same solute molecule (intramolecular). Removing the explicit solvent the solute molecules interact differently. Since all atoms are subject to the mutual attractive vdW interactions, therefore, MD simulations in absence of solvent leads to aggregation of different solute molecules and for flexible solute molecules to self-aggregation resulting in too compact solute conformers. The influence of solute-solvent vdW interactions on molecular solvation energies, ligand binding and protein docking was discussed many times. However, it

was first pointed out in (Pitera, J. W. et al. 2001) that this effect can also influence molecular conformations significantly.

The hydrophobic effect (Kauzmann, W. 1959; Tanford, C. 1968; Hummer, G. et al. 1996; Southall, N. T. et al. 2001; Ben-Naim, A. 1980) is based on the entropy content of the water structure, which assumes a maximum for undisturbed bulk water. In contrast to the mutual attraction of solute molecules in IS models, which is an artefact of the unbalanced attractive vdW interaction, the hydrophobic effect has a physical basis. It causes an effective attraction between different solute molecules and in case of flexible solute molecules also between atoms of the same solute molecule. It is the hydrophobic effect, which is for instance the driving force of protein folding (Dill, K. A. 2002; Privalov, P. L. et al. 1993; Honig, B. et al. 1995). This effect has been considered properly in explicit water MD simulations (Makarov, V. A. et al. 1998; Cheatham Iii, T. E. et al. 2003). However, in implicit water simulations, this effective attractive hydrophobic interaction between solute molecules is absent, while conversely the lack of balance in the attractive vdW interactions between solute-solute and solute-solvent contacts leads to an artificial attraction between solute molecules. Hence, under the conditions of an IS model these two effects partially compensate each other.

Hence, in an IS model, proper tuning is required for the surface energy replacing the hydrophobic effect and the strength of the unbalanced vdW attractions. For similar reasons it may be necessary to adjust Coulomb interactions for specific atom pairs of the solute to compensate for the absence of explicit H-bonds with water molecules, which may have not been considered appropriately by the continuum electrostatic approach. Similar but more subtle effects can be due to specific H-bond pattern between solute and water molecules, which may stabilize specific solute conformations and thus change the equilibrium distribution between them. To account for all these effects appropriately, not only specific atom pair Coulomb interactions but also specific torsion potentials need to be adjusted.

4 Polyethylene glycol: model system for IS

Polyethylene glycol (PEG) (Bailey, F. E. et al. 1976; Harris, J. M. 1992) involving the ethylene glycol repeat unit (-CH$_2$-CH$_2$-O-) (Fig. 1) has been considered as model system for the study. The IS model parameter have been adjusted first for 1,2-dimethoxy-ethane (DME), the monomer unit of PEG and are then transferred to PEG.
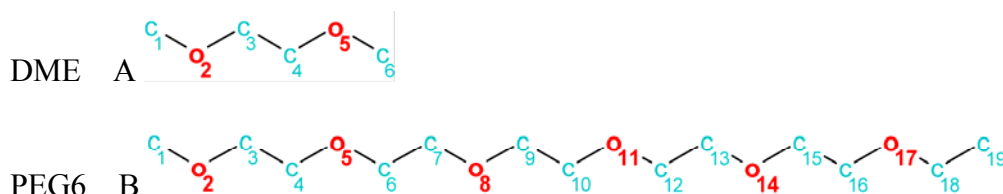
DME A

PEG6 B

Figure 1: Schematic structure representations of DME (A) and PEG6 (B). Atoms are numbered to refer to specific numbering used in text.

The DME structure can be characterized by three dihedral angles defining the local arrangement of a set of four covalently connected atoms (C-O-C-C, O-C-C-O and C-C-O-C). Each of the torsion angles can be in one of the three possible torsion angle intervals, which are centered around 180^0 called trans (T) conformer or around $+60^0$ or -60^0 corresponding to the two possible gauche (G, G') conformers (Fig. 2).

Eclipsed	Staggered	
	Trans (T)	Gauche (G or G´)
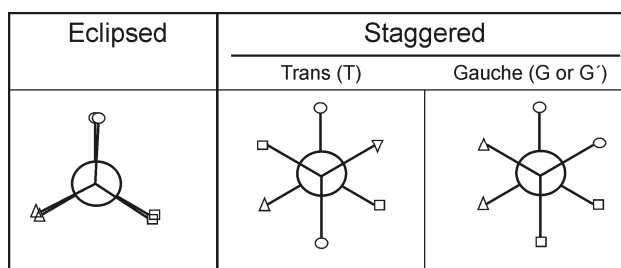
Figure 2: Newman projection of chemical conformations assessing the torsion angle around a chemical bond between atoms represented by dot (front) and circle (back), both have three distinct atoms represented by different symbols. In eclipsed conformation the torsion angle is minimized, in trans (T) it is 180^0 whereas in gauche it is $+60^0$ (-60^0) for G (G').

Based on different combinations of dihedral angles, DME has ten idealized conformers (Fig. 3). Amongst these TGT, TGG, TTT, TGG' and TTG are the principal conformers. *Ab intio* quantum chemical calculations of DME, where solvent effects are accounted for by continuum electrostatic models (Mueller-Plathe, F. et al. 1994) as well as MD simulations with umbrella sampling in explicit water (Liu, H. et

al. 1995) indicate that DME prefers the TGT conformer, which is also evident from Raman (Goutev, N. et al. 2000), IR (Yoshida, H. et al. 1992) and NMR studies (Tasaki, K. et al. 1985).
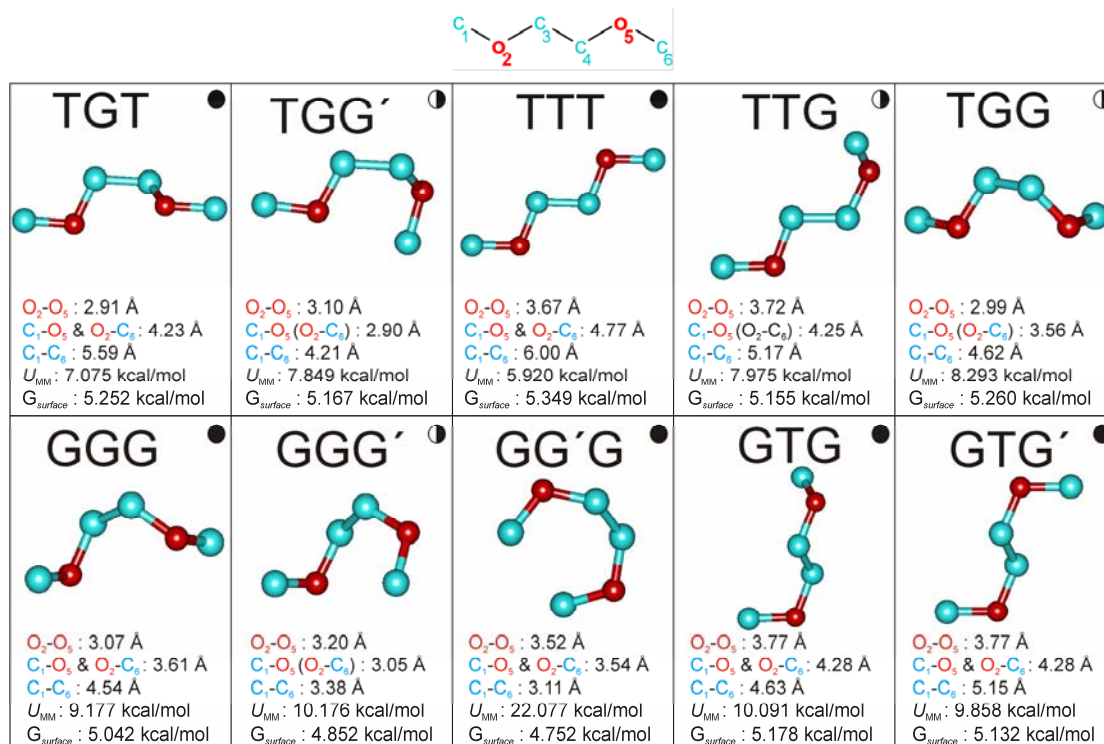
Figure 3: Structure of the ten idealized DME conformers. These conformers are named according to their three consecutive dihedral angles (C-O-C-C, O-C-C-O and C-C-O-C) centered at either 180^0 trans (T) or gauche $+60^0$ (G) or -60^0 (G'). The filled (half) filled circle on the upper right corner labels symmetric (non-symmetric) DME conformers. The values below the structures are: 1-4 oxygen-oxygen (**O_2-O_5**) atom pair distances, 1-5 oxygen-carbon (C_1-O_5, O_2-C_6), end-to-end atom pair (C_1-C_6) distances in Å, U_{MM} in kcal/mol represents the vacuum molecular mechanics potential energy of the conventional CHRAMM35 force field and $G_{surface}$ in kcal/mol is the surface energy of the standard IS model of CHARMM35 for the idealized DME conformers. Torsion angles are exactly at the values of 180^0 and $\pm60^0$ corresponding to the idealized geometries.

Explicit solvent MD simulations of DME, adopts a mixture of mainly four conformers. There are with decreasing order population TGT, TGG', TGG, TTT (Fig. 3, and Fig. 4). These DME conformers have been observed in various solvents like water (Masatoki, S. et al. 1995), methanol and carbon tetrachloride (Begum, R. et al. 1998). Among these four principal conformers, it has been noticed that TGT has the lowest free energy. In the experiments (Masatoki, S. et al. 1995; Begum, R. et al. 1998), one observes an interesting inverse correlation between gauche C-C bond and C-O bond populations, which is prevalent in DME and PEG chains and thus corroborates the choice of DME as model compound to adjust the PEG force field (Lee, H. et al. 2008).
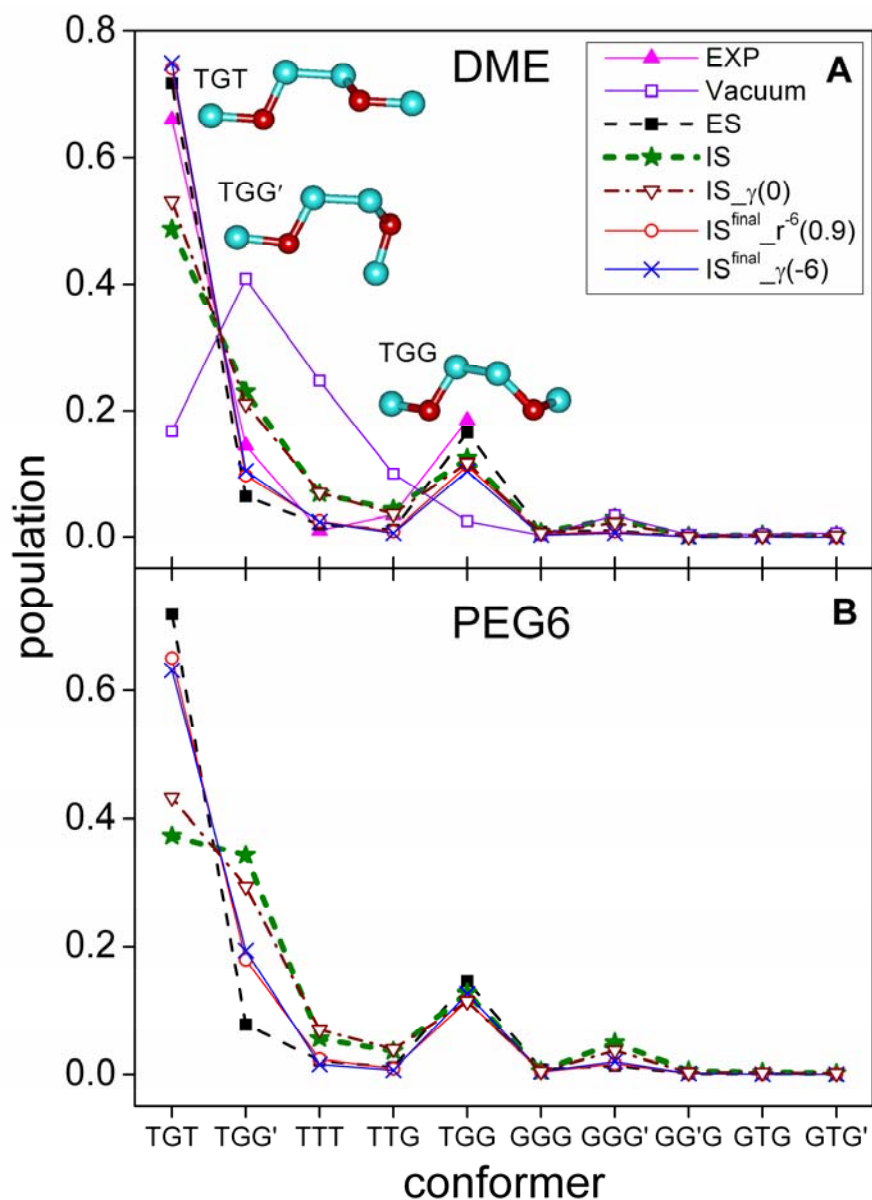
Figure 4: Population of local conformers of DME (part A) defined in Fig. 3 and PEG6 (part B). The experimental results (magenta solid line with filled up triangles in part A) are based on Raman spectra of DME in mixtures of water and DME (Goutev, N. et al. 2000), which were extrapolated to pure water. Vacuum simulation data for DME is presented as well (violet solid line with open squares in part A). Also shown are MD simulation data based on the ES model (black dashed line with filled squares) and based on four IS models. The conventional IS model with surface energy $[\gamma = 30 \text{ cal/(mol Å}^2)]$ and no further parameter tuning is shown as green dashed line with filled stars. Results obtained with the GB model (IS with vanishing surface energy) are shown as wine dashed-dotted line with open triangles. The MD simulation data of the two optimized IS models (see Table I) are also shown. One optimized IS model is with vanishing surface energy and 90% of the attractive vdW interaction (red solid line with 'o') The other with negative surface energy $[\gamma = -6 \text{ cal/(mol Å}^2)]$ and vdW interactions unchanged (blue solid line with 'x'). The experiments and the MD simulations data show predominant populations of the DME structures in the conformers TGT, TGG´ and TGG whose schematic structures are shown in the insert.

Crystallized and amorphous PEG adopt predominantly helix-like conformers (Fig. 5) (Takahashi, Y. et al. 1973; Matsuura, H. et al. 1969). They involve the most prevalent local structure TGT (Fig. 3 and Fig. 4) found in DME and PEG. The helical structure of PEG is stabilized by water bridges between nearest neighbor oxygen atoms in PEG. This has been reported by NMR (Lusse, S. et al. 1996), *ab intio* quantum chemical method (Wang, R. L. C. et al. 1997) and MD simulation (Tasaki, K. 1996; Depner, M. et al. 1991; Heymann, B. et al. 1999) studies.
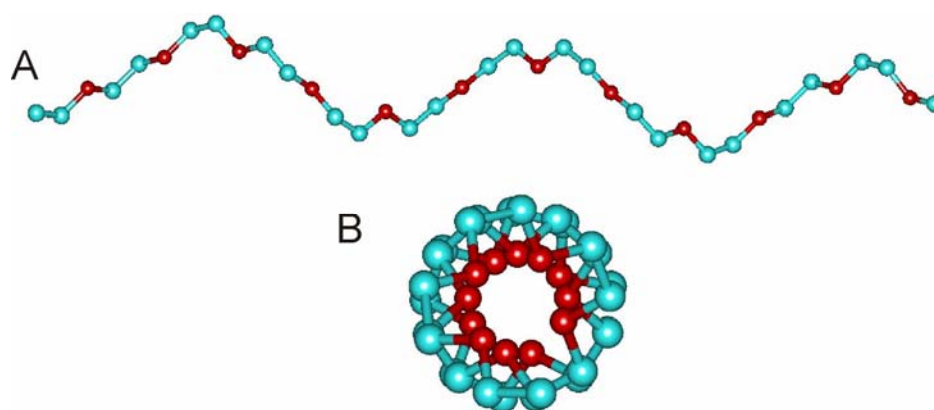
Figure 5: Side (A) and top (B) view of the PEG helix conformer composed of 12 monomer units in local TGT conformations comprising two helix turns. TGT is the most dominant local structure of PEG as well as of DME (Fig. 3 and Fig. 4).

PEG is water soluble, non-toxic, degrades slowly by metabolic enzymes and possesses low immunogenicity (Pasut, G. et al. 2004; Mero, A. et al. 2009). Therefore, it is widely used as excipient in different pharmaceutical formulations, foods, and cosmetics (Fuertges, F. et al. 1990; Harris, J. M. et al. 2003). PEG is also used as precipitant agent for protein crystallization (Lascombe, M.-B. et al. 2000). Being flexible and water-soluble, PEG can be used to create high osmotic pressure (Herrmann, A. et al. 1985; Zimmerberg, J. et al. 1986). Coating gene therapy vectors with PEG reduces innate immune responses (O'Riordan, C. R. et al. 1999).

Interaction of water with PEG gives rise to dominant steric repulsive forces which is crucial for protein resistance, therefore, self assembled mono-layers on a gold surface resist protein adsorption from aqueous solution, if terminated with PEG (Prime, K. L. et al. 1993; Pertsin, A. J. et al. 1998; Harder, P. et al. 1998; Wang, R. L. C. et al. 1997). PEG is of interest for the pharmaceutical industry to encapsulate drugs in nano-particle structures or dendrimers made of PEG based polymeric material (Veronese, F. et al. 2002; Roberts, M. J. et al. 2002; Harris, J. M. et al. 2003;

Guillaudeu, S. J. et al. 2008; Veronese, F. et al. 2005). It can also be used to connect ligands serving as drugs to form dimeric or even multimeric ligands to enhance drug activity by the multivalent binding effect (Mammen, M. et al. 1998; Kramer, R. H. et al. 1998). In multivalent ligand binding the chain entropy of PEG plays a key role to understand the effect quantitatively. So far only a simple Gaussian chain model was used to describe the influence of the PEG linker on bivalent ligand binding (Kramer, R. H. et al. 1998; Diestler, D. J. et al. 2008, 2009).

5 Methodology

5.1 General overview

Because of the importance of PEG it was in the focus of many experimental (Pertsin, A. J. et al. 1998; Oesterhelt, F. et al. 1999) and computational studies (Heymann, B. et al. 1999). Recently the PEG force field parameters (Vorobyov, I. et al. 2007; Lee, H. et al. 2008) for explicit water MD simulations were optimized in CHARMM (Brooks, B. R. et al. 2009; MacKerell, A. D. J. et al. 1998). The present study on IS models is using results of MD simulations with ES based on this improved force field as a reference. The main object of the study is to demonstrate how large deviations in local and global conformational features of PEG can be if one uses state of the art IS models for a flexible molecule and what one has to change to obtain a faithful force field for IS MD simulations.

In the recent years, much effort was placed in the development of theoretical models explaining the conformational preference of 1,2-dimethoxy-ethane (DME), the monomer unit of PEG (Bailey, F. E. et al. 1976). For this purpose specialized classical force field parameters for MD simulations of DME and PEG were developed (Neyertz, S. et al. 1994; Smith, G. D. et al. 1993) and applied (Smith, G. D. et al. 1995; Bedrov, D. et al. 1998) indicating the existence of specific DME conformers in the liquid phase. Recently, the CHARMM force field has been complemented with parameters for ether compounds (Vorobyov, I. et al. 2007), which were further adjusted in CHARMM35 (Lee, H. et al. 2008) using DME Raman spectra (Goutev, N. et al. 2000).

There is always a trade-off between the accuracy and speed when proposing a new computational model in molecular modeling applications. Since, the direct solution to Poisson equation is computationally expensive, therefore the generalized Born (GB) formalism has been commonly employed as an efficient approximation to numerical solutions of the Poisson equation at much reduced computational cost (Feig, M. et al. 2004). Inspired by the Born equation to compute solvation energies of ions (Born, M. 1920), a number of GB models (Still, W. C. et al. 1990; Edinger, S. R. et al. 1997; Dominy, B. N. et al. 1999; Bashford, D. et al. 2000; Im, W. et al. 2003; Zhu, J. et al. 2002; Mongan, J. et al. 2007) have been developed in the past years.

In GB models, the effective non-polar interactions related to the hydrophobic effect are proportional to the solvent accessible surface area (SASA) (Wesson, L. et al. 1992; Levy, R. M. et al. 2003). In the actual CHARMM force field version 35 (CHARMM35) (Lee, H. et al. 2008) the module for IS MD simulations uses as default value the proportionality constant $\gamma = 30$ cal/(mol Å^2) for the surface energy term. To evaluate the electrostatic interaction of solute with implicit solvent, generalized Born with a simple SWitching (GBSW) model as implemented in CHARMM (Im, W. et al. 2003; Im, W. et al. 2003; Chen, J. et al. 2006) has been used along with the ether force field parameterization of CHARMM35 (Lee, H. et al. 2008). Henceforth, we refer to MD simulations using CHARMM35 with the TIP3P water model (Jorgensen, W. L. et al. 1983) as explicit solvent (ES) model and in absence of solvent with standard CHARMM35 parameterization as IS model. The different IS models considered in the present study are also labeled by IS with additional information where they deviate from the default CHARMM35 parameter settings.

5.2 Comparing ES and IS models simulation

To compare the behavior of the DME and PEG force fields using the ES and the IS models, specific atom pair distance distributions, which were evaluated with standard histogram techniques using the conformers from MD simulation data was considered. For IS models, the attractive r^{-6} terms of LJ atom-pair interactions have been varied. Alternatively, the surface energy described by the surface tension coefficient γ (Gilson, M. K. et al. 1993; Hermann, R. B. 1972), whose value is

30 cal/(mol Å^2) in the CHARMM35 IS force field has also been varied. Fine tuning of the energy function of the IS model required also to change the Coulomb interactions of the 1-4 atom-pairs and of the 1-5 O-H (e.g. **O$_2$**–H$_6$ and **O$_5$**-H$_1$ where the subscripts at H refer to the particular carbon atoms to which the hydrogen atoms are attached, see Fig. 1) atom pairs. Furthermore, optimized polyether torsion potentials $V_{O-C-C-O}(\phi)$ and $V_{C-O-C-C}(\phi)$ describing the rotation barrier for the C-C and O-C bonds, for ES simulations in units of kcal/mol yield (Vorobyov, I. et al. 2007; Lee, H. et al. 2008)

$$V_{O-C-C-O}(\phi) = 0.59(1+\cos(\phi-\pi))+1.16(1+\cos(2\phi)) \tag{1a}$$

$$V_{C-O-C-C}(\phi) = 0.57(1+\cos(\phi))+0.29(1+\cos(2\phi))+0.43(1+\cos(3\phi)) \tag{1b}$$

To match the torsion angle distributions between ES and IS simulations, the parameter of the first term in $V_{O-C-C-O}(\phi)$ was increased from 0.59 to 1.09, while the parameter of the second term in $V_{C-O-C-C}(\phi)$ was reduced from 0.29 to 0.20. These were obtained by first evaluating the probability distribution of the torsion angles from the conformers of the MD simulation data by a histogram method. These probability distributions were then transformed to free energies by taking the negative logarithm of these probabilities and multiplying them by $k_B T$ with T = 300K.

5.3 Scaling parameters/force field for IS model

To visualize the effects of force field and the default parameters on the conformations, end-to-end distance distribution of DME is obtained from MD simulations under condition of vacuum, ES and IS models (Fig. 6). It is not surprising that the differences of DME conformations between explicit water and vacuum are enormous. As well as the conformational distributions of DME obtained with standard IS simulation conditions differ considerably from results of ES MD simulations, particularly the most prominent TGT conformer is less populated with IS than with ES simulation conditions (compare green and black dashed lines in Fig. 6).
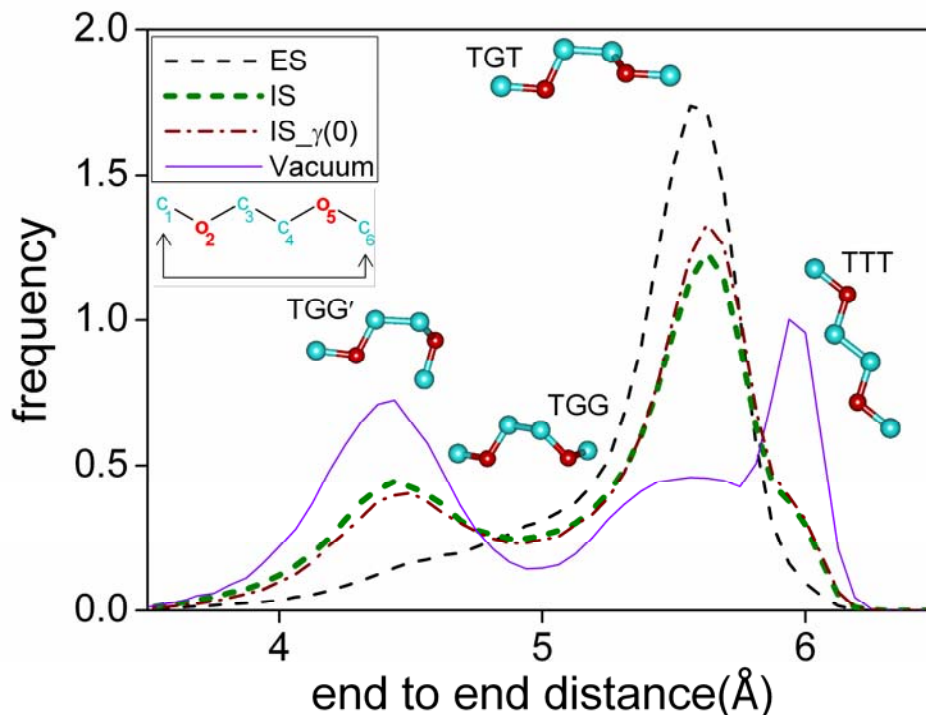
Figure 6: Conformational distribution of DME monitored by the $C_1 - C_6$ end-to-end distance obtained from MD simulations under different force field conditions. Violet solid line: vacuum condition; black dashed line: explicit water; wine dashed-dotted line: IS with GB only (vanishing surface energy); green dashed line: standard IS (GB and surface energy).

With the standard IS model, DME conformers are generally too compact due to the attractive surface energy term and the unbalanced vdW attractive r^{-6} terms that preferentially populate the TGG' and TGG conformers. Although DME conformer distributions obtained with the standard IS model show large deviations relative to results from the ES model, the TGT conformer corresponding to the helix conformation of PEG (Fig. 5) remains to be the most populated. Hence, the electrostatic screening effect due to unspecific H-bond pattern of DME-water interaction in ES model, which is modelled by GB electrostatics, approximates qualitatively the most prominent features of the DME conformer distribution. Astonishingly, the DME conformers obtained with the bare GB model (vanishing surface energy) (wine dashed-dotted line in Fig. 6) show nearly the same distribution as with the standard IS model (green dashed line in Fig. 6). Apparently the four most populated DME conformers (TGT, TGG', TGG, TTT) possess nearly the same surface and therefore their populations do not depend on the surface energy term (refer to $G_{surface}$ values in Fig. 3).

To bridge the gap between ES and IS models, the energy function for the IS model was optimized first considering DME, the monomeric unit of PEG. For this purpose, the dependencies of the IS model for DME on the strength of the vdW attraction and the surface energy term were monitored, as well as the dependence on the strength of specific Coulomb interactions. These factors were found to influence the MD simulation data for the IS DME models. These interactions were optimized before fine tuning the energy function with respect to the torsion potentials. Finally, we investigated whether the DME optimized energy function can be transferred to PEG.

5.3.1 Role of vdW attraction and surface energy

To establish the effect of vdW attraction and surface energy on the DME conformers, the differences between end-to-end distance distributions of the DME atoms C_1 and C_6 (Fig. 1) obtained from the conventional ES and the standard IS DME model is demonstrated (Fig. 7) and it is clearly visible that with the IS model (green dashed line in Fig. 7), the compact DME structures at 4.5 Å are considerably more populated at the expense of the more extended structures at about 5.5 Å. Moreover, in the IS model a shoulder appears at 6.0 Å end-to-end distances, which is absent in the ES model (black dashed line in Fig. 7). The enhanced occurrences of the compact structures in the IS model are believe to be due to unbalanced vdW interactions.

One way to qualitatively correct for this effect is by decreasing the attractive long range wing of the LJ interaction (except for the 1-4 atom pair interaction), while another way is by decreasing the surface energy term into the negative regime (red and blue solid lines in top and bottom parts of Fig. 7, respectively). The negative surface energies that account for unbalanced vdW interactions were also considered to characterize the energetics of protein mutants (Lopes, A. et al. 2007). However, negative surface energies were occasionally also used for IS models to account for dielectric screening, in the absence of more expensive electrostatic approaches like GB or PB (Ferrara, P. et al. 2002).
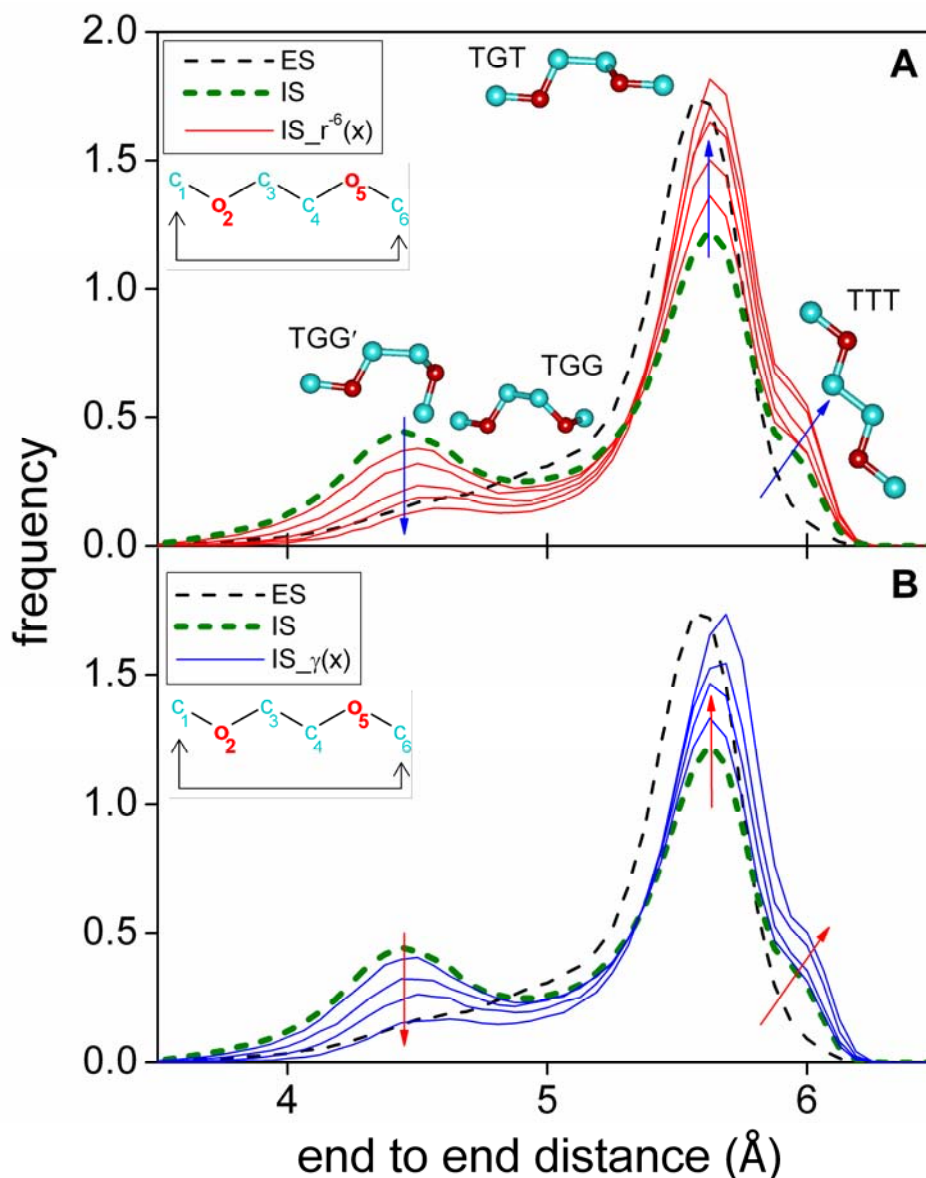
Figure 7: Role of attractive vdW and surface energy. The end-to-end distance distributions of DME obtained from MD simulation are shown for ES (black dashed line) and standard IS (green dashed line) solvent models. (A) The r^{-6} vdW attraction (except the 1-4 atom pair interaction) is scaled down by the factors 0.9, 0.7, 0.5, 0.3 and 0.1 (red solid lines). The four most dominant DME structures are shown for the maxima at 4.5 and 5.5 Å and the shoulder at 6.0 Å, corresponding to structures where subsequent three torsion angles (C-O-C-C, O-C-C-O, C-O-C-C) correspond to TGT, TGG', TGG, TTT, respectively. (B) The surface tension coefficient γ is varied from 0, -50, -100 and -200, cal/(mol Å2) (blue solid lines). Note that negative values of γ favor conformers with large surfaces. The arrows indicate the direction of the changes in the distribution occurring by decreasing either the vdW attraction (A) or the surface tension coefficient (B).

With a surface tension parameter γ as low as -200 cal/(mol Å^2) the effect from the surface energy term saturates yielding mainly the extended DME structure. However, even after these corrections, the resulting end-to-end distance distributions

of DME in implicit solvent still show significant deviations relative to the reference distribution from ES MD simulations (black dashed line in Fig. 7). The side maximum at 4.5 Å remains, while it appears in the ES MD simulation as a shoulder only. Furthermore, by lowering the vdW attractive r^{-6} or the surface energy term the shoulder at 6.0 Å becomes even more pronounced and the main maximum is shifted to larger distances in particular when lowering the surface energy.

5.3.2 Role of 1-4 and 1-5 atom pair interactions

Reduction in vdW attractive r^{-6} term as one of the options to improve the agreement between ES and IS models of DME clearly indicates that it would be useful to first optimize the non-bonded interactions. Exploring it further, the dependence of the end-to-end distance distribution of DME on the 1-4 atom pair Coulomb interaction and the 1-5 oxygen-hydrogen (O-H) (e.g. O_2-H_6 and O_5-H_1, H_i is attached to C_i) Coulomb interaction were investigated (magenta solid lines in top and bottom parts of Fig. 8, respectively). For these atom pairs because of the lack of competition with explicit solvent molecules the influence from these direct interactions may be overestimated. Therefore, these Coulomb interactions were scaled down by the factors 0.9, 0.7, 0.5, 0.3 and 0.1.

Down scaling of the 1-4 atom pair Coulomb interactions (Fig. 8A) removes the shoulder at 6.0 Å and simultaneously partially fills the valley between the side (4.5 Å) and main maximum (5.5 Å) of the end-to-end distribution. The former is mainly due to the 1-4 oxygen-oxygen (O-O) electrostatic repulsion. Furthermore, while the height of the main maximum remains essentially invariant; its position is shifted slightly to lower distances.
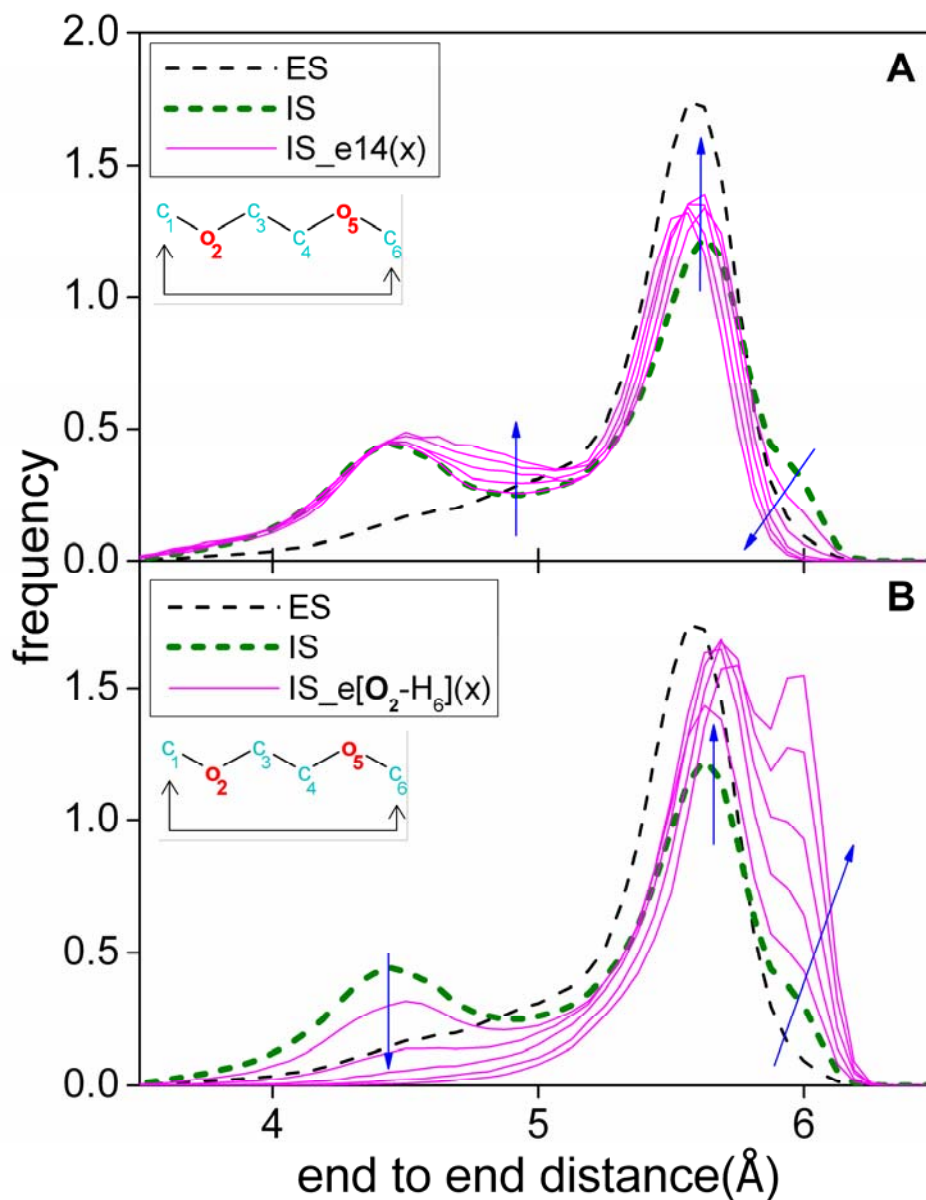
Figure 8: Role of 1-4 and 1-5 O-H atom pair interactions for the end-to-end distance distributions of DME obtained from MD simulation. In (A) and (B) the black dashed line refers to explicit (ES), the green dashed line to standard implicit solvent (IS) model. (A) The 1-4 atom pair Coulomb interaction is scaled down by the factors 0.9, 0.7, 0.5, 0.3 and 0.1 (magenta solid lines). (B) The 1-5 oxygen-hydrogen (O_2-H_6, O_5-H_1) Coulomb interaction is scaled down by the factors 0.9, 0.7, 0.5, 0.3 and 0.1 (magenta solid lines). The arrows indicate the direction of the changes in the distribution occurring by decreasing either the 1-4 atom pair (A) or the oxygen-hydrogen (B) electrostatic interaction.

The 1-5 O-H interaction governs the strength of the weak H-bond between O_2 (O_5) and the methyl group at position 6 (1). This direct 1-5 O-H interaction is suspected to be weaker in explicit solvent. Scaling down this interaction converts the side maximum at 4.5 Å to a shoulder similar to the ES model but of lower intensity. Simultaneously the shoulder at 6.0 Å corresponding to all-trans stretched

conformation (Fig. 8B) grows considerably and the main maximum shifts to larger distances. The latter is unfavorable but in combination with the scaling of the 1-4 Coulomb interaction it can be useful. All these changes are useful to correct deficiencies that remained or appeared after diminishing vdW attraction or surface energy.

5.4 Fine tuning of non-bonding energy terms

To account for a cumulative effect of the above discussed four different non-bonded energy terms: (1) attractive vdW term, (2) surface energy, (3) 1-4 atom pair Coulomb energy and (4) 1-5 O-H Coulomb energy, appropriately all these terms need to be adjusted. This was done in a tedious manual optimization procedure, by monitoring end-to-end distance distribution of DME (see Fig. 9A). Two possible nearly equivalent solutions were found. The first option is with vanishing surface energy ($\gamma = 0$) where the vdW attractive r^{-6} term is reduced to 90% (red solid line with open circles 'o' in Fig. 9). The second option is to assume a slightly repulsive surface energy term [$\gamma = -6$ cal/(mol Å^2)] with no change in the vdW attractive interaction (blue solid line with 'x' in Fig. 9). Both the solutions have in common that 1-4 atom pair and 1-5 O-H Coulomb energies are reduced to 90%. The corresponding end-to-end distance distributions are already close to the results obtained with the ES model (black dashed line in Fig. 9) which is used as reference, but there is still room for improvement, for which the torsion potentials were subjected to optimization.
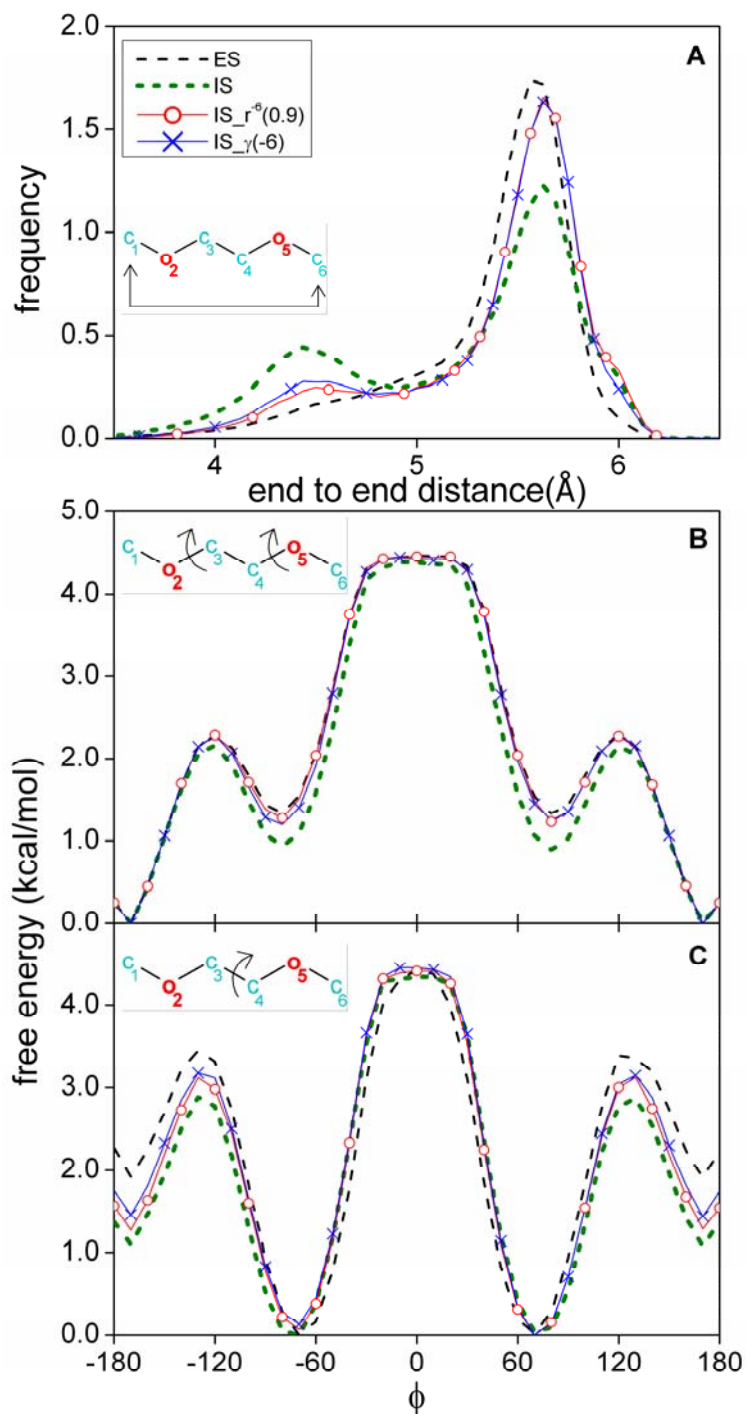
Figure 9: DME simulation data. Black (green) dashed lines are the results based in the ES (standard IS) model. All other data have in common that the 1-4 and 1-5 O-H atom pair Coulomb interactions are reduced to 90%. Red (blue) lines with 'o' ('x') symbols display MD simulation data where the surface energy vanishes ($\gamma = 0$) and the attractive vdW interaction is reduced to 90% (the surface energy is negative [$\gamma = -6$ cal/(mol Å^2)] and the vdW interaction unchanged). (A) End-to-end (C_1-C_6) distance distribution of DME. (B) Free energy profiles averaged over the two C-O-C-C torsion angles (C_1-O_2-C_3-C_4 and C_3-C_4-O_5-C_6). (C) Free energy profile of the O-C-C-O torsion angle (O_2-C_3-C_4-O_5).

5.5 Fine tuning of torsion potentials

Free energy profiles of the torsion angles as derived from MD simulation data of ES and IS models of DME were examined. The results of MD simulations for the torsion potentials involving rotations around two C-O bonds and one C-C bond of DME are shown in part B and C of Fig. 9, respectively. C-C bond rotation is the most critical, where IS models (blue and red lines in Fig. 9C) still show larger deviations from the MD simulation data obtained with ES model (black dashed line in Fig. 9C).

To obtain an agreement in the torsion angle distributions between ES and IS simulations, the parameter of the first term in $V_{O\text{-}C\text{-}C\text{-}O}(\phi)$ of eq. (1a) was increased from 0.59 to 1.09, while the parameter of the second term in $V_{C\text{-}O\text{-}C\text{-}C}(\phi)$ of eq. (1b) was reduced from 0.29 to 0.20.

5.6 Optimized IS model parameters

With this choice of appropriate torsion potentials, it is possible to essentially close the gap appearing in the effective C-C bond torsion potential (Fig. 11B) between the results of ES reference model and IS models. Slight improvements have also been observed for C-O torsion angle distribution (Fig. 11A). End-to-end, 1-5 O-C and 1-4 O-O distance distributions of IS models for DME agree well with the MD simulation data of ES model (Fig. 10). The final set of optimized parameters for IS models of DME including also the torsion potential corrections are described in Table I.

Table I: Parameters of DME for explicit (ES) and implicit solvent (IS) models.

force field term	CHARMM35 force field		adjusted CHARMM35 force field	
	ES	IS (standard)	IS^{final}_r^{-6}(0.9)	IS^{final}_γ(-6)
O-C-C-O (eq. 1a)	0.59/1/180	0.59/1/180	1.09/1/180	1.09/1/180
C-O-C-C (eq. 1b)	0.29/2/0	0.29/2/0	0.20/2/0	0.20/2/0
vdW (no14)	1.0	1.0	0.9	1.0
e14fac	1.0	1.0	0.9	0.9
elec O(1)-H(6)	1.0	1.0	0.9	0.9
γ [cal/mol.Å^2]	N/A	30	0	-6

The two adjusted CHARMM35 (Lee, H. et al. 2008) ether force fields. Both involve down scaling of 1-4 all atom and 1-5 O-H atom pair Coulomb interactions to 90% and optimized parameters of torsion potentials (eq. 1a & 1b). The IS^{final}_r^{-6}(0.9) and IS^{final}_γ(-6) force fields refer to the cases of (i) vanishing surface energy ($\gamma = 0$) and 90% of attractive vdW interaction and (ii) negative surface energy [$\gamma = -6$ cal/(mol Å^2)] and vdW interaction unchanged.
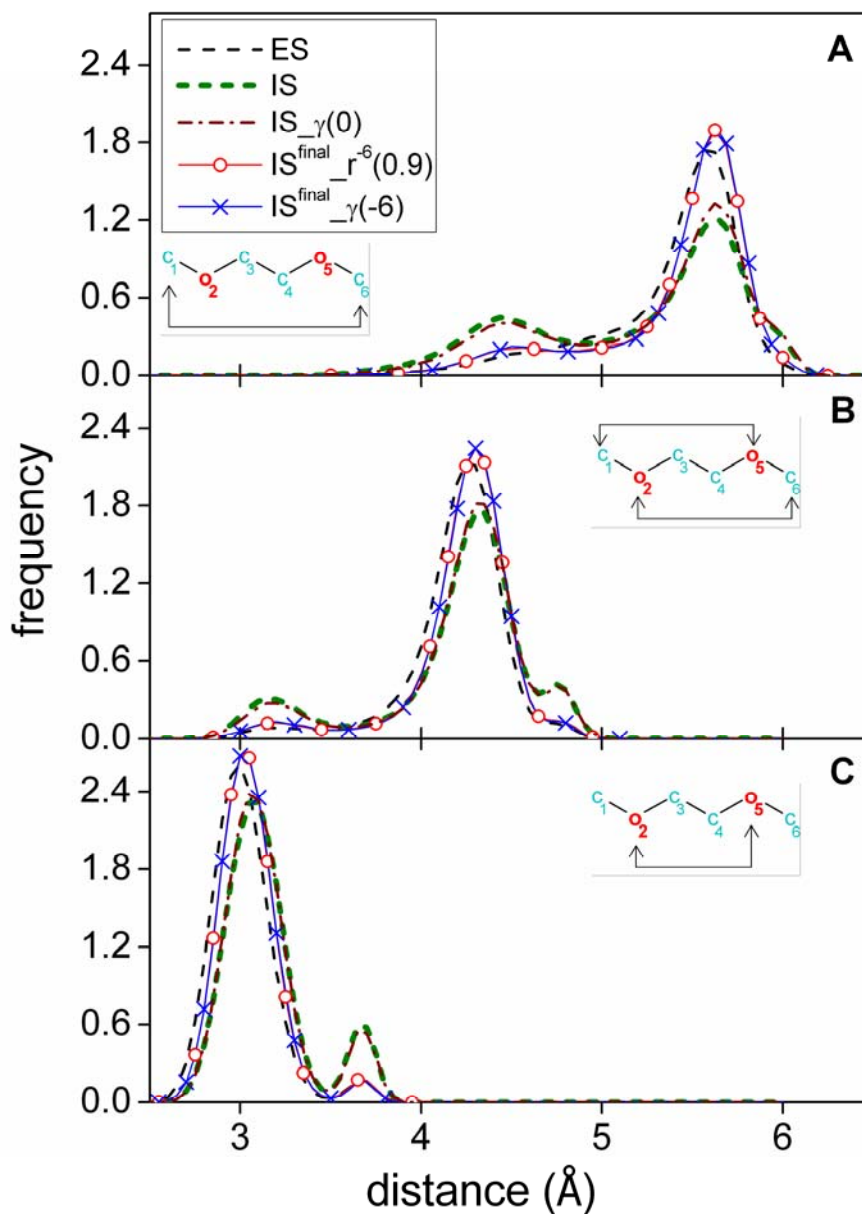
Figure 10: Atom pair distance distributions of DME: (A) end-to-end (C_1-C_6); (B) 1-5 O-C atom pairs (averaged over C_1-O_5, and O_2-C_6); (C) 1-4 O-O atom pair (O_2-O_5). The distributions were obtained from MD simulation data of ES (black dashed lines), standard IS i.e. GB and surface energy (green dashed lines) and GB only (vanishing surface energy) (wine dashed-dotted lines) models. The red [blue] solid lines with 'o' ['x'] symbols show the distributions of two different optimized IS models obtained from MD simulation data where the surface energy vanishes ($\gamma = 0$) and the attractive vdW interactions are reduced to 90% [the surface energy is negative ($\gamma = -6$ cal/(mol Å^2)) and the vdW interaction unchanged] with 1-4 and 1-5 O-H atom pair Coulomb interactions reduced to 90% and torsion angle correction as described in Table I.

Since the surface energy practically does not vary between the four most populated DME conformers (TGT, TGG', TGG, TTT) ($G_{surface}$ values in Fig. 3), the populations of DME conformers obtained with the bare GB and the standard IS

models for the atom pair distances and torsion potentials are nearly identical (compare wine dashed-dotted and green dashed lines in Fig. 10 and 11, respectively). This is also evident from the population of the local DME conformers (Fig. 4A).
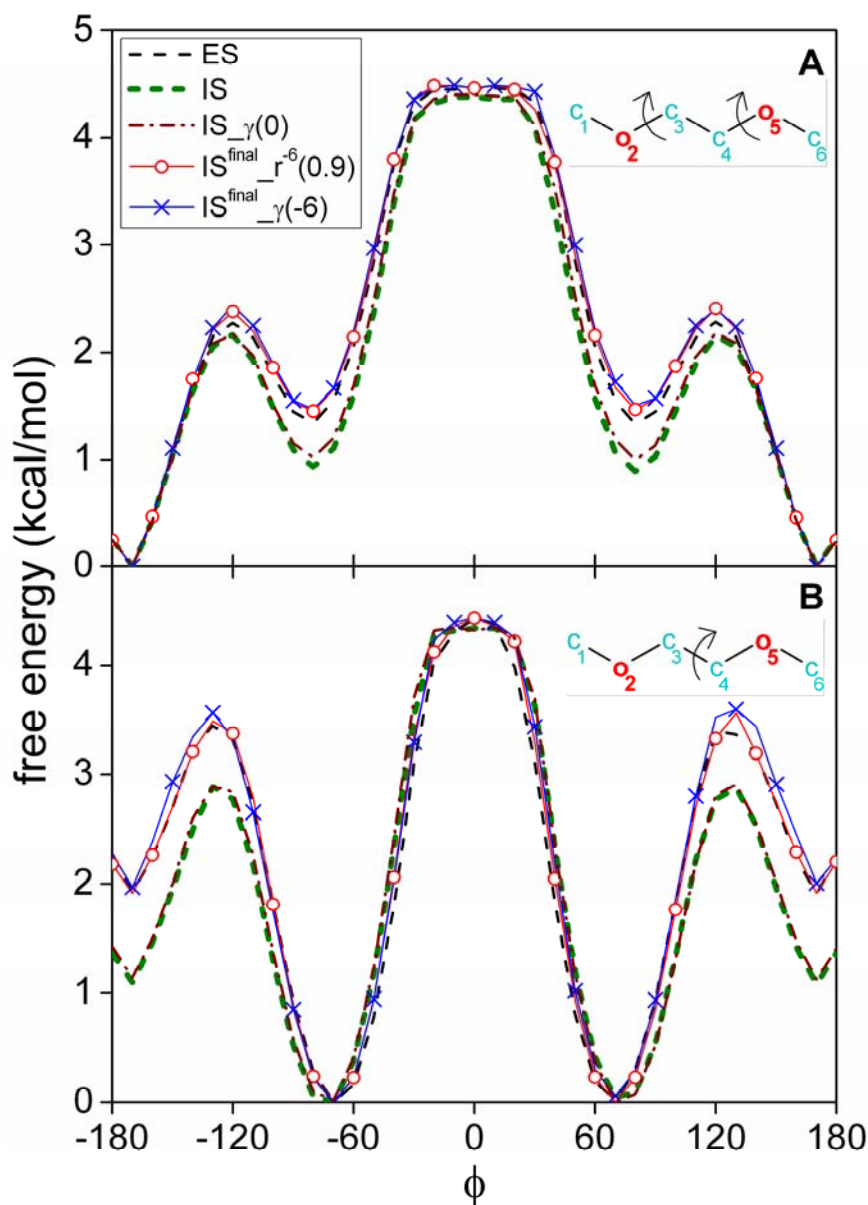
Figure 11: DME simulation data. Free energy profiles (A) average over two C-O-C-C torsion angles (e.g., C_1-O_2-C_3-C_4 and C_3-C_4-O_5-C_6) and (B) O-C-C-O (O_2-C_3-C_4-O_5) torsion angles obtained with ES (black dashed line), standard IS model (green dashed line) and GB only (vanishing surface energy) (wine dashed-dotted line) models. The red [blue] solid lines with 'o' ['x'] symbols show the distributions of two different optimized IS models obtained from MD simulation data where the surface energy vanishes ($\gamma = 0$) and the attractive vdW interactions are reduced to 90% [the surface energy is negative ($\gamma = -6$ cal/(mol Å^2)) and the vdW interaction unchanged] with 1-4 and 1-5 O-H atom pair Coulomb interactions reduced to 90% and torsion angle correction as described in Table I.

6 Transferring IS model from DME to PEG

In order to validate whether the IS model developed for DME can be transferred to PEG, we compare MD simulation data of PEG6 (with 6 monomer units of DME) based on the ES model with data of PEG6 obtained from the two IS models optimized for DME. For this, local conformers and global behavior of PEG6 were analyzed.
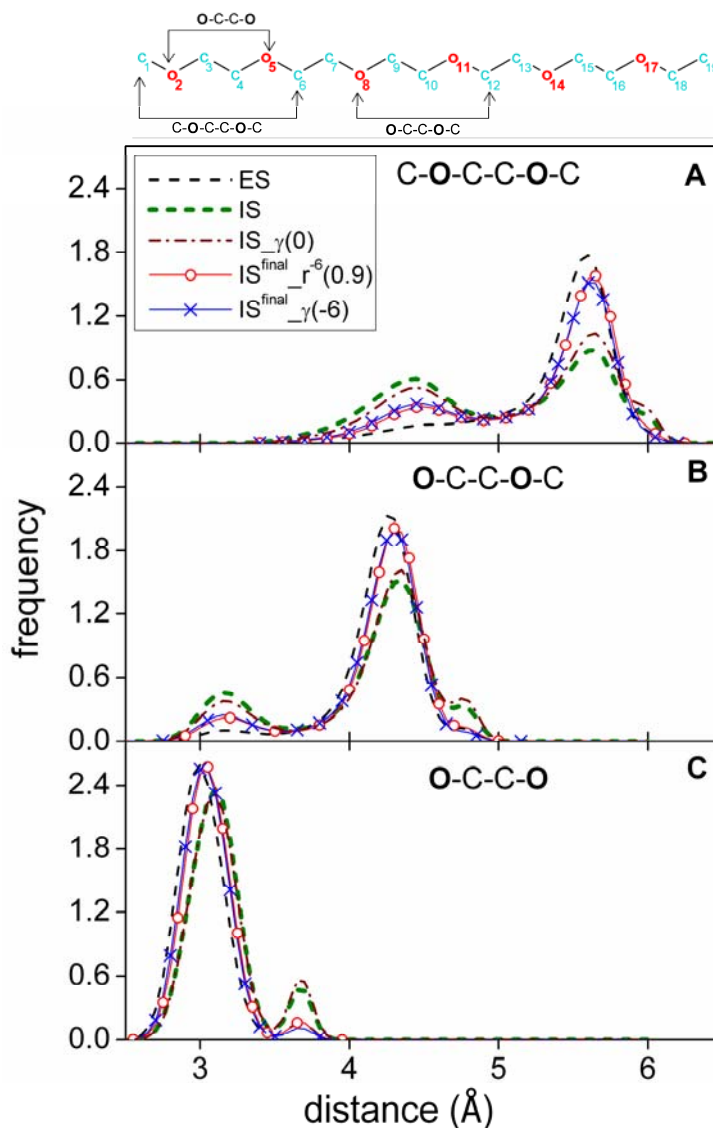
Figure 12: Atom pair distance distributions of PEG6: (A) 1-6 C-C (average of C_1-C_6, C_4-C_9, C_7-C_{12}, C_{10}-C_{15}, C_{13}-C_{18}); (B) 1-5 O-C (average of O_2-C_6, O_5-C_1, O_5-C_9, O_8-C_4, O_8-C_{12}, O_{11}-C_7, O_{11}-C_{15}, O_{14}-C_{10}, O_{14}-C_{18}, O_{17}-C_{13}); (C) 1-4 O-O (average of O_2-O_5, O_5-O_8, O_8-O_{11}, O_{11}-O_{14}, O_{14}-O_{17}). The distributions were obtained from MD simulation data of ES (black dashed line), standard IS i.e. GB and surface energy (green dashed line) and GB only (vanishing surface energy) (wine dashed-dotted line) models. The red [blue] solid lines with 'o' ['x'] symbols show the distributions of two different optimized IS models obtained from MD simulation data where the surface energy vanishes ($\gamma = 0$) and the attractive vdW interactions are reduced to 90% [the surface energy is negative ($\gamma = -6$ cal/(mol Å^2)) and the vdW interaction unchanged] with 1-4 and 1-5 O-H atom pair Coulomb interactions reduced to 90% and torsion angle correction as described in Table I.

The C_1–C_6, 1-5 O-C (**O_2-C_6**) and 1-4 O-O (**O_2-O_5**) atom pair distance distributions that reflect the local conformers of PEG6 (Fig. 12) are virtually identical to the corresponding atom pair distribution of DME (Fig. 10). Also the torsion potentials obtained for PEG6 (Fig. 13) correspond very well to the DME data (Fig. 11). With both the optimized force fields, PEG6 shows practically the same agreement with data obtained from its ES model as for DME. The corresponding distance distributions obtained from both the IS models PEG6 for more distant O-O atom pairs agreed well with the ES model as shown in Figure 14.
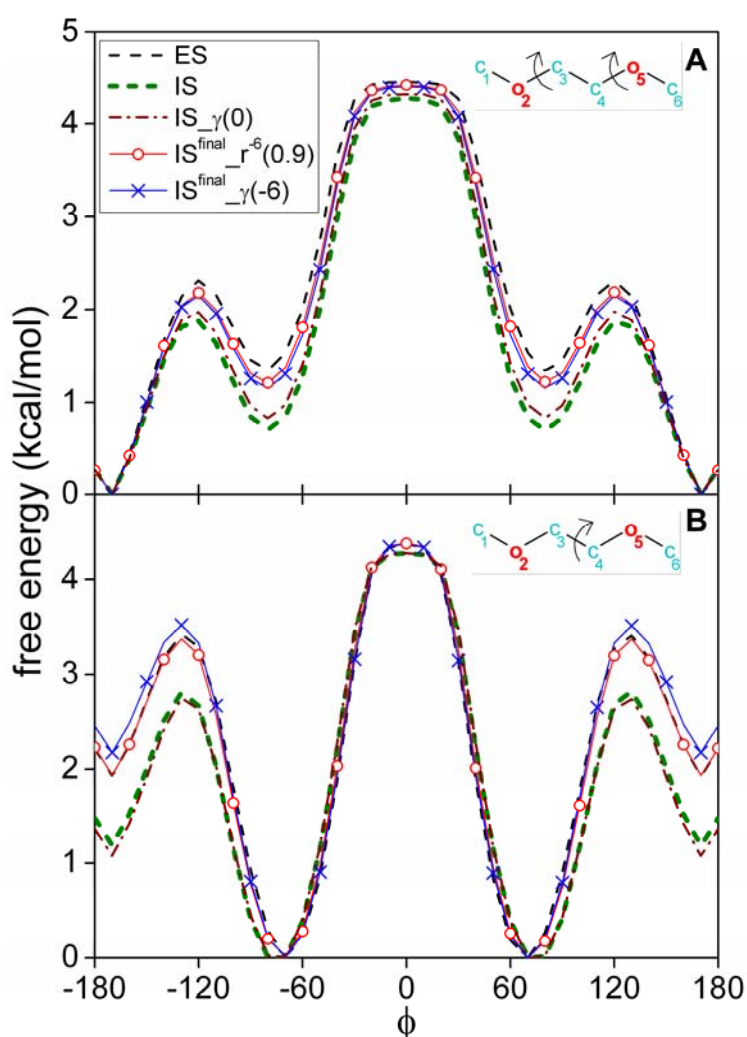
Figure 13: Average free energies of all possible (A) C-O-C-C (e.g., C_1-**O_2**-C_3-C_4 and C_3-C_4-**O_5**-C_6 etc) and (B) O-C-C-O (e.g., **O_2**-C_3-C_4-**O_5**, **O_5**-C_6-C_7-**O_8** etc) torsion angles of PEG6 obtained from ES (black dashed line), standard IS (green dashed line) and GB only (vanishing surface energy) (wine dashed-dotted line) models. The red [blue] solid lines with 'o' ['x'] symbols show the distributions of two different optimized IS models obtained from MD simulation data where the surface energy vanishes ($\gamma = 0$) and the attractive vdW interactions are reduced to 90% [the surface energy is negative ($\gamma = -6$ cal/(mol Å^2)) and the vdW interaction unchanged] with 1-4 and 1-5 O-H atom pair Coulomb interactions reduced to 90% and torsion angle correction as described in Table I.
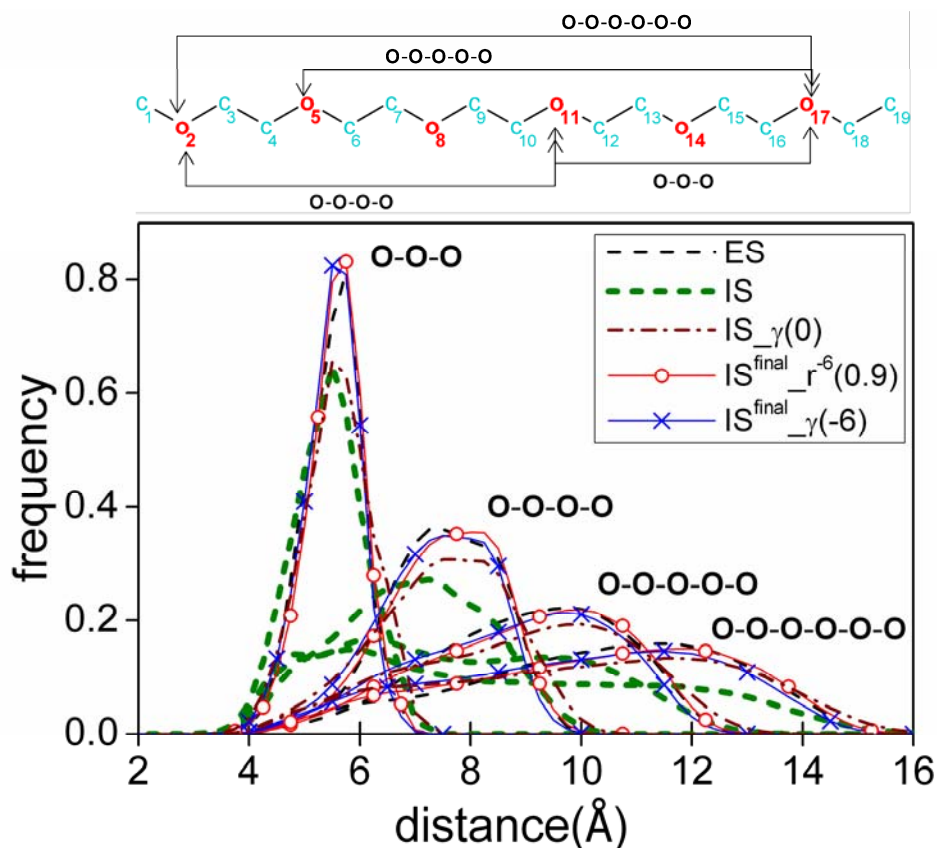
Figure 14: Average O-O distance distribution of PEG6 with one oxygen atom in between (O_2-O_8, O_5-O_{11}, O_8-O_{14}, O_{11}-O_{17}), two oxygen atoms in between (O_2-O_{11}, O_5-O_{14}, O_8-O_{17}), three oxygen atoms in between (O_2-O_{14}, O_5-O_{17}) and four oxygen atoms in between (O_2-O_{17}). The black dashed lines refer data obtained with ES, green dashed lines with standard IS and wine dashed-dotted lines with GB only (vanishing surface energy) models. The red [blue] solid lines with 'o' ['x'] symbols show the distributions of two different optimized IS models obtained from MD simulation data where the surface energy vanishes ($\gamma = 0$) and the attractive vdW interactions are reduced to 90% [the surface energy is negative ($\gamma = -6$ cal/(mol Å^2)) and the vdW interaction unchanged] with 1-4 and 1-5 O-H atom pair Coulomb interactions reduced to 90% and torsion angle correction as described in Table I.

The global behavior of the PEG6 conformers is probed by the end-to-end (C_1–C_{19}) atom pair distance distribution. Agreement with the global behavior of conformers obtained with the ES model is much more demanding for the larger PEG chain molecules. The agreement of the optimized IS models with the results of the ES reference model is fairly good while the standard IS model fails (compare red [blue] solid lines and green dashed line with black dashed line in Fig. 15, respectively). The maximum of the end-to-end distance distribution is at 13.5 Å for the ES model, it is only at 7 Å for the standard IS model. The most probable distance of PEG6 in the ES model is considerably smaller than the end-to-end distance in the ideal TTT conformation and the ideal TGT helix conformation, which are 24 Å and 18.5 Å, respectively.
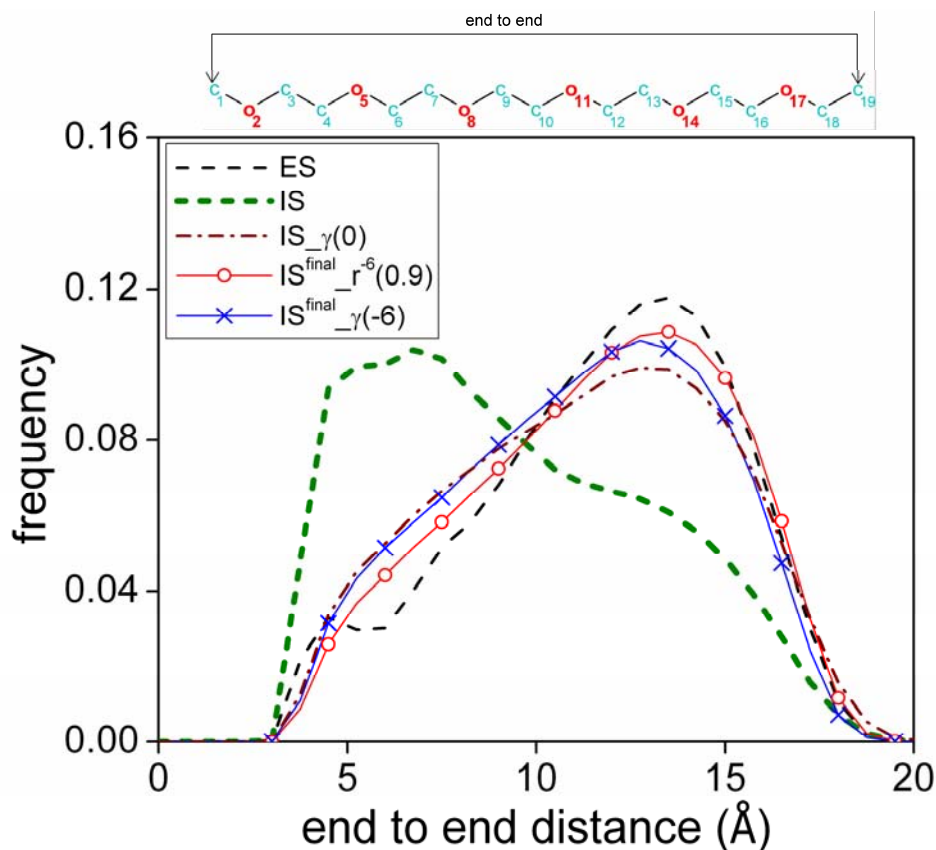
Figure 15: End-to-end (C_1-C_{19}) distance distribution obtained from MD simulation data for PEG6 using ES and different IS models. Black (green) dashed lines are the results based in the ES (standard IS) model. The wine dashed-dotted line is obtained with GB only (vanishing surface energy). All other data have in common that the 1-4 and 1-5 O-H atom pair Coulomb interactions are reduced to 90% and the torsion potentials are optimized as described in Table I. The red [blue] solid lines with 'o' ['x'] symbols show final optimized MD simulation data where the surface energy vanishes ($\gamma = 0$) and the attractive vdW interaction is reduced to 90% (the surface energy is negative [$\gamma = -6$ cal/(mol Å^2)] and the vdW interaction unchanged).

The end to end distance of the energy minimized PEG6 helix conformer remains with about 17.2 Å for the standard and the optimized IS models close to value of the ideal helix structure of PEG6. In addition to end-to-end distance distribution, the radius of gyration is also measured to assess global the structure of PEG6 as shown in Figure 16.

It is also observed that the bare GB model without surface energy term applied to PEG6 yields an end-to-end distance distribution very close to the results obtained with the ES model (refer to wine dashed-dotted and black dashed line in Fig. 15, respectively). However, the local conformations obtained with the bare GB model differ from the ES model for both DME and PEG6 in the same way (refer to wine dashed-dotted and black dashed lines, and compare Figs. 10 & 11 with Figs. 12 & 13,

respectively) interestingly. It has been reported that using small or nearly vanishing surface energies agrees with recent applications to compute solvation energies (Wagoner, J. A. et al. 2006; Sitkoff, D. et al. 1994; Onufriev, A. et al. 2002; Mongan, J. et al. 2007; Zhu, J. et al. 2002).
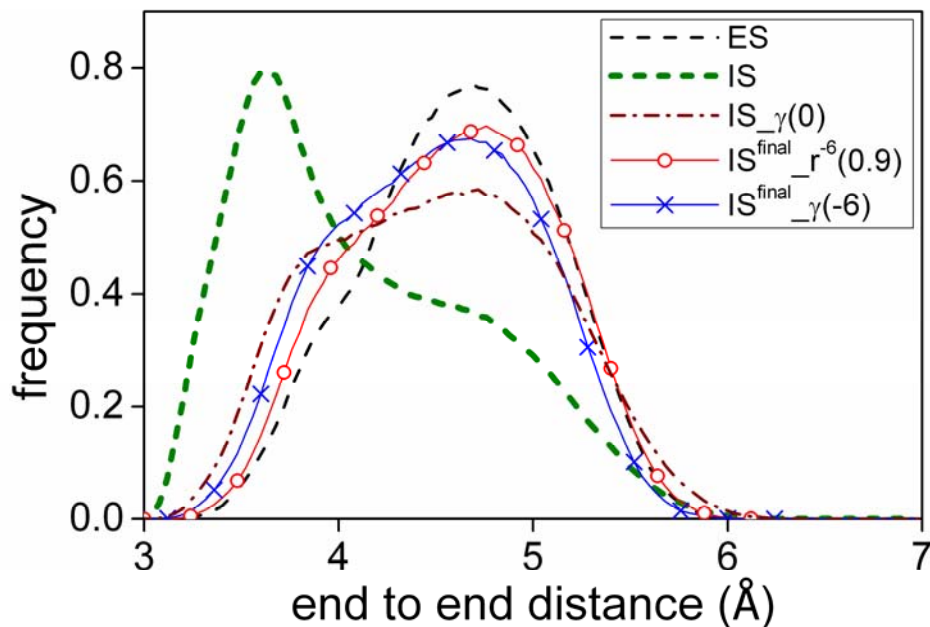
Figure 16: Mass weighted radius of gyration obtained from MD simulations of PEG6 with ES (black dashed line), standard IS (green dashed line) and GB only (vanishing surface energy) (wine dashed-dotted line) models. The red [blue] solid lines with 'o' ['x'] symbols show the distributions of two different optimized IS models obtained from MD simulation data where the surface energy vanishes (γ = 0) and the attractive vdW interactions are reduced to 90% [the surface energy is negative ($\gamma = -6$ cal/(mol Å^2)) and the vdW interaction unchanged] with 1-4 and 1-5 O-H atom pair Coulomb interactions reduced to 90% and torsion angle correction as described in Table I.

It was observed that the standard IS model of PEG6 yields conformers that are generally much more compact due to the attractive surface energy term and the unbalanced vdW attractive r^{-6} terms and both predominantly populate the TGG' and TGG conformers (see green dashed line in Fig. 15). The attractive Coulomb interactions of the 1-5 O–H atom pairs that preferentially populate the TGG' state of DME (refer to side maximum at 4.5 Å of green dashed line in Fig. 6) may contribute to the more compact PEG6 conformers of the standard IS model. However, reduction in the Coulomb interactions for all atom pairs of the type 1-5 O–H to 90% lead to much too extended PEG6 conformers. Hence, the best agreement with the ES model of PEG is obtained, if the force field optimized for the DME IS model is directly transferred to PEG.

7 Conformational sampling

It is interesting to compare how efficient the ES and IS models of PEG explore the conformational space with time. As expected, the behavior of the two models will differ, since in the absence of explicit solvent, the IS dynamics is not slowed down by solvent viscosity, whereas in the ES model, solvent viscosity slows down conformational transitions. To evaluate the efficiency of MD simulations exploring PEG conformational space, we examined end-to-end distance distribution of PEG6 with two approaches: (1) time evolution of the integral square deviations $[\Delta g]^2$ and (2) autocorrelation function $c(\Delta t)$ of the time dependent end to end distance.

7.1 Integral of square deviation

The time evolution of end-to-end atom pair distance distributions $g^{(t)}(x)$ were monitored for the total time span t used for the ensemble averages. A suitable quantity to measure how fast the limit distribution $g^{(\infty)}(x)$ of the ensemble is approached with the MD simulation time is given by integral of the square deviation

$$[\Delta g]^2 = \int [g^{(t)}(x) - g^{(\infty)}(x)]^2 \, dx . \tag{2}$$

In our study, we started with all trans stretched conformer of PEG6 and monitored how the end-to-end distance distribution $g^{(t)}(x)$ approaches the limit distribution $g^{(\infty)}(x)$ of the ES model obtained with the maximum available time span of 1 μs. This is shown in Fig. 17 in a double logarithmic plot where the time decay appears approximately linear corresponding to a $t^{-\alpha}$ power law with $\alpha = \frac{2}{3}$.

The end-to-end distance distribution of PEG6 seems to converge after about 50 ns using the ES model, however for the two IS models it is converged earlier at about 10 ns (black dashed and red (blue) solid lines in Fig. 17, respectively). For the two IS models beyond this time span until the maximum time span of 400 ns the integral square deviations of the end-to-end distributions are approximately constant at a low value that corresponds to the small deviations of the distributions that are based on the ES and the optimized IS models.
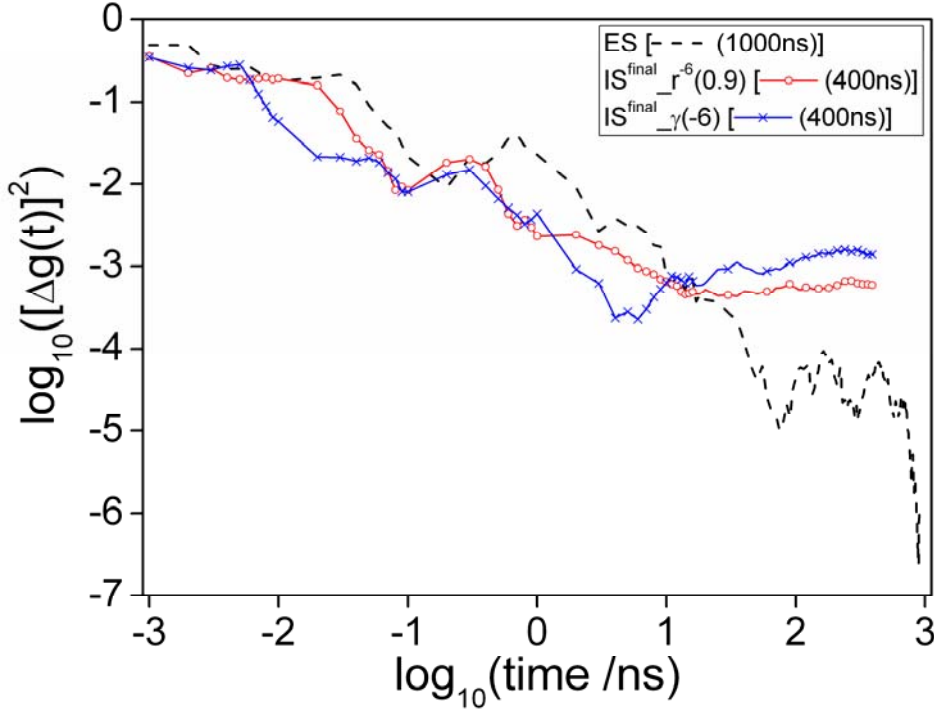
Figure 17: Efficiency of MD simulations to explore the conformational space of PEG6. Integral square deviation $[\Delta g]^2$, eq. (2), of end-to-end atom pair distance distribution as a function of the time span used to obtain the ensemble average. Each time span starts at t = 0 where the initial PEG6 conformer is stretched all-trans. The lengths of the time spans to evaluate the averages of the distance distributions range from 1 ps to up to 1 μs. Black dashed line displays $[\Delta g]^2$ for the ES model. Red (blue) solid lines show the time evolution of IS models with a maximum time span of 400 ns. The symbols 'o' ('x') refer to IS models with vanishing surface energy and 90% of vdW attraction (negative surface energy $[\gamma = -6$ cal/(mol Å2)] and unchanged vdW interaction). The reference ensemble average distribution $g^{(\infty)}(x)$ is for all three cases considered here is the distribution of the ES model obtained for the maximum time span of 1 μs. This is why $[\Delta g]^2$ vanishes at long times for the ES model, while it remains constant at a low value for the two IS models.

7.2 Autocorrelation function

Alternatively, end-to-end distance autocorrelation functions of PEG6 can be evaluated according to eq. (3), which exhibits a power law decay behavior $t^{-\alpha}$ with $\alpha = \frac{1}{2}$ (Fig. 18A) for the IS models, while for the ES model the decay behavior obeys a stretched exponential $\exp(-t^{-\beta})$ with $\beta = \frac{2}{3}$ (Fig. 18B). The end-to-end distance autocorrelation function of PEG6 decays by two orders of magnitude in about 300 ps for the IS and in about 1 ns for the ES model.

$$c(\Delta t) = \frac{\left\langle [d(t)-\overline{d}][d(t+\Delta t)-\overline{d}]\right\rangle_t}{\left\langle [d(t)-\overline{d}]^2\right\rangle_t}. \tag{3}$$
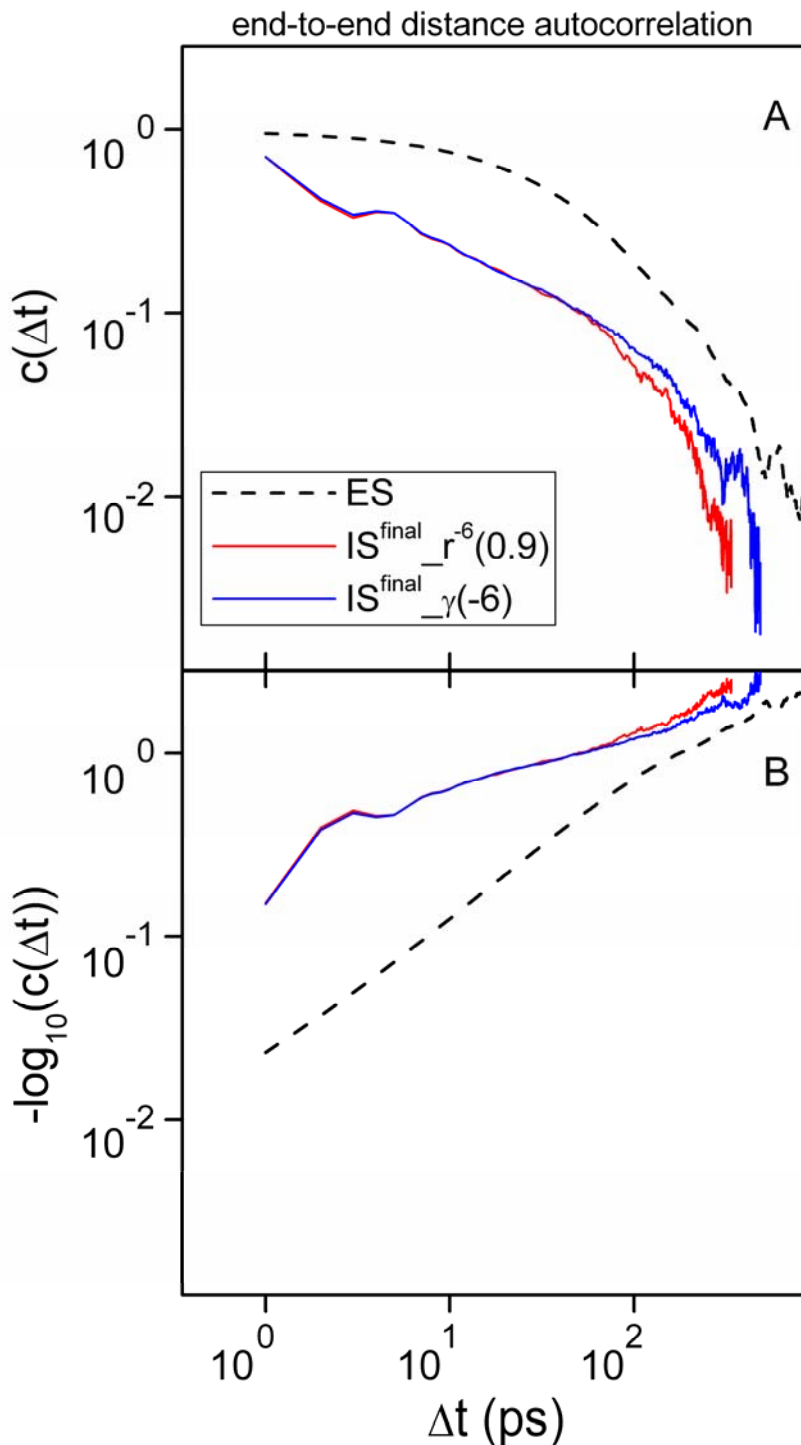
Figure 18: End-to-end distance autocorrelation functions of PEG6 according to eq. (3) in main text. (A) IS models exhibit a decay behavior that obeys the power law $t^{-\alpha}$ with $\alpha = \frac{1}{2}$. (B) The ES model exhibits decay behavior that obeys the stretched exponential $\exp(-t^{-\beta})$ with $\beta = \frac{2}{3}$. Black dashed line is the results based on the ES model. The red [blue] solid lines refers to two different optimized IS models obtained from MD simulation data where the surface energy vanishes ($\gamma = 0$) and the attractive vdW interactions are reduced to 90% [the surface energy is negative ($\gamma = -6$ cal/(mol Å^2)) and the vdW interaction unchanged] with 1-4 and 1-5 O-H atom pair Coulomb interactions reduced to 90% and torsion angle correction as described in Table I.

8 Simulation protocols

8.1 Explicit solvent simulation

The program CHARMM was used to prepare the explicit solvent simulation setup with the CHARMM35 ether force field (Lee, H. et al. 2008). A fully extended DME (PEG6) was used as initial conformer embedded in 199 (1477) water molecules in a periodic cubic box of 18 Å (35 Å) edge length. Before the MD simulation was started, energy minimization was performed in three stages: (1) minimizing only water molecules with 500 steps of conjugated gradient (CONG) and 2000 steps of adopted basis Newton-Raphson (ABNR), (2) minimizing all hydrogens only with 500 steps of CONG and 2000 steps of ABNR, (3) minimizing all atoms with 500 steps of CONG and 4000 steps of ABNR. We used the leap frog Verlet integrator (Hockney, R. W. 1970) implemented in CHARMM. The temperature was maintained at 300 K and the pressure at 1 atm using the Nose-Hoover thermostat (Nosé, S. et al. 1983; Hoover, W. G. 1985) and the Andersen-Hoover barostat (Andersen, H. C. 1980; Hoover, W. G. 1985) respectively. The TIP3P model (Jorgensen, W. L. et al. 1983; Durell, S. R. et al. 1994) was used for water. SHAKE (Ryckaert, J.-P. et al. 1977) was used to constrain covalent bonds involving hydrogens allowing a time step of 2 fs.

The MD simulation was started heating the molecular system to 300 K by velocity rescaling for 20 ps followed by equilibration for 20 ps. Subsequently, a first short production run of 100 ps was performed. The final set of coordinates was used as input for long term MD simulation with NAMD (Phillips, J. C. et al. 2005) which was performed for 100 ns with DME and 1 μs with PEG6. Atomic coordinates were saved every 0.2 ps. For both systems the first 1 ns was discarded to account for a more extensive equilibration. All parameters and conditions for the long term simulation with NAMD were the same as used before with CHARMM except that with NAMD temperature was controlled by a Langevin thermostat with friction coefficient $\beta = 2$ ps^{-1} and the electrostatic interactions were evaluated using particle mesh Ewald method (Darden, T. et al. 1993). Using the program NAMD, the MD simulation of PEG6 for 1 μs in explicit water took ~870 h on an IBM p595 machine (AIX 5.3, 1.9 GHz, dual core) using 16 CPUs with 32 processes. The corresponding time for 100 ns MD simulation of DME was ~50 h. The simulations with NAMD

were performed on the high performance computing facility "ABACUS4" of the ZEDAT, Freie Universität Berlin.

8.2 Implicit solvent simulation

All implicit solvent simulations were carried out with the generalized Born with a simple SWitching (GBSW) module (Dominy, B. N. et al. 1999) implemented in CHARMM using the ether force field (Lee, H. et al. 2008) and optimized PB atomic radii (Nina, M. et al. 1999) for smoothed dielectric boundary. A fully extended DME (PEG6) was used as initial conformer. The entire system was minimized with 2000 steps of steepest decent (SD) and 4000 steps of ABNR. A canonical NVT ensemble was used maintaining the temperature at 300K with the Nose-Hoover thermostat (Nosé, S. et al. 1983; Hoover, W. G. 1985). SHAKE (Ryckaert, J.-P. et al. 1977) was used to constrain covalent bonds involving hydrogens allowing a time step of 2 fs.

The MD simulation was started heating the molecular system to 300 K by velocity rescaling for 6 ps and then equilibrated for 20 ps followed by the production run of 100 ns for DME and 400 ns of PEG6. The first 1 ns was discarded to account for equilibration. Atomic coordinates are saved every 0.2 ps. For the implicit solvent simulations we used a single CPU of Intel Xeon L5420 2.5 GHz, where 100 ns of DME took ~16 h and 400 ns of PEG6 took ~150 h. Taking the performance difference in the CPUs and the different numbers of CPUs into account, the MD simulation of PEG6 with implicit solvent was about a factor of 20 more efficient than with explicit solvent. However, due to lack of solvent viscosity in the IS model, the sampling of PEG6 conformers was about five times faster than with ES model.

9 Conclusions

The accurate modeling of biomolecules in molecular environment requires realistic models for the interaction of the solvent with the biomolecule. In order to model such systems faithfully by molecular mechanics, both explicit solvent and implicit models have been developed. Explicit solvent (ES), the most widely used models to carry out MD simulations in solvent are computationally expensive due to the evaluation of a large number of atom pair interactions as solvent atoms are

represented explicitly. The ES model though offers elaborate atomic details of the solute-solvent interactions but at the same time limits its use to perform long term simulations as it involves the integration over numerous solvent degrees of freedom and thus converges slowly. An alternative to ES models are implicit solvent (IS) models, which make a compromise between efficiency and accuracy. The IS models offer several significant advantages over ES models, including lower computational costs and faster conformational search.

In an IS model, the electrostatic solute-solvent interactions are often modeled by simplified GB electrostatics and the hydrophobic effect is modeled by a surface energy term. In ES models, the attractive vdW interactions between solute atoms compete with corresponding interactions of solute-solvent atom pairs thus leading to balanced and effectively weaker solute-solute interactions. However, in IS models, where discrete solvent molecules are replaced by high dielectric continuum medium, the competition of attractive van der Waals (vdW) interactions between solute-solute and solute-solvent atom pairs are absent. As a consequence the unbalanced attractive vdW interactions between solute atom pairs yield too compact solute structures, which is clearly demonstrated by comparing MD simulation data of PEG based on ES and standard IS models (involving attractive surface energy terms).

On the other hand, the bare GB model (with vanishing surface energy) yields PEG conformer distributions, which are globally realistic but have local deficiencies, which are similar as observed for DME. The DME conformers obtained with the bare GB model (vanishing surface energy) show nearly the same distribution as with the standard IS model, suggesting that for small molecular units like DME the surface energy plays a minor role, since the molecular surface varies negligibly with the conformers and therefore their populations do not depend on the surface energy. The same holds true for local conformers of PEG6 which are represented by 1-6 C-C atom pair distance distributions.

Adjusting the force field to remove these deficiencies, it turns out that in order to balance for lack of solute-solvent attractive vdW interactions either the surface energy term must vanish and the attractive part of the vdW interactions be reduced to 90% or alternatively, to use even a slightly negative surface energy, if the vdW interactions remain unchanged, which suggests that the influence from the unbalanced attractive vdW interactions is even stronger than the hydrophobic effect.

Additionally small but nevertheless significant reductions of the 1-4 and 1-5 O-H Coulomb interactions and adjustments of the torsion potentials were applied to obtain faithful local DME conformers. The former force field adjustment was necessary, since the GB model has generally a tendency to overestimate Coulomb interactions at short distances. Interestingly, the force field originally adjusted for DME (the monomeric unit of PEG), could be transferred to PEG unchanged. The resulting energy function for PEG in implicit water is as accurate as in explicit water and allows simulating PEG dynamics several orders of magnitude more efficiently.

The CPU time per time step for PEG6 is about a factor of 20 larger with the ES than with the IS model. Conformational space exploration of PEG6 in explicit and implicit water obeys power law $t^{-\alpha}$ in time t with exponent $\alpha = \frac{2}{3}$. The dynamics of the end-to-end distance distribution of PEG6 obeys a stretched exponential decay law for the ES model while a power law was found for the IS model demonstrating differences in the long term behavior between these models. Due to the lack of viscosity in the IS model the PEG6 conformations are explored about five times faster with the IS than with ES model. Hence, the conformational space exploration of PEG6 for the IS model is in total about a factor of 100 faster than for the ES model.

In summary, IS models in molecular systems are important developments to enhance the efficiency of MD simulations. However, still better characterization is needed to further enhance the accuracy and computational efficiency of IS models. In this contribution, the adjusted parameters for IS models addresses issues like unbalanced attractive solute-solute vdW interactions and inefficient electrostatic screening between solute atoms. Both these deficiencies in IS models are traced back to the absence of solvent molecules.

PEG is an important molecular component in medicine and pharmacy. With the successful application of the proposed optimized IS force fields for DME and PEG along with the increased efficiency achieved with IS models, it would be rewarding to simulate the molecule of interest (PEG in our case) with IS models that are as realistic as ES model and several orders of magnitude more efficient.

10 References

Abagyan, R. and Totrov, M. 1994. Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins. *J. Mol. Biol.* **235**: 983-1002.

Andersen, H. C. 1980. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72**: 2384-2393.

Bailey, F. E. and Koleske, J. V. (1976). Poly(ethylene oxide). New York, Academic Press.

Baker, N. A. 2005. Improving implicit solvent simulations: a Poisson-centric view *Curr. Opin. Struct. Biol.* **15**: 137-143

Bashford, D. and Case, D. A. 2000. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**: 129-152.

Bedrov, D., Borodin, O. and Smith, G. D. 1998. Molecular Dynamics Simulations of 1,2-Dimethoxyethane/Water Solutions. 1. Conformational and Structural Properties. *J. Phys. Chem. B* **102**: 5683–5690.

Begum, R., Sagawa, T., Masatoki, S. and Matsuura, H. 1998. Infrared spectroscopic study of conformational properties of short chain poly(oxyethylene)s in methanol and carbon tetrachloride. *J. Mol. Struct.* **442**: 243-250.

Ben-Naim, A. (1980). Hydrophobic interactions. New York, Plenum Press.

Born, M. 1920. Volumen und Hydratationswärme der Ionen. *Z. Phys.* **1**: 45-48.

Brooks, B. R., Brooks, C. L., MacKerell, A. D. J., Nilsson, L., Petrella, R. J., Roux, B., et al. 2009. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**: 1545-1614.

Cai, W., Deng, S. and Jacobs, D. 2007. Extending the fast multipole method to charges inside or outside a dielectric sphere. *J. Comput. Phys.* **223**: 846-864.

Calimet, N., Schaefer, M. and Simonson, T. 2001. Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins: Struct., Funct., Bioinf.* **45**: 144-158.

Cheatham Iii, T. E. and Kollman, P. A. 2003. Molecular Dynamics Simulations of Nucleic Acids. *Annu. Rev. Phys. Chem.* **51**: 435-471.

Chen, J. and Brooks, C. L. 2008. Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. *Phys. Chem. Chem. Phys.* **10**: 471-481.

Chen, J., Brooks, C. L. and Khandogin, J. 2008. Recent advances in implicit solvent-based methods for biomolecular simulations *Curr. Opin. Struct. Biol.* **18**: 140-148.

Chen, J., Im, W. and Brooks, C. L. 2006. Balancing Solvation and Intramolecular Interactions:Toward a Consistent Generalized Born Force Field. *J. Am. Chem. Soc.* **128**: 3728–3736.

Corbeil, C. R. and Moitessier, N. 2009. Docking Ligands into Flexible and Solvated Macromolecules. 3. Impact of Input Ligand Conformation, Protein Flexibility, and Water Molecules on the Accuracy of Docking Programs. *J. Chem. Inf. Model.* **49**: 997-1009.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**: 5179-5197.

Cramer, C. J. and Truhlar, D. G. 1999. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **99**: 2161–2200.

Darden, T., York, D. and Pedersen, L. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**: 10089-10092.

Depner, M., Schürmann, B. L. and Auriemma, F. 1991. Investigation of a poly(oxyethylene) chain by a molecular dynamics simulation in an aqueous solution and by Langevin dynamics simulations. *Mol. Phys.* **74**: 715 - 733.

Diestler, D. J. and Knapp, E. W. 2008. Statistical Thermodynamics of the Stability of Multivalent Ligand-Receptor Complexes. *Phys. Rev. Lett.* **100**: 178101-4.

Diestler, D. J. and Knapp, E. W. 2009. Statistical Mechanics of the Stability of Multivalent Ligand Receptor Complexes. *J. Phys. Chem. C*.http://dx.doi.org/10.1021/jp904258c

Dill, K. A. 2002. Dominant forces in protein folding. *Biochemistry* **29**: 7133-7155.

Dominy, B. N. and Brooks, C. L. 1999. Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids. *J. Phys. Chem. B* **103**: 3765–3773.

Durell, S. R., Brooks, B. R. and Ben-Naim, A. 1994. Solvent-Induced Forces between Two Hydrophilic Groups. *J. Phys. Chem.* **98**: 2198–2202.

Edinger, S. R., Cortis, C., Shenkin, P. S. and Friesner, R. A. 1997. Solvation Free Energies of Peptides:Comparison of Approximate Continuum Solvation Models with Accurate Solution of the Poisson−Boltzmann Equation. *J. Phys. Chem. B* **101**: 1190–1197.

Eisenberg, D. and McLachlan, A. D. 1986. Solvation energy in protein folding and binding. *Nature* **319**: 199-203.

Feig, M. and Brooks, C. L. 2004. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **14**: 217-224.

Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A. and Brooks, C. L. 2004. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **25**: 265-284.

Ferrara, P., Apostolakis, J. and Caflisch, A. 2002. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Struct., Funct., Bioinf.* **46**: 24-33.

Fraczkiewicz, R. and Braun, W. 1998. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19**: 319-333.

Franks, F., Ed. (1972-1982). <u>Water: A Comprehensive Treatise</u>. New York, Plenum Pub Corp.

Frauenfelder, H., Fenimore, P. W., Chen, G. and McMahon, B. H. 2006. Protein folding is slaved to solvent motions. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 15469-15472.

Froloff, N., Windemuth, A. and Honig, B. 1997. On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. *Protein Sci.* **6**: 1293-1301.

Fuertges, F. and Abuchowski, A. 1990. The clinical efficacy of poly(ethylene glycol)-modified proteins. *J. Controlled Release* **11**: 139-148.

Gilson, M. K. 1995. Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.* **5**: 216-223.

Gilson, M. K., Davis, M. E., Luty, B. A. and McCammon, J. A. 1993. Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *J. Phys. Chem.* **97**: 3591–3600.

Goutev, N., Ohno, K. and Matsuura, H. 2000. Raman Spectroscopic Study on the Conformation of 1,2-Dimethoxyethane in the Liquid Phase and in Aqueous Solutions. *J. Phys. Chem. A* **104**: 9226–9232.

Grant, J. A., Pickup, B. T. and Nicholls, A. 2001. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comp. Chem.* **22**: 608-640.

Gregory, R. B., Ed. (1995). <u>Protein-Solvent Interactions</u>. New York, Marcel Dekker, Inc.

Guillaudeu, S. J., Fox, M. E., Haidar, Y. M., Dy, E. E., Szoka, F. C. and Fréchet, J. M. J. 2008. PEGylated Dendrimers with Core Functionality for Biological Applications. *Bioconjugate Chem.* **19**: 461-469.

Harder, P., Grunze, M., Dahint, R., Whitesides, G. M. and Laibinis, P. E. 1998. Molecular Conformation in Oligo(ethylene glycol)-Terminated Self-Assembled Monolayers on Gold and Silver Surfaces Determines Their Ability To Resist Protein Adsorption. *J. Phys. Chem. B* **102**: 426–436.

Harris, J. M., Ed. (1992). <u>Poly(Ethylene Glycol) Chemistry: Biotechnical and Biomedical Applications</u>. New York, Plenum Publishing.

Harris, J. M. and Chess, R. B. 2003. Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discovery* **2**: 214-221.

Hermann, R. B. 1972. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.* **76**: 2754–2759.

Herrmann, A., Arnold, K. and Pratsch, L. 1985. The effect of osmotic pressure of aqueous PEG solutions on red blood cells. *Biosci. Rep.* **5**: 689-696.

Heymann, B. and Grubmüller, H. 1999. Elastic properties of poly(ethylene-glycol) studied by molecular dynamics stretching simulations. *Chem. Phys. Lett.* **307**: 425-432.

Hockney, R. W. 1970. Potential calculation and some applications *Meth. Comput. Phys.* **9**: 136-211.

Honig, B. and Nicholls, A. 1995. Classical electrostatics in biology and chemistry. *Science* **268**: 1144-1149.

Honig, B. and Yang, A.-S. 1995. Free Energy Balance in Protein Folding. <u>Adv. Protein Chem.</u> Anfinsen, Richards, Edsall and Eisenberg, Academic Press. **46:** 27-58.

Hoover, W. G. 1985. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**: 1695-1697.

Hornak, V., Okur, A., Rizzo, R. C. and Simmerling, C. 2006. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 915-920.

Hummer, G., Garde, S., García, A. E., Pohorille, A. and Pratt, L. R. 1996. An information theory model of hydrophobic interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 8951-8955.

Im, W., Beglov, D. and Roux, B. 1998. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* **111**: 59-75.

Im, W., Feig, M. and Brooks, C. L. 2003. An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophys. J.* **85**: 2900-2918.

Im, W., Lee, M. S. and Brooks, C. L. 2003. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* **24**: 1691-1702.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**: 926-935.

Jorgensen, W. L., Maxwell, D. S. and Tirado-Rives, J. 1996. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**: 11225-11236.

Jorgensen, W. L. and Tirado-Rives, J. 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**: 1657-1666.

Karplus, M. and McCammon, J. A. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* **9**: 646-652.

Kauzmann, W. 1959. Some Factors in the Interpretation of Protein Denaturation. <u>Adv. Protein Chem.</u> C.B. Anfinsen, Anson, Bailey and Edsall, Academic Press. **Volume 14:** 1-63.

Klebe, G. 2006. Virtual ligand screening: strategies, perspectives and limitations. *Drug. Discov. Today* **11**: 580-594.

Knapp, E. W. and Muegge, I. 1993. Heterogeneous diffusion of water at protein surfaces: application to BPTI. *J. Phys. Chem.* **97**: 11339-11343.

Kosikov, K. M., Gorin, A. A., Lu, X.-J., Olson, W. K. and Manning, G. S. 2002. Bending of DNA by Asymmetric Charge Neutralization: All-Atom Energy Simulations†. *J. Am. Chem. Soc.* **124**: 4838-4847.

Kramer, R. H. and Karpen, J. W. 1998. Spanning binding sites on allosteric proteins with polymer-linked ligand dimers. *Nature* **395**: 710-713.

Lascombe, M.-B., Milat, M.-L., Blein, J.-P., Panabières, F., Ponchet, M. and Prangé, T. 2000. Crystallization and preliminary X-ray studies of oligandrin, a sterol-carrier elicitor from Pythium oligandrum. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**: 1498-1500.

Lazaridis, T. and Karplus, M. 1999. Effective energy function for proteins in solution. *Proteins: Struct., Funct., Bioinf.* **35**: 133-152.

Lee, B. and Richards, F. M. 1971. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55**: 379-400.

Lee, H., Venable, R. M., MacKerell, A. D. and Pastor, R. W. 2008. Molecular Dynamics Studies of Polyethylene Oxide and Polyethylene Glycol: Hydrodynamic Radius and Shape Anisotropy. *Biophys. J.* **95**: 1590-1599

Levy, R. M., Zhang, L. Y., Gallicchio, E. and Felts, A. K. 2003. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute−Solvent Interaction Energy. *J. Am. Chem. Soc.* **125**: 9523–9530.

Liu, H., Müller-Plathe, F. and van Gunsteren, W. F. 1995. A molecular dynamics simulation study with a combined quantum mechanical and molecular mechanical potential energy function: Solvation effects on the conformational equilibrium of dimethoxyethane. *J. Chem. Phys.* **102**: 1722-1730.

Lopes, A., Alexandrov, A., Bathelt, C., Archontis, G. and Simonson, T. 2007. Computational sidechain placement and protein mutagenesis with implicit solvent models. *Proteins: Struct., Funct., Bioinf.* **67**: 853-867.

Luo, R., David, L. and Gilson, M. K. 2002. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comp. Chem.* **23**: 1244-1253.

Lusse, S. and Arnold, K. 1996. The Interaction of Poly(ethylene glycol) with Water Studied by 1H and 2H NMR Relaxation Time Measurements. *Macromolecules* **29**: 4251-4257.

MacKerell, A. D. J., Bashford, D., Bellott, M., R. L. Dunbrack, J., Evanseck, J. D., Field, M. J., et al. 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**: 3586–3616.

Makarov, V. A., Feig, M., Andrews, B. K. and Pettitt, B. M. 1998. Diffusion of Solvent around Biomolecular Solutes: A Molecular Dynamics Simulation Study. *Biophys. J.* **75**: 150-158.

Mammen, M., Choi, S.-K. and Whitesides, G. M. 1998. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chem. Int. Ed.* **37**: 2754-2794.

Masatoki, S., Takamura, M., Matsuura, H., Kamogawa, K. and Kitagawa, T. 1995. Raman Spectroscopic Observations of Anomalous Conformational Behavior of Short Poly(oxyethylene) Chains in Water. *Chem. Lett.* **24**: 991-992.

Matsuura, H. and Miyazawa, T. 1969. Vibrational analysis of molten poly(ethylene glycol). *J. Polym. Sci. A-2* **7**: 1735-1744.

Mero, A., Schiavon, O., Pasut, G., Veronese, F. M., Emilitri, E. and Ferruti, P. 2009. A Biodegradable Polymeric Carrier Based on PEG for Drug Delivery. *J. Bioact. Compat. Polym.* **24**: 220-234.

Mongan, J., Simmerling, C., McCammon, J. A., Case, D. A. and Onufriev, A. 2007. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* **3**: 156–169.

Mueller-Plathe, F. and van Gunsteren, W. F. 1994. Can Simple Quantum-Chemical Continuum Models Explain the Gauche Effect in Poly(ethylene oxide)? *Macromolecules* **27**: 6040–6045.

Neyertz, S., Brown, D. and Thomas, J. O. 1994. Molecular dynamics simulation of crystalline poly(ethylene oxide). *J. Chem. Phys.* **101**: 10064-10073.

Nina, M., Im, W. and Roux, B. 1999. Optimized atomic radii for protein continuum electrostatics solvation forces. *Biophys. Chem.* **78**: 89-96.

Nosé, S. and Klein, M. L. 1983. A study of solid and liquid carbon tetrafluoride using the constant pressure molecular dynamics technique. *J. Chem. Phys.* **78**: 6928-6939.

O'Riordan, C. R., Lachapelle, A., Delgado, C., Parkes, V., Wadsworth, S. C., Smith, A. E., et al. 1999. PEGylation of Adenovirus with Retention of Infectivity and Protection from Neutralizing Antibody in Vitro and in Vivo. *Hum. Gene Ther.* **10**: 1349-1358.

Oesterhelt, F., Rief, M. and Gaub, H. E. 1999. Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water. *New J. Phys.* **1**: 6.1-6.11.

Onufriev, A., Case, D. A. and Bashford, D. 2002. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.* **23**: 1297-1304.

Pasut, G., Guiotto, A. and Veronese, F. 2004. Protein, peptide and non-peptide drug PEGylation for therapeutic application. *Expert Opin. Ther. Pat.* **14**: 859-894.

Pertsin, A. J., Grunze, M. and Garbuzova, I. A. 1998. Low-Energy Configurations of Methoxy Triethylene Glycol Terminated Alkanethiol Self-Assembled Monolayers and Their Relevance to Protein Adsorption. *J. Phys. Chem. B* **102**: 4918–4926.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**: 1781-1802.

Pitera, J. W. and Swope, W. 2003. Understanding folding and design: Replica-exchange simulations of ``Trp-cage'' miniproteins. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 7587-7592.

Pitera, J. W. and van Gunsteren, W. F. 2001. The Importance of Solute−Solvent van der Waals Interactions with Interior Atoms of Biopolymers. *J. Am. Chem. Soc.* **123**: 3163-3164.

Prabhu, N. V., Zhu, P. and Sharp, K. A. 2004. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *J. Comp. Chem.* **25**: 2049-2064.

Prime, K. L. and Whitesides, G. M. 1993. Adsorption of proteins onto surfaces containing end-attached oligo(ethylene oxide): a model system using self-assembled monolayers. *J. Am. Chem. Soc.* **115**: 10714–10721.

Privalov, P. L. and Makhatadze, G. I. 1993. Contribution of Hydration to Protein Folding Thermodynamics : II. The Entropy and Gibbs Energy of Hydration. *J. Mol. Biol.* **232**: 660-679.

Ramstein, J. and Lavery, R. 1988. Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. U. S. A.* **85**: 7231-7235.

Richmond, T. J. 1984. Solvent accessible surface area and excluded volume in proteins : Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **178**: 63-89.

Roberts, M. J., Bentley, M. D. and Harris, J. M. 2002. Chemistry for peptide and protein PEGylation. *Adv. Drug Delivery Rev.* **54**: 459-476.

Roux, B. and Simonson, T. 1999. Implicit solvent models. *Biophys. Chem.* **78**: 1-20.

Ryckaert, J.-P., Ciccotti, G. and Berendsen, H. J. C. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**: 327-341.

Scarsi, M., Apostolakis, J. and Caflisch, A. 1997. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A* **101**: 8098-8106.

Sigalov, G., Fenley, A. and Onufriev, A. 2006. Analytical electrostatics for biomolecules: Beyond the generalized Born approximation. *J. Chem. Phys.* **124**: 124902-14.

Simmerling, C., Strockbine, B. and Roitberg, A. E. 2002. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **124**: 11258-11259.

Sitkoff, D., Sharp, K. A. and Honig, B. 1994. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **98**: 1978-1988.

Smith, G. D., Jaffe, R. L. and Yoon, D. Y. 1993. Force field for simulations of 1,2-dimethoxyethane and poly(oxyethylene) based upon ab initio electronic structure calculations on model molecules. *J. Phys. Chem.* **97**: 12752–12759.

Smith, G. D., Jaffe, R. L. and Yoon, D. Y. 1995. Conformations of 1,2-Dimethoxyethane in the Gas and Liquid Phases from Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **117**: 530–531.

Soares, C. M., Teixeira, V. H. and Baptista, A. M. 2003. Protein Structure and Dynamics in Nonaqueous Solvents: Insights from Molecular Dynamics Simulation Studies. *Biophys. J.* **84**: 1628-1641.

Southall, N. T., Dill, K. A. and Haymet, A. D. J. 2001. A View of the Hydrophobic Effect. *J. Phys. Chem. B* **106**: 521-533.

Spassov, V. Z., Yan, L. and Szalma, S. 2002. Introducing an Implicit Membrane in Generalized Born/Solvent Accessibility Continuum Solvent Models. *J. Phys. Chem. B* **106**: 8726-8738.

Still, W. C., Tempczyk, A., Hawley, R. C. and Hendrickson, T. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**: 6127–6129.

Takahashi, Y. and Tadokoro, H. 1973. Structural Studies of Polyethers, (-(CH2)m-O-)n. X. Crystal Structure of Poly(ethylene oxide). *Macromolecules* **6**: 672–675.

Tanford, C. 1968. Protein Denaturation. <u>Adv. Protein Chem.</u> C.B. Anfinsen, Anson, Edsall and Frederic, Academic Press. **Volume 23:** 121-282.

Tasaki, K. 1996. Poly(oxyethylene)−Water Interactions: A Molecular Dynamics Study. *J. Am. Chem. Soc.* **118**: 8459-8469.

Tasaki, K. and Abe, A. 1985. NMR Studies and Conformational Energy Calculations of 1,2-Dimethoxyethane and Poly(oxyethylene). *Polym. J.* **17**: 641-655.

Teeter, M. M. 1991. Water-Protein Interactions: Theory and Experiment. *Annu. Rev. Biophys. Biophys. Chem.* **20**: 577-600.

Tironi, I. G., Sperb, R., Smith, P. E. and van Gunsteren, W. F. 1995. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **102**: 5451-5459.

Tomasi, J. 2004. Thirty years of continuum solvation chemistry: a review, and prospects for the near future. *Theor. Chem. Acc.* **112**: 184-203.

Tsui, V. and Case, D. A. 2000. Molecular Dynamics Simulations of Nucleic Acids with a generalized Born Solvation Model. *J. Am. Chem. Soc.* **122**: 2489-2498.

Vaitheeswaran, S., Yin, H., Rasaiah, J. C. and Hummer, G. 2004. Water clusters in nonpolar cavities. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 17002-17005.

Veronese, F. and Harris, J. M. 2002. Introduction and overview of peptide and protein pegylation *Adv. Drug Delivery Rev.* **54**: 453-456.

Veronese, F. and Pasut, G. 2005. PEGylation, successful approach to drug delivery. *Drug Discovery Today* **10**: 1451-1458.

Vorobjev, Y. N., Almagro, J. C. and Hermans, J. 1998. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins: Struct., Funct., Bioinf.* **32**: 399-413.

Vorobyov, I., Anisimov, V. M., Greene, S., Venable, R. M., Moser, A., Pastor, R. W., et al. 2007. Additive and Classical Drude Polarizable Force Fields for Linear and Cyclic Ethers. *J. Chem. Theory Comput.* **3**: 1120–1133.

Wagoner, J. A. and Baker, N. A. 2006. Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 8331-8336.

Wang, L., Hingerty, B. E., Srinivasan, A. R., Olson, W. K. and Broyde, S. 2002. Accurate Representation of B-DNA Double Helical Structure with Implicit Solvent and Counterions. *Biophys. J.* **83**: 382-406.

Wang, R. L. C., Kreuzer, H. J. and Grunze, M. 1997. Molecular Conformation and Solvation of Oligo(ethylene glycol)-Terminated Self-Assembled Monolayers and Their Resistance to Protein Adsorption. *J. Phys. Chem. B* **101**: 9767–9773.

Warshel, A. and Aqvist, J. 1991. Electrostatic Energy and Macromolecular Function. *Annu. Rev. Biophys. Biophys. Chem.* **20**: 267-298.

Warwicker, J. and Watson, H. C. 1982. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.* **157**: 671-679.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., et al. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**: 765-784.

Weiner, S. J., Kollman, P. A., Nguyen, D. T. and Case, D. A. 1986. An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**: 230-252.

Wesson, L. and Eisenberg, D. 1992. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1**: 227-235.

Yoshida, H., Kaneko, I., Matsuura, H., Ogawa, Y. and Tasumi, M. 1992. Importance of an intramolecular 1,5-CH…O interaction and intermolecular interactions as factors determining conformational equilibria in 1,2-dimethoxyethane: a matrix-isolation infrared spectroscopic study. *Chem. Phys. Lett.* **196**: 601-606.

Zagrovic, B., Snow, C. D., Shirts, M. R. and Pande, V. S. 2002. Simulation of Folding of a Small Alpha-helical Protein in Atomistic Detail using Worldwide-distributed Computing. *J. Mol. Biol.* **323**: 927-937.

Zhou, R. and Berne, B. J. 2002. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water? *Proc. Natl. Acad. Sci. U. S. A.* **99**: 12777-12782.

Zhu, J., Shi, Y. and Liu, H. 2002. Parametrization of a Generalized Born/Solvent-Accessible Surface Area Model and Applications to the Simulation of Protein Dynamics. *J. Phys. Chem. B* **106**: 4844–4853.

Zimmerberg, J. and Parsegian, V. A. 1986. Polymer inaccessible volume changes during opening and closing of a voltage-dependent ionic channel. *Nature* **323**: 36-39.

This page intentionally left blank

Summary

The indirect drug design approach outlined in the first section provides an alternative way to model drugs efficiently in case structural knowledge about the target protein is absent. The procedure employs a specialized energy function to be used with molecular dynamics (MD) simulations of ligand pairs binding to same target, allowing the ligands to exchange the information they have about the target. This yields ligand conformers that are most similar to bound ligand conformers. The proposition in the second section is about successful application of an optimized ether force field for polyethylene glycol (PEG), which is used to connect ligands serving as drugs to form dimeric or even multimeric ligands and thus enhancing drug activity by the multivalent binding effect. The proposed ether force field, specifically optimized for implicit solvent (IS) simulation allows to perform molecular dynamics of ligands connected by PEG that are as realistic as explicit solvent (ES) simulation but at much reduced cost.

Structure based drug design involves structural knowledge on the target protein and its binding pocket. However, in absence of that and if the structural information of a set of ligands binding to the same target is available, one can still use indirect drug-design approach proposed in the present study. However, this requires a set of ligands with chemically different architecture binding to the same target protein. There exist several indirect drug design methods. The present approach yields coordinates of the ligand that is the bound ligand conformer in binding pocket (BLC), which can then be taken further ahead where different strategies of drug design can be pursued.

The assumption that a set of ligands with chemically different architecture binding to the same target protein carry hidden information of their corresponding true BLCs has been exploited in the indirect drug-design approach. Pairwise flexible structure alignment (FSA) using molecular dynamics simulations with specialized energy function employed for this procedure extract the hidden information derived from the molecular similarity of the ligands that allows the ligands to adopt the same space. The conformations of different ligand pairs obtained from FSA once combined yield consensus ligand conformers (CLCs) that should be similar to BLCs. This proposed procedure has been validated on 44 HIV-1 protease (HIV-P) bound ligands of sufficiently diverse chemical composition, the crystal structures of which are

available in PDB. The described four different cluster of HIV-P BLCs based on similarity measures are consistent with the H-bond patterns of the ligands bound to HIV-P in the crystal structures exhibiting four different binding modes.

MD simulations of biomolecules with realistic representations of cellular environments remain challenging. Explicitly considering water molecules during molecular modeling studies is the most realistic theoretical approach but is computationally expensive. Implicit solvent (IS) approaches, which include the effect of water in a potential of mean force have a clear advantage over explicit solvent (ES) in efficiency. For the attractive features of the IS methodology one has to pay a price in making certain approximations, which may not be sufficient to provide agreement between ES and IS approaches.

In our study, an accurate polyether force field for IS MD simulations was developed that matches local and global conformations of 1,2-dimethoxy-ethane (DME) and polyethylene glycol (PEG), respectively. In the IS model, since solvent molecules are missing, the competition of attractive van der Waals (vdW) interactions between solute-solute and solute-solvent atom pairs is absent. As a consequence the remaining attractive vdW interactions between solute atom pairs yield too compact structures. To regain balance of attractive interactions in the IS model, surface energy terms are turned off and attractive vdW interactions are reduced to 90% or as an alternative even slightly negative surface energies with no change in the vdW attractive interactions are employed. However, to obtain quantitatively the same local and global distributions of PEG conformers as in ES, additional force field adjustments involving torsion potentials and 1-4 and 1-5 atom pair Coulomb interactions are also suggested. The proposed ether force field, specifically optimized for IS simulation conditions is equally valid for monomeric and polymeric ethylene glycol.

Zusammenfassung

Der im ersten Absatz beschriebene indirekte Ansatz erlaubt eine effiziente Alternative für das *drug design* im Falle nicht vorhandener Strukturinformation. Die Methode nutzt eine spezialisierte Energiefunktion für Molekulardynamik (MD) Simulationen von Ligandenpaaren die das selbe Zielprotein binden, die es den Liganden erlaubt ihre Informationen über das Zielprotein auszutauschen. Dies führt zu Ligandenkonformationen die extreme Ähnlichkeit mit denen gebundener Liganden aufweisen. Der Vorschlag im zweiten Absatz beschreibt die erfolgreiche Anwendung optimierter Etherkraftfelder für polyethylenglykol (PEG). PEG findet im *drug design* als Linker zwischen Liganden Anwendung, um sie zu Dimeren oder Multimeren zusammenzufassen und dadurch die Wirkung über multivalente Bindungen zu erhöhen. Das vorgestellte Etherkraftfeld, erlaubt MD Simulationen mit implizitem Lösemittel (*implicit solvent*: IS) für Liganden die über PEG verknüpft sind mit der selben Güte wie mit explizitem Lösemitteln (*explicit solvent*: ES) aber mit deutlich verringerter benötigter Rechnerleistung.

Für strukturbasiertes *design* von Medikamenten (*drugs*) benötigt man Strukturinformationen über das Zielprotein und dessen Bindungstasche. Falls diese Informationen nicht verfügbar sind, kann man immer noch eine indirekte Methode des *drug designs* einsetzen. Dazu benötigt man allerdings ein Ensemble von Liganden unterschiedlicher chemischer Architektur, die alle an dasselbe Zielprotein binden. Es gibt zahlreiche Methoden des indirekten *drug design*. Neu an dieser Arbeit ist, dass mit der hier verwendeten Methode des *drug design* explizit die Koordinaten der Liganden konstruiert werden, die dessen Struktur in der Bindungstasche wiedergeben [BLC *(bound ligand conformer)*]. Diese Strukturinformation ermöglicht es, neue Wege beim dem design von Medikamenten zu beschreiten.

Die Arbeit beruht auf der Annahme, dass ein Ensemble von Liganden, deren chemischer Aufbau sich unterscheidet, die aber an dasselbe Protein binden, versteckte Informationen ihres entsprechenden BLC in sich tragen. Die hier eingesetzte Methode des paarweisen flexiblen Strukturalignment (FSA) nutzt Molekulardynamik (MD) Simulationen von Ligandenpaaren mit für diese Aufgaben maßgeschneiderten Energiefunktionen, um die versteckten Strukturähnlichkeiten der Liganden zu extrahieren. Diese Energiefunktionen sind so konstruiert, dass die Liganden beim Strukturalignment das gleiche Volumen einnehmen können. Die Konformationen

verschiedener Ligandenpaare, die über FSA erhalten werden ergeben in Kombination eine Konsensuskonformation für die Liganden (*consensus ligand conformers*: CLCs), die dann Ähnlichkeit mit dem BLC aufweisen. Der vorgeschlagene Ansatz ist anhand von 44 der PDB entnommenen HIV-1 Proteaseliganden (HIV-P) von ausreichend unterschiedlichem chemischem Aufbau überprüft worden. Die auf Grundlage der Strukturähnlichkeiten gefundenen vier unterschiedlichen Cluster von HIV-P BLCs sind konsistent mit den vier in den Kristallstrukturen vorgefundenen Mustern an Wasserstoffbrücken zwischen den gebundenen Formen der Liganden und der HIV-P.

MD Simulation von Biomolekülen mit einer realistischen Repräsentation der zellulären Umgebung ist eine Herausforderung für computergestützte theoretische Methoden. Die Verwendung von explizit modelliertem Wasser (*explicit solvent*: ES) ist der realistischste Ansatz, erfordern aber enorme Rechenzeiten. MD Simulationsverfahren mit implizit modelliertem Wasser (*implicit solvent*: IS), bei denen den Effekt des Wassers durch ein angepaßtes Kraftfeld ersetzt wird, haben gegenüber ES-Verfahren einen großen Vorteil, da sie deutlich weniger Rechenzeit benötigen. Der Preis den man für diesen Vorteil zahlt, sind aber Abweichungen in den Simulationsdaten, die durch die Näherungen im IS Kraftfeld bedingt sind.

Im Rahmen dieser Arbeit wurde ein Polyether-Kraftfeld für IS MD Simulationen entwickelt, das lokale Konformationen von 1,2-dimethoxy-ethane (DME) und globale Konformationen von Polyethylenglykol (PEG) akkurat beschreibt. Im IS Model fehlen die Wassermoleküle. Dies hat zur Folge, dass die anziehenden van der Waals (vdW) Wechselwirkungen der Atompaare in PEG einerseits und der Atompaare zwischen Wassermolekülen und PEG andererseits nicht ausbalanciert ist. Die im IS Model verbleibenden anziehenden vdW Wechselwirkungen zwischen den Atompaaren in PEG führen, verglichen mit einem ES Model, zu kompakteren Strukturen. Um die Balance der anziehenden vdW Kräfte im IS Model wieder herzustellen, werden entweder die Oberflächenenergieterme abgeschaltet und die vdW Wechselwirkungen auf 90% reduziert oder alternativ die Oberflächenenergieterme leicht negativ gesetzt, während die vdW Wechselwirkungen unverändert bleiben. Jedoch werden, um quantitativ die gleichen PEG Konformationen wie im ES-Model zu erreichen, zusätzliche Kraftfeldänderungen für die Torsionspotentiale und die Coulomb Wechselwirkungen der 1-4 und 1-5 Atompaare vorgeschlagen. Das vorgestellte IS Etherkraftfeld für MD Simulationen ist für Monomere wie für Polyether gleichermaßen gültig.

Curriculum vitae

For reasons of data protection,
the curriculum vitae is not included in the online version

Publications

1.	**Merging implicit with explicit solvent simulations: Polyethylene glycol**; <u>Alok Juneja</u>, Jorge Numata, Lennart Nilsson and E. W. Knapp; *J. Chem. Theory Comput.*, 2010 (DOI: 10.1021/ct100075m)
2.	**Bound ligand conformer revealed by flexible structure alignment in absence of crystal structures: Indirect drug design probed for HIV-1 protease inhibitors**; <u>Alok Juneja</u>, Henning Riedesel, Milan Hodozcèk and E. W. Knapp; *J. Chem. Theory Comput.*, 2009, 5 (3), pp 659–673 (DOI**:** 10.1021/ct8004886) *Accomplishment:* Article figure on the cover page of JCTC http://pubs.acs.org/action/showLargeCover?issue=338107928
3.	**Discerning Relationships Among Human Cytochrome P450s by Computational Analyses**; <u>Alok Juneja</u>, Kulvinder Singh Saini and Rajeev Soni; *Bioinformatics Trends*, 2006, 1(2), pp 1-12

Posters & Conferences

1.	**Indirect approach to drug design using MD for flexible structure alignment: Application to HIV-1 protease inhibitors;** *<u>Alok Juneja</u>, Henning Riedesel, Milan Hodozcèk, Ernst Walter Knapp*; Seventh European workshop in drug design; Certosa di Pontignano, Siena (Italy); May 24 - 30, 2009
2.	**Indirect drug design using MD for flexible structure alignment, application to HIV-1 protease inhibitors;** *<u>Alok Juneja</u>, Henning Riedesel, Milan Hodozcèk, Ernst Walter Knapp*; Latest Advances in Drug Discovery Informatics; Bryn Mawr College, Philadelphia, USA; October 13-16, 2008
3.	**Flexible structure alignment of HIV-1 protease inhibitors: Indirect drug design approach;** *<u>Alok Juneja</u>, Henning Riedesel, Milan Hodozcèk, Ernst Walter Knapp;* The 8th International Conference on Chemical Structures; The Netherlands; June 01-05, 2008
4.	**Indirect drug design by flexible structure alignment applied to HIV-1 protease inhibitors;** *<u>Alok Juneja</u>, Milan Hodozcèk, Ernst Walter Knapp*; 3. German Conference on Chemoinformatics - 21. CIC-Workshop Goslar; November 11-13, 2007
5.	**Drug design by flexible structure alignment applied to HIV-1 protease inhibitors: Indirect approach;** *<u>Alok Juneja</u>, Milan Hodozcèk, Ernst Walter Knapp*; German Conference on Bioinformatics 2007 Potsdam; September 26-28, 2007
6.	**Discerning relationships among cytochrome P450s of different animal species by computational analyses;** *<u>Alok Juneja</u>, Rajeev Soni, Kulvinder Singh Saini;* American Society of Biochemistry and Molecular Biology (ASBMB) Annual Meeting and 8th IUBMB Conference; Boston, Massachusetts, USA; June 12-16, 2004